

Marcel Tolentino Pinheiro de Oliveira

Análise Comportamental de Bovinos Baseada em Trajetórias Semânticas Aplicada à Pecuária de Precisão

Campo Grande - Mato Grosso do Sul
26 de outubro de 2013

Marcel Tolentino Pinheiro de Oliveira

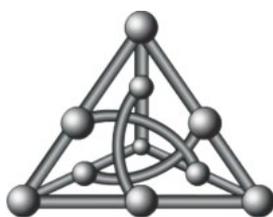
Análise Comportamental de Bovinos Baseada em Trajetórias Semânticas Aplicada à Pecuária de Precisão

Dissertação apresentada à Faculdade de
Computação da Universidade Federal do
Mato Grosso do Sul como parte dos requisi-
tos para obtenção do título de **Mestre em
Ciência da Computação**.

Orientação: Prof. Dra. Hana Karina Salles Rubinsztein

Co-orientação: Dr. Pedro Paulo Pires

Área de Concentração: Sistemas de Computação



Faculdade de Computação
Universidade Federal de Mato Grosso do Sul
Campo Grande - Mato Grosso do Sul
26 de outubro de 2013

*Para Irene e Persio, meus avós paternos,
que hoje moram em um lugar melhor!*

“Existem muitas hipóteses em ciência que estão erradas. Isso é perfeitamente aceitável, eles são a abertura para achar as que estão certas.”

Carl Sagan

Sumário

Agradecimentos	7
Resumo	8
Abstract	9
Lista de Figuras	10
Lista de Tabelas	12
Lista de Siglas	14
1 Introdução	16
1.1 Contextualização	16
1.2 Motivação e Objetivos	17
1.3 Estrutura da Dissertação	18
2 Sistemas de Sensores Sem Fio	20
2.1 Rede de Sensores Sem Fio	20
2.2 Características de uma Rede de Sensores	21
2.2.1 Auto-organização	22
2.2.2 Comunicação Sem Fio <i>Ad-Hoc</i>	22
2.2.3 Comunicação <i>Multi-hop</i>	23
2.2.4 Grande Quantidade de Nós Distribuídos	24
2.2.5 Limitação Energética	24
2.2.6 Esforço Cooperativo	24

2.3	Roteamento em uma Rede de Sensores	25
2.3.1	Roteamento Hierárquico	26
2.3.2	Roteamento Plano	26
2.4	Redes Oportunísticas	27
2.4.1	Roteamento Oportunístico	27
2.4.1.1	Roteamento Baseado em Disseminação	28
2.4.1.2	Roteamento Baseado em Contexto	28
2.4.1.3	Roteamento Baseado em Infraestrutura Fixa	29
2.4.1.4	Roteamento Baseado em Infraestrutura Móvel	29
2.5	Considerações Finais	29
3	Trajetórias Semânticas	30
3.1	Mineração de Dados	30
3.1.1	Processo de Descoberta de Conhecimento	31
3.1.2	Mineração de Dados Espaço-Temporais	33
3.2	Trajetórias	34
3.2.1	Trajetória Semântica	35
3.2.2	Algoritmos de Trajetórias Semânticas	36
3.2.2.1	SMoT ou IB-SMoT	37
3.2.2.2	CB-SMoT	38
3.2.2.3	Padrão <i>Avoidance</i>	38
3.3	Considerações Finais	41
4	Trabalhos Correlatos	42
4.1	Monitoramento Animal	42
4.2	Mineração de Dados na Agropecuária	44
4.3	Trabalhos realizados entre Embrapa e FACOM	45
4.4	Considerações Finais	47
5	Sistema de Análise de Trajetórias Bovinos	48
5.1	Motivação	48

5.2	Arquitetura do Sistema	49
5.2.1	Camada de Apresentação	49
5.2.1.1	Alimentação do Sistema com Trajetórias dos Animais . . .	49
5.2.1.2	Alimentação do Sistema com Dados Referentes ao Ambiente	50
5.2.1.3	Visualização de Objetos Espaciais em Mapa	51
5.2.2	Camada de Controle	51
5.2.2.1	Cálculos com Tipos Espaciais	51
5.2.2.2	Incorporação de Semântica às Trajetórias	52
5.2.2.3	Localização de Regiões de Alta Concentração	52
5.2.2.4	Identificação de Áreas Evitadas pelos Animais	53
5.2.2.5	Manipulação de Arquivos de Entrada sobre Trajetórias Animais	54
5.2.2.6	Manipulação de Arquivos KML	56
5.2.3	Camada de Modelagem	56
5.2.3.1	Operações no Banco de Dados	56
5.2.3.2	Operações sobre Dados Espaciais	57
5.3	Implementação	57
5.3.1	Tecnologias Envolvidas	57
5.3.2	Funcionalidades Implementadas	58
5.3.2.1	Submissão das Trajetórias Bovinas	58
5.3.2.2	Gerenciamento dos Dados Referentes ao Ambiente	59
5.3.2.3	Adicionar Semântica para Trajetórias	61
5.3.2.4	Verificação da Presença Animal em uma Área de Interesse	62
5.3.2.5	Verificação da Presença Animal em uma Área de Sombra .	63
5.3.2.6	Distância Percorrida, Duração e Velocidade Média de uma Trajetória	64
5.3.2.7	Identificação de Regiões de Concentração	64
5.3.2.8	Verificar Existência de Pontos de Concentração em Áreas de Sombra	66
5.3.2.9	Identificação Áreas Evitadas pelos Animais	67
5.3.2.10	Visualização de Trajetórias	69

5.4	Considerações Finais	70
6	Estudo de Caso	71
6.1	Definição de Estudo	71
6.1.1	Uso de Dados Reais	73
6.1.2	Ambiente Simulado	74
6.2	Técnicas de Trajetórias Semânticas em Dados Reais e Simulados	77
6.2.1	Verificar o período em que cada animal permaneceu em uma determinada região	77
6.2.1.1	Resultados Obtidos com Dados Reais	78
6.2.1.2	Resultados Obtidos com Dados Simulados	79
6.2.2	Localizar Regiões de Concentração na Área de Pastagem	80
6.2.2.1	Resultados Obtidos com Dados Reais	80
6.2.2.2	Resultados Obtidos com Dados Simulados	84
6.2.3	Identificar Áreas Evitadas pelos Animais	88
6.2.3.1	Resultados Obtidos com Dados Reais	88
6.2.3.2	Resultados Obtidos com Dados Simulados	89
6.3	Considerações Finais	91
	Conclusão	92
	Referências	95

Agradecimentos

A minha família pela força e pela presença nos momentos felizes e principalmente nos difíceis. Meus pais, Natanael e Eliana, meus avós maternos, Helvécio e Jenica, meu irmão e grande amigo, André, e a minha cunhadinha querida, Juliana.

Um agradecimento especial à minha querida namorada, Ariane, pelo seu apoio, carinho e por sempre acreditar na minha capacidade em superar desafios.

Aos meus companheiros de mestrado, Geraldo, Vinícius, Márcio, Pedro Neves, Fabrício, Leandro Oliveira e Leandro Ishi pelo companheirismo durante as complicadas disciplinas em nossa pós-graduação.

Aos meus professores da FACOM/UFMS pelos ensinamentos, tanto em minha graduação quanto em meu mestrado. Agradeço até mesmo aqueles que dificultaram o caminho, afinal são as dificuldades que nos fazem crescer. Agradeço também aos meus alunos de Estágio de Docência, pela incrível experiência que foi lecionar a vocês.

Um agradecimento especial a minha orientadora, Dra. Hana Karina Salles Rubinsztejn, por ter aceitado o desafio da minha proposta de projeto e por me orientar sempre pelo caminho correto.

Aos meus companheiros da Embrapa, Quintino, Camilo, Daniel, Bianca e Ricardo pela disposição em sempre ajudar em minhas necessidades.

Um agradecimento especial ao Dr. Pedro Paulo Pires, por acreditar em minha capacidade e me apresentar esse projeto fascinante. Espero não ter desapontado.

Ao grupo de pesquisa de Trajetórias de Objetos Móveis da UFSC e da UFRGS, em especial a Dra. Vania Bogorny e a Me. Alisson Moscato Loy pelos conhecimentos compartilhados da área e, principalmente, por se mostrarem sempre abertos a ajudar em minhas dúvidas.

Ao pesquisador Norte Americano, Edward Raff, pelos conhecimentos em Mineração de Dados compartilhados.

Por fim, agradeço a Deus nosso PAI, pela sua infinita bondade.

Resumo

Com o mercado exigindo cada vez mais informação sobre os produtos que consome, a agricultura e a pecuária passaram a desenvolver novos meios de controle e de produção a fim de atender essas exigências. Com objetivo de cumprir essas demandas, o conceito do monitoramento animal surgiu para a pecuária como uma solução para o controle de qualidade. Assim, a ideia é monitorar cada animal da propriedade com a finalidade de fornecer ao produtor o resumo diário das atividades do bovino. Na maioria dos casos, o monitoramento é realizado através de colares com um sensor GPS embutido, armazenando o posicionamento do animal em um determinado ambiente. A proposta desta pesquisa é a construção de um sistema computacional que seja capaz de identificar padrões comportamentais e informações relevantes através das trajetórias dos animais, produzidas durante o pastejo. Para isso, são utilizadas metodologias de trajetória semântica. Essas descobertas irão auxiliar os pecuaristas e pesquisadores na tomada de decisões. Este projeto consiste em mais uma parceria entre Embrapa - Gado de Corte e a FACOM/UFMS, na área de Pecuária de Precisão.

Palavras-Chave: Pecuária de Precisão, Monitoramento Bovino, Trajetória Semântica.

Abstract

With the market demanding more information about the products they consume, agriculture and livestock started to develop new means of production and control in order to meet their requirements. To meet these demands, the concept of cattle monitoring emerged as a solution for quality control. The idea is to monitor each animal in order as a means the producer the daily summary of the activities of cattle. In most cases, monitoring is carried out by means of collars with a built-in GPS sensor, storing the position of the animal in a given environment. The purpose of this research is to build a computational system that is able to identify behavioral patterns and relevant information through the paths of animals, produced during grazing. With this purpose, methodologies for trajectory semantics are used. These discoveries will help researchers and farmers in decision making. This project consists of another partnership between Embrapa - Gado de Corte and FACOM/UFMS, in Precision Livestock area.

Key-Words: Precision Livestock, Cattle Monitoring, Trajectory Semantics.

Lista de Figuras

2.1	Um exemplo simplificado das unidades funcionais de um sensor sem fio.	21
2.2	Redes com Infraestrutura x Redes Sem Infraestrutura (<i>Ad-Hoc</i>).	22
2.3	Um exemplo de uma rede de sensores utilizando a comunicação <i>multi-hop</i>	23
2.4	Roteamento Centrado em Endereços x Roteamento Centrado em Dados [23].	25
2.5	Roteamento Hierárquico de dois grupos de rede de sensores sem fio.	26
2.6	Roteamento Plano com um nó como sorvedouro.	27
2.7	Taxonomia do Roteamento Oportunístico.	28
3.1	Fases do <i>Knowledge Discovery in Databases</i> (KDD) [38].	31
3.2	Fase de Preparação dos Dados [43].	32
3.3	Todos os tipos relações espaciais ilustrados [47].	34
3.4	Exemplos de padrões de comportamento em grupos de trajetórias.	35
3.5	Exemplos de Trajetórias sem e com semântica.	36
3.6	Esquerda: um conjunto de pontos sem semântica. Direita: o mesmo conjunto após cruzar com os dados geográficos fornecidos.	37
3.7	(1) é resultado do algoritmo IB-SMoT e a (2) é resultado do CB-SMoT [53].	38
3.8	Exemplo do Padrão <i>Avoidance</i>	39
3.9	Um exemplo com uma região de interesse sobre objeto-alvo.	39
3.10	Exemplo de subtrajetórias direcionadas ao alvo.	40
3.11	Exemplo de uma trajetória que passa pela região de incremento de confiança.	40
4.1	Um colar do programa Zebranet sendo colocado em uma zebra.	43
4.2	Bovinos com os colares no experimento do <i>Belmont Research Station</i> [57].	43
4.3	Leitora de RFID da Destron	46

4.4	Alguns transponders TXUM00	46
5.1	Arquitetura da Plafatorma.	49
5.2	O Colar GPS (A) e a Estação Base que armazena os dados dos colares (B).	50
5.3	Formula de Haversine.	52
5.4	Exemplo do formato dos arquivos texto criado por um colar da OTAG.	54
5.5	Exemplo do formato dos dados em um ARFF.	55
5.6	Exemplo da formatação de código de um KML.	56
5.7	Janela referente a submissão de arquivos de trajetórias.	58
5.8	Janela Principal das seções Área de Pastagem e Área de Sombra.	59
5.9	Janela de Adição de Área de Pastagem e Área de Sombra.	60
5.10	ARFF com as coordenadas de uma área.	61
5.11	Tela de adição de semântica para trajetórias.	61
5.12	As principais perguntas da seção Perguntas e Respostas.	62
5.13	Relação dos dias e do momento em que as trajetórias do colar 0006 passaram pela área selecionada.	63
5.14	Relação dos momentos em que as trajetórias do colar 0006 interceptaram uma área de sombra.	63
5.15	Janela principal da seção Padrões em Trajetórias.	64
5.16	Janela principal da seção Agrupamentos em Trajetórias.	64
5.17	Janela de resultado do agrupamento executado.	65
5.18	Regiões de Concentração representadas em um mapa.	66
5.19	Tela com a resposta da pergunta “Existe pontos se concentrando na Sombra?”	66
5.20	Janela principal da seção Padrões em Trajetórias.	67
5.21	Janela com o resultado do padrão aplicado.	67
5.22	Áreas evitadas pelas trajetórias em mapa.	68
5.23	Janela principal da seção Observar Trajetória.	69
5.24	Uma trajetória animal representada em um mapa.	69
6.1	A Área de Pastagem do Mangueiro Digital	73
6.2	Equação para a seleção do próximo ponto.	75

6.3	O Simulador de Trajetórias do SATB.	75
6.4	A área de pastagem do Mangueiro Digital dividida em quatro áreas de interesse.	77
6.5	Regiões de concentração do estudo empírico (parâmetros: 20 e 6).	81
6.6	Regiões de concentração do estudo empírico (parâmetros: 15 e 8).	84
6.7	Regiões de concentração da simulação (parâmetros: 20 e 6).	85
6.8	Regiões de concentração da simulação (parâmetros: 15 e 8).	86
6.9	Regiões de concentração da simulação (parâmetros: 10 e 10).	87
6.10	Área de Pastagem após a aplicação do padrão <i>avoidance</i> sobre o conjunto de trajetórias reais.	89
6.11	Área de Pastagem após a aplicação do padrão <i>avoidance</i> sobre o conjunto de trajetórias simuladas.	90

Lista de Tabelas

6.1	Relação dos dados de posicionamento coletados em campo.	72
6.2	Trajетórias reais selecionadas pelo critério estabelecido.	73
6.3	Relação final dos pontos usados no Estudo de Caso.	74
6.4	Diferença entre os quatro conjuntos de trajetórias.	76
6.5	Dados Reais - Período de tempo (em %) em cada trajetória permaneceu em cada área de interesse do ambiente de pastagem.	78
6.6	Dados Simulados - Período de tempo (em %) em cada trajetória permaneceu em cada área de interesse do ambiente de pastagem.	79
6.7	Relação de pontos em área de sombra do colar 0004	82
6.8	Relação de pontos em área de sombra do colar 0006	83

Lista de Siglas

API *Application Programming Interface*

ARFF *Attribute-Relation File Format*

ASCII *American Standard Code for Information Interchange*

CAR *Context-Aware Routing*

CB-SMoT *Clustering-Based Stops and Moves of Trajectories*

CRUD *Create Remote Update Delete*

DBSCAN *Density-Based Spatial Clustering of Applications with Noise*

DC-DC *Direct Current to Direct Current*

ECTC *Eletronic Collars to Track Cattle*

EMBRAPA *Empresa Brasileira de Pesquisa Agropecuária*

FACOM *Faculdade de Computação*

GB *Gigabyte*

GPS *Global Positioning System*

IBGE *Instituto Brasileiro de Geografia e Estatística*

IB-SMoT *Intersection-Based Stops and Moves of Trajectories*

IP *Internet Protocol*

JDOM *Java Document Object Model*

JDBC *Java Database Connectivity*

KDD *Knowledge-Discovery in Databases*

KML *Keyhole Markup Language*

LEACH *Low-Energy Adaptive Clustering Hierarchy*

MEMS *Micro-Electro-Mechanical Systems*

MULE *Mobile Ubiquitous LAN Extension*

MV *Meetings and Visits*

OTAG *Operational Management and Geodecisional Prototype to Track and Trace Agricultural Production*

P2P *Peer-to-peer*

PC *Personal Computer*

PEGASIS *Power-Efficient Gathering in Sensor Information Systems*

RFID *Radio-Frequency IDentification*

RSSF *Rede de Sensores Sem Fio*

SATB *Sistema de Análise de Trajetórias Bovinas*

SAPI *Sistema Agropecuário de Produção Integrada*

SAR *Sequential Assignment Routing*

SGBD *Sistema de Gerenciamento de Banco de Dados*

SI *Sistema de Informação*

SMoT *Stops and Moves of Trajectories*

SPIN *Sensor Protocols for Information via Negotiation*

SQL *Structured Query Language*

SWIM *Shared Wireless Infostation Model*

TEEN *Threshold sensitive Energy Efficient sensor Network*

UFMS *Universidade Federal de Mato Grosso do Sul*

USB *Universal Serial Bus*

XML *Extensible Markup Language*

Capítulo 1

Introdução

1.1 Contextualização

A pecuária é uma atividade econômica de vital importância para a economia nacional. Tendo em vista as exigências de nações importadoras e da própria competição existente no setor, o Brasil cada vez mais tem investido na bovinocultura a fim de, não somente cumprir as metas impostas pelo mercado exterior, mas também melhorar a qualidade da carne e dos produtos derivados.

Segundo o Departamento de Agricultura dos Estados, em 2011 o Brasil ocupava a segunda posição mundial em rebanho bovino, sendo a primeira ocupada pela Índia. Na produção de carne bovina, no entanto, o quadro é um pouco diferente, sendo os Estados Unidos o maior produtor mundial e o Brasil, o segundo. A produção per capita de carcaça bovina brasileira ficou em torno de 35 kg/hab/ano [1].

Ainda em 2011, segundo dados do IBGE, o efetivo de bovinos teve aumento de 1,6% em relação ao ano de 2010, contabilizando cerca de 212,798 milhões de cabeças. Este efetivo encontra-se disperso por todo o território nacional, embora seja encontrado em maior número na região Centro-Oeste do país (34,1%). As demais regiões apresentam os seguintes percentuais de participação: Norte (20,3%), Sudeste (18,5%), Nordeste (13,9%) e Sul (13,1%). O Estado de Mato Grosso possuía o maior efetivo de bovinos, 13,8%; seguido por Minas Gerais, com 11,2%; Goiás, com 10,2%; e Mato Grosso do Sul, com 10,1% [1].

Diversos fatores foram determinantes para a conquista brasileira no comércio internacional da carne bovina, dentre eles, as ações desenvolvidas e em continuidade, em prol da erradicação da febre aftosa. Outra característica de valorização é a produção de alimento seguro, já que a maior parte do rebanho brasileiro é alimentada em pasto. Além disso, a iniciativa da rastreabilidade da carne bovina destinada à exportação, especificamente para União Europeia, tem contribuído significativamente para o aumento das expectativas dos consumidores internacionais, quanto à segurança dos alimentos brasileiros [2].

Visando a qualidade dentro da agropecuária, a Embrapa – Gado de Corte produziu o

manual de **Boas Práticas Agropecuárias - Bovinos de Corte** [2], um instrumento destinado a orientar o produtor a como produzir para a indústria e para o mercado consumidor, em sistemas produtivos sustentáveis. A adoção das boas práticas tem o objetivo principal de garantir a produção de alimentos seguros e com atributos de qualidade que atendam aos interesses dos grandes mercados [3].

Neste contexto, a pecuária de precisão diz respeito a inovações tecnológicas que monitoram o animal no seu ambiente, sendo uma forma eficiente de gerenciar os sistemas de produção animal. Consiste, também, na medição de diferentes parâmetros dos animais (como por exemplo, o tempo de permanência de um animal em uma região do pasto), na modelagem desses dados para selecionar a informação desejada e no uso desses modelos em tempo real, visando o monitoramento e controle de animais e rebanhos [4].

A base da proposição da pecuária de precisão está no monitoramento individual e remoto dos animais e pastagens, possibilitado pelos novos avanços nas áreas de tecnologia. As aplicações dessas tecnologias são potencialmente imensas, inclusive podendo contribuir para a rastreabilidade dos produtos, uma vez que oferece registros quanto à origem dos produtos e o seu meio de produção [5].

Uma das possibilidades de monitoramento mais conhecidas é por meio do uso de sensores. Um sensor é um dispositivo que recebe e responde a sinais ou estímulos e pode ser usado para medir grandezas físicas e, caso possua um meio de transmissão, pode transmitir essas informações para outro lugar. O uso de sensores para monitoramento existe há décadas, sendo usados em aplicações de climatologia, biologia, militares e indústria.

Na Pecuária de Precisão, o GPS (*Global Position System*) é um dos sensores mais utilizados no monitoramento animal. Embutido em um aparelho carregado pelo animal (geralmente na forma de um colar sobre o pescoço), o sensor recebe em um determinado intervalo de tempo, a posição geográfica do animal. Usando metodologias provenientes da Inteligência Artificial, um conjunto de dados georreferenciados coletados por esse GPS pode, além mostrar a trajetória do bovino, revelar padrões comportamentais e novos conhecimentos a respeito daquele animal. Por meio de uma análise em um conjunto de dados espaço-temporais, é possível, por exemplo, interpretar as atitudes de um bovino em uma pastagem (monitorado por um colar com um GPS embutido). Se o animal vai até o cocho lambar sal, se frequenta certas regiões do pasto, se evita alguma região do pasto, se ele tem preferência pela sombra de uma determinada árvore, entre outros [6].

Ao longo dos anos, instituições de pesquisas como a Embrapa, executaram vários projetos de monitoramento animal e ambiental com uso de sensores. A Embrapa - Gado de Corte e a FACOM/UFMS, por exemplo, já desenvolveram algumas soluções para problemas na área de pecuária de precisão. Essas soluções envolvem estudos na área de monitoramento animal e identificação eletrônica por radiofrequência.

1.2 Motivação e Objetivos

Nos últimos anos, a exploração da bovinocultura de corte vem sofrendo profundas interferências em todos os segmentos que compõem a cadeia de produção da carne bovina. A

mudança de comportamento do consumidor, que passou a exigir produtos de maior qualidade, a intensificação na produção de suínos e aves, que se tornaram mais competitivos, e a abertura do Brasil ao mercado externo, são alguns precursores de tal mudança [7].

A necessidade de se produzir de forma eficiente e competitiva, exige desse setor, o estabelecimento de um novo conceito, ou seja, a produção de carne de qualidade. O produto final é definido como alimento e tem suas características definidas pelo consumidor. Apesar de o consumo da carne bovina ser influenciado pela renda per capita da população, pelo seu valor e pelo preço das outras carnes, a preferência do consumidor pode ser influenciada por campanhas que procuram enfatizar determinadas qualidades das carnes concorrentes, especialmente aquelas relacionadas com aspectos de saúde [8].

Para atingir o nível de exigência do mercado consumidor, é necessário que o pecuarista possua um conhecimento apurado sobre cada animal de sua propriedade. Entretanto, levando em conta a quantidade de animais, conhecer o dia-a-dia de cada um, sem qualquer auxílio de ferramentas de monitoramento remoto, é uma tarefa quase impossível para o produtor. E além das ferramentas, é necessário que os dados coletados sejam analisados e interpretados para assim se transformarem em informação relevante ao dono dos animais.

Uma alternativa para a análise e interpretação de dados oriundos de monitoramento é o uso do conceito de Trajetórias Semânticas. Por meio de um conjunto de dados espaço-temporais produzidos pelas ferramentas de monitoramento animal, além de conferir semântica (significado, sentido) aos dados, as técnicas dessa área são capazes de extrair informação relevante e padrões ocultos sobre o comportamento dos animais monitorados.

Portanto, o principal objetivo deste estudo é a utilização de metodologias e tecnologias de manipulação e análise comportamental de dados espaço-temporais (ou georreferenciáveis) de bovinos a fim de extrair desses conjuntos um conhecimento apurado sobre os animais monitorados pelo produtor/pesquisador.

Alguns objetivos mais específicos são:

- Desenvolvimento de um *software* que disponibilize ao usuário as metodologias e tecnologias necessárias para o estudo;
- Integrar ao *software* um sistema gerenciador de banco de dados capaz de armazenar os objetos espaciais e dados organizados pelo sistema;
- Utilizar técnicas de trajetórias semânticas para a extração de conhecimento relevante dos dados brutos;
- Elaborar um modo de visualização para os resultados das técnicas de análise de trajetórias.

1.3 Estrutura da Dissertação

Neste capítulo foi apresentado uma breve introdução ao assunto, a motivação e os objetivos para o desenvolvimento desta pesquisa.

Os próximos capítulos estão organizados da seguinte maneira: o Capítulo 2 apresenta a fundamentação teórica a respeito de Sistemas de Sensores Sem Fio, o Capítulo 3 apresenta a fundamentação teórica a respeito de Mineração de Dados e Trajetórias Semânticas, o Capítulo 4 lista os trabalhos correlatos referentes às áreas de monitoramento animal e mineração de dados na agropecuária, além de alguns trabalhos realizados entre a Embrapa - Gado de Corte e a FACOM/UFMS, o Capítulo 5 apresenta o projeto de pesquisa e o sistema desenvolvido, o Capítulo 6 apresenta alguns estudos de caso para a plataforma e finalmente, são apresentadas as conclusões da pesquisa e os possíveis trabalhos futuros, e por fim as referências bibliográficas que compõem este estudo.

Capítulo 2

Sistemas de Sensores Sem Fio

Neste capítulo será formalizado o conceito e a fundamentação teórica de Sistemas de Sensores Sem Fio (*Wireless Sensor Systems*).

2.1 Rede de Sensores Sem Fio

Sendo componente principal desta tecnologia, um sensor é um aparelho que produz uma resposta a uma alteração mensurável em uma condição física. Logo, ele pode ser projetado para medir a temperatura, a pressão do ar, a intensidade de luz, entre outras grandezas físicas.

Segundo Akyildiz *et al.* [9], os recentes avanços em comunicações possibilitaram o desenvolvimento de sensores pequenos e baratos baseados em *Micro-Electro-Mechanical Systems* (MEMS), com capacidade de comunicação via rede sem fios e processadores de baixo consumo de energia. O uso de dispositivos pequenos e baratos possibilita o uso em larga escala para diversos tipos de sistemas. Um exemplo conhecido da tecnologia é o GPS, um aparelho móvel com um sistema de navegação por satélite embutido que envia informações sobre o posicionamento geográfico do equipamento em qualquer horário e em qualquer condição climática.

Conceitualmente, uma Rede de Sensores Sem Fio (RSSF) consiste em um grande número dispositivos compactos e autônomos (chamados nós sensores) densamente distribuídos em uma região de interesse [10, 11]. Uma RSSF é utilizada como uma ferramenta de sensoriamento distribuído de fenômenos (entidades de interesse que estão sendo monitorados), processamento e disseminação de dados coletados e informações processadas para um ou mais observadores (usuário final interessado pela informação) [12].

Em uma RSSF, cada nó é capaz de realizar as seguintes funções: sensoriamento do ambiente, processamento da informação, armazenamento de informação e comunicação com outros sensores através do canal de comunicação sem fio. A Figura 2.1 exemplifica as unidades funcionais de um nó sensor com base em suas funções.

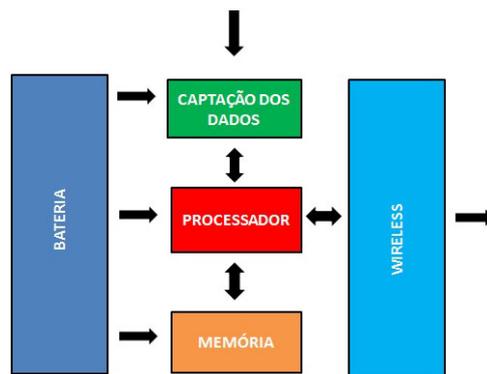


Figura 2.1: Um exemplo simplificado das unidades funcionais de um sensor sem fio.

A posição dos nós em uma rede de sensores não necessita ser predeterminada. Isso permite, por exemplo, a implantação aleatória em terrenos inacessíveis ou em operações de resgate. Por outro lado, isso também significa que os protocolos de rede de sensores e os algoritmos devem possuir capacidades de auto-organização.

A comunicação entre uma RSSF e outras redes ocorre por meio de nós ponto de acesso (*gateway*). O ponto de acesso pode ser implementado em um nó da rede que será conhecido como nó sorvedouro (*sink node*) ou em uma estação base (*base station*). As informações coletadas percorrem a RSSF até chegar a um ponto de acesso que irá encaminhá-las, por uma rede como a Internet, até a máquina onde estará a aplicação que utilizará as informações [11, 13].

Uma rede de sensores sem fio é um sistema dependente da aplicação. Qualquer projeto ou solução proposta que envolva o uso de RSSFs deve considerar os requisitos da aplicação a ser construída, as características e restrições dos componentes dos nós sensores, assim como as características do ambiente onde a rede será aplicada [12].

Uma grande característica dos sensores é a sua versatilidade, uma vez que podem ser empregados nas mais variadas áreas e funções. Por exemplo, na área Ambiental, pode-se usar sensores para monitorar variáveis ambientais em locais internos como prédios e residências, e locais externos como florestas, desertos, oceanos, vulcões, etc. Na área Militar pode-se usar sensores para detectar movimentos inimigos, explosões, a presença de gás venenoso ou radiação, etc. Por fim, na Saúde pode-se usar sensores para monitorar o funcionamento de um órgão específico, detectar a presença de substâncias que revelem a presença ou surgimento de alguma anomalia.

2.2 Características de uma Rede de Sensores

Redes de Sensores Sem Fio diferem das redes de computadores convencionais em diversos aspectos. Elas possuem um grande número de nós distribuídos, têm restrições de energia, devem possuir mecanismos para autoconfiguração e adaptação devido a problemas como falhas de comunicação e perda de nós [14]. Nesta seção, as principais características desta tecnologia serão detalhadas.

2.2.1 Auto-organização

Os nós de uma RSSF podem falhar por diferentes causas: más condições físicas, falta de energia, problemas de comunicação envolvendo o canal de comunicação sem fio. Além das falhas, uma rede de sensores pode apresentar problemas em situações onde um nó sensor, antes inativo, torna-se ativo e em ocasiões onde um novo nó é adicionado ao campo de sensores [14, 15].

Tendo em vista esses exemplos, é altamente desejável que a rede seja tolerante a falhas, isto é, é necessário que a rede possua mecanismos de auto-organização para que ela continue a executar a sua função, mesmo diante de um revés. Essa configuração deve ser automática e periódica já que a configuração manual não é viável devido a problemas de escalabilidade [14, 15].

2.2.2 Comunicação Sem Fio *Ad-Hoc*

Em várias situações, o ambiente a ser monitorado não possui uma infraestrutura necessária para suportar conexões cabeadas fim-a-fim (*end-to-end*). Dessa forma, os sensores devem usar os canais de comunicação sem fio (*wireless*) para solucionar esta adversidade. As redes sem fio podem ser divididas em duas distintas categorias. A primeira abrange as redes com **infraestrutura**, nas quais toda a comunicação é realizada através de um ponto de acesso. A segunda categoria engloba as redes **sem infraestrutura** (*ad-hoc*), nas quais as estações se comunicam diretamente, não existindo um ponto de acesso [16].

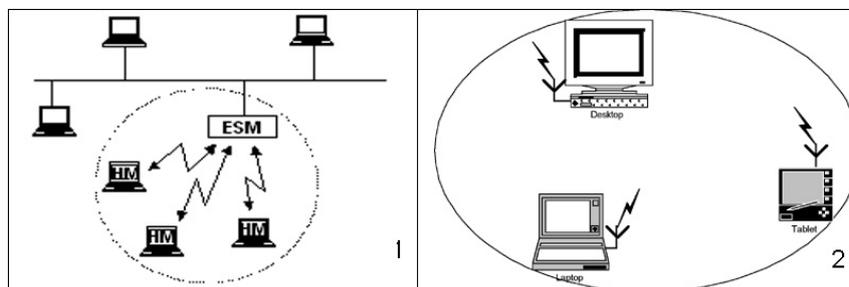


Figura 2.2: Redes com Infraestrutura x Redes Sem Infraestrutura (*Ad-Hoc*).

A Figura 2.2 apresenta a diferença entre as duas categorias de rede. A imagem número 1 ilustra uma Rede com Infraestrutura, na qual a comunicação entre os componentes da rede é realizada com um ponto de acesso (representado na imagem pela sigla ESM). Já a imagem número 2 ilustra uma Rede *Ad-Hoc* na qual os três componentes da rede (o *desktop*, o *notebook* e o *tablet*) se comunicam diretamente.

Para a implementação de muitas aplicações que envolvem RSSFs, é necessário o uso de técnicas do domínio de redes *ad-hoc*. Entretanto, embora muitos protocolos e algoritmos tenham sido propostos para as tradicionais redes *ad-hoc*, eles não são bem adequados para as características únicas e requisitos de aplicações das redes de sensores [9].

Em comparação com as redes *ad-hoc* tradicionais, as principais diferenças de uma rede de sensores são: o número de nós em uma rede de sensores pode ser muito maior do que os

nós de uma rede *ad-hoc* tradicional, os nós sensores são densamente implantados, os nós sensores são propensos a falhas, a topologia de uma rede de sensores muda frequentemente, os nós sensores utilizam principalmente o paradigma de comunicação *broadcast* enquanto a maioria das redes *ad-hoc* tradicional são baseadas na comunicação ponto-a-ponto, os nós sensores possuem limitações de potência e capacidades computacionais, e por fim, os nós sensores podem não possuir uma identificação global devido à grande quantidade de sobrecarga e o alto número de sensores [17].

2.2.3 Comunicação *Multi-hop*

A comunicação de múltiplos saltos (*multi-hop*) é uma forma de comunicação de dados presente nas RSSFs. Neste tipo de comunicação, caso um sensor (nó origem) precise transmitir os dados de sua memória para uma estação base ou outro sensor da rede (nó destino) fora de seu alcance de transmissão, ele transmitirá os dados para um sensor mais próximo que, caso o destino esteja fora de seu alcance, transmitirá a outro sensor até alcançar o destino. A transmissão de dados de um sensor para outro é conhecida como salto (*hop*). Logo, em uma rede de sensores *multi-hop*, os nós atuam em dois papéis: criador e roteador de dados [9].

A razão para o uso do roteamento *multi-hop* em uma rede de sensores se deve a limitação do alcance de transmissão dos próprios sensores. Por exemplo, suponha que um nó esteja distante da estação base e precise enviar os dados coletados. Como o seu alcance de transmissão não é o suficiente para se comunicar com a base, ele precisará enviar os dados para outros nós que estejam mais próximos da base. A Figura 2.3 ilustra o uso desta técnica em um campo de sensores.

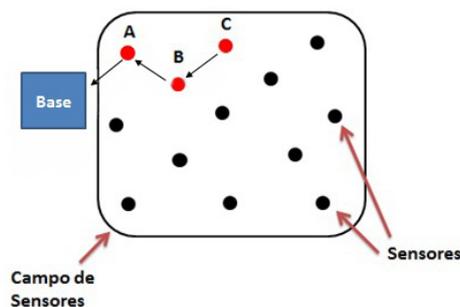


Figura 2.3: Um exemplo de uma rede de sensores utilizando a comunicação *multi-hop*.

A Figura 2.3 apresenta um campo de sensores com uma estação base responsável por coletar os dados produzidos pelos aparelhos. No exemplo, para transmitir os dados para a estação base que estava fora de seu alcance, os dados de C realizam um salto para o sensor B. Como a estação também encontra-se fora do alcance sensor B, os dados de B realizam um salto para o sensor A. Como a estação base encontra-se ao alcance de transmissão do sensor A, ele transmite os dados originários do sensor C para a base.

O método de comunicação *multi-hop* é uma das principais formas de economia no consumo de energia nas redes de sensores sem fio. Isso acontece uma vez que a energia

necessária para a comunicação entre dois nós arbitrários X e Y é dependente da distância entre os dois nós [11].

2.2.4 Grande Quantidade de Nós Distribuídos

A quantidade de nós de sensores implantados no ambiente pode ser da ordem de centenas ou milhares. Dependendo da aplicação, o número pode chegar a um valor extremo de milhões de sensores. Logo, escalabilidade é uma característica desejável, uma vez que é necessário uma habilidade para manipular uma porção crescente de trabalho de forma uniforme, ou estar preparado para crescer [9, 18].

2.2.5 Limitação Energética

Como os sensores utilizam energia provida de uma bateria embarcada, eles possuem uma quantidade limitada de recursos energéticos, os quais determinam seu tempo de vida. O tempo de vida de uma RSSF depende fortemente dos tempos de vida dos nós que constituem a rede. A vida útil de um nó sensor depende basicamente de dois fatores: quanto de energia ele consome ao longo do tempo e o quanto de energia está disponível para uso [19].

Segundo Akyildiz *et al.* [9], a quantidade predominante de energia de um nó sensor é consumida durante as atividades de detecção, comunicação e processamento de dados. Uma vez que é inviável parar o serviço para recarregar milhares de nós, cada nó deve ser tão energeticamente eficiente quanto possível. Ou seja, aplicações, protocolos, e algoritmos para redes sensores sem fio não podem ser escolhidos considerando apenas sua elegância e capacidade, mas, principalmente, pela quantidade de energia consumida [14].

Portanto, sabendo que uma RSSF só pode cumprir sua finalidade desde que ela esteja energeticamente ativa, a energia é o principal recurso e o grande gargalo desta tecnologia.

2.2.6 Esforço Cooperativo

Os sensores tendem a serem projetados com pequenas dimensões. Esta limitação de tamanho acaba impondo limitações nos recursos dos dispositivos, tais como capacidade da fonte de energia, processador e transceptor. Mas apesar dos nós possuírem individualmente pouca capacidade computacional e de energia, um esforço colaborativo entre os mesmos permite a realização de uma tarefa maior [12].

Devido às restrições das RSSFs, os dados tendem a ser fundidos/agregados ou sumarizados para melhorar o desempenho no processo de detecção de eventos. A **sumarização** é o conjunto de técnicas que visa a compactação de um determinado conjunto de dados em um conjunto de menor tamanho, mantendo o máximo de informação possível. Essas técnicas são dependentes da aplicação que está sendo executada [14, 20].

A **fusão de dados** é uma técnica na qual os nós sensores realizam o processamento

das informações recebidas de diferentes vizinhos a fim de explorar a correlação entre as diferentes informações e reduzir a quantidade de informação a ser transmitida [21].

2.3 Roteamento em uma Rede de Sensores

O processo de roteamento em uma RSSF consiste na escolha entre os vários caminhos possíveis para se enviar os dados coletados pelos sensores. O nó destino poderá ser o nó que processará os dados ou um ponto de acesso que enviará os dados para a aplicação em questão.

Nas Redes Tradicionais, o roteamento segue o modelo **Centrado em Endereços** (*Address-Centric*), onde o foco está na busca pelo menor caminho entre pares (origem e destino) de nós endereçáveis. Nesse modelo, cada nó envia seus dados ao nó sorvedouro de forma independente, ou seja, sem a necessidade de uma requisição do sorvedouro para que o nó envie os dados. Já nas Redes de Sensores, a abordagem mais usada é a **Centrada em Dados** (*Data-Centric*), onde a estação base envia consultas para regiões específicas da RSSF e aguarda pelos dados dos sensores localizados nas regiões selecionadas. O modelo ainda permite o processo sumarização, evitando assim a redundância dos dados [13, 22].

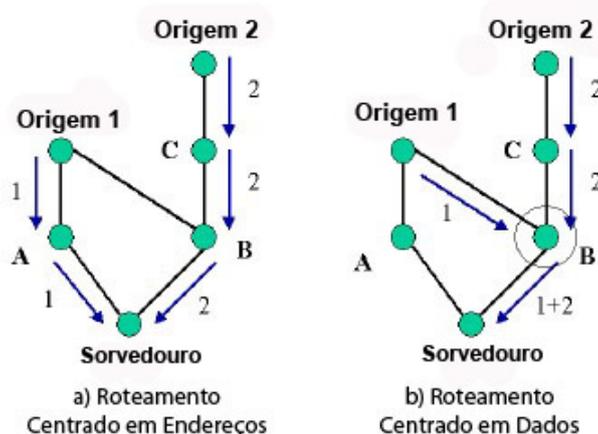


Figura 2.4: Roteamento Centrado em Endereços x Roteamento Centrado em Dados [23].

A Figura 2.4 exemplifica a diferença entre as duas abordagens. No modelo Centrado em Endereços, cada nó origem envia suas informações de forma separada para o nó sorvedouro (o nó origem 1 encaminha os dados rotulados como “1” através do nó A, e o nó origem 2 encaminha os dados rotulados como “2” através dos nós C e B). No modelo Centrado em Dados, os dados dos nós origem 1 e 2 são agregados pelo nó B, combinados (rotulados como “1 + 2”) e por fim enviados de B para o nó sorvedouro. A abordagem Centrada em Dados resulta em uma economia de energia com poucas transmissões para enviar as informações dos nós origem para o sorvedouro.

A comunicação centrada em dados define um novo paradigma de endereçamento, no qual os nós são endereçados pelos atributos dos dados que eles geram, um par (atributo,

valor), ou por sua localização geográfica, em vez de serem endereçados pela sua localização topológica na rede (como em endereços IP) [23].

Nas RSSFs, existem duas principais famílias de protocolos: Hierárquicos ou Planos.

2.3.1 Roteamento Hierárquico

No roteamento hierárquico, os nós sensores são organizados em grupos (*clusters*). Em cada grupo, existem dois tipos de nós: os nós fontes e o líder de grupo (*cluster head*). Os nós fontes simplesmente coletam e enviam os dados para o líder de seu grupo que pode executar uma fusão/agregação de dados antes de enviá-lo para o ponto de acesso. O ponto de acesso pode ser implementado em uma estação base (*base station*).

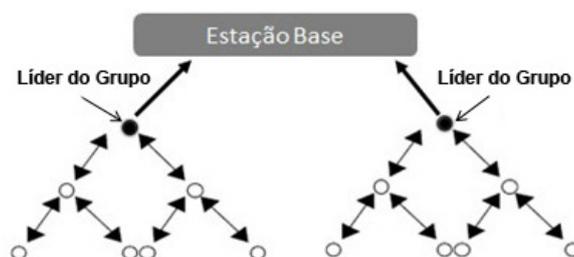


Figura 2.5: Roteamento Hierárquico de dois grupos de rede de sensores sem fio.

A Figura 2.5 ilustra dois grupos de *clusters* realizando o roteamento hierárquico. Cada nó dentro do *cluster* comunica-se apenas com o seu nó pai. Cada líder de grupo integra a informação dos nós fontes abaixo dele por meio de alguma técnica de agregação e reporta os resultados acima na hierarquia. No caso da Figura 2.5, os líderes de cada *cluster* enviam as informações para a estação base que representa o topo da hierarquia. Os problemas desta classe de protocolos estão na seleção dos nós líderes e na configuração da hierarquia de forma a garantir um balanceamento no gasto de energia dos nós [24].

Entre os principais algoritmos, pode-se destacar o LEACH (*Low-Energy Adaptive Clustering Hierarchy*) [25], o PEGASIS (*Power-Efficient Gathering in Sensor Information Systems*) e o TEEN (*Threshold sensitive Energy Efficient sensor Network*) [13].

2.3.2 Roteamento Plano

No Roteamento Plano (*Flat Routing*), todos os sensores são considerados iguais, ou seja, a atividade de roteamento é tratada de forma idêntica por todos os sensores da rede. Cada nó é autônomo e conectado a alguns ou a todos os outros nós, dependendo do alcance do seu rádio. Não existem *cluster heads* ou estações bases [24], logo os sorvedouros (*sink nodes*) serão os pontos de acesso, responsáveis pelo processamento da informação coletada pela rede e pelo envio dos dados para a aplicação.

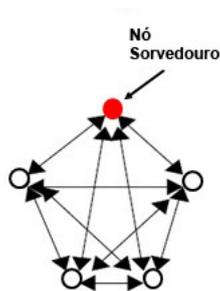


Figura 2.6: Roteamento Plano com um nó como sorvedouro.

A Figura 2.6 ilustra um exemplo de uma RSSF de roteamento plano. Quando um nó sensor dispõe de uma informação que deve ser transmitida, ele deve encontrar uma rota até o nó sorvedouro. Caso o sensor fonte seja vizinho do sensor sorvedouro, a conexão será direta, caso contrário, a rota será constituída de múltiplos saltos [22].

Entre os principais algoritmos, pode-se destacar o *Directed Diffusion* [15], o SPIN (*Sensor Protocols for Information via Negotiation*) [25] e o SAR (*Sequential Assignment Routing*).

2.4 Redes Oportunísticas

Apesar de providenciar eficientes metodologias para comunicação em RSSFs, os protocolos de Roteamento Hierárquico e Plano restringem-se apenas às redes nas quais os sensores permanecem estáticos no ambiente. Em uma rede na qual os nós sensores possuem mobilidade, determinar uma topologia é inviável e, logo, o emprego de protocolos das duas famílias de roteamento também.

A solução para isso fica por conta do uso de algoritmos de Redes Oportunísticas (*Opportunistic Networks*). Segundo Perusi *et al.* [26], uma Rede Oportunística é uma evolução das redes *ad-hoc multi-hop*. Já para Lilien *et al.* [27], elas podem ser compreendidas como redes que estão dentro de uma intersecção formada pelas Redes *Ad hoc*, Sistemas P2P (*Peer-to-peer*) e Redes de Sensores.

Nesse tipo de rede, nenhuma suposição é feita a respeito da existência de um caminho completo entre dois nós. Nós origem e destino podem nunca estar conectados com a mesma rede, ao mesmo tempo. Entretanto, as técnicas de rede oportunistas permitem tais nós trocarem mensagens entre si. Normalmente, isso vem ao preço de um atraso adicional na entrega de mensagens, uma vez que as mensagens são frequentemente armazenadas na rede à espera de um caminho em direção ao destino que esteja disponível [26].

2.4.1 Roteamento Oportunístico

O desempenho do roteamento melhora consideravelmente quando mais conhecimento sobre a topologia da rede pode ser explorada [28]. Infelizmente, em uma rede oportunística

esse tipo de informação não é facilmente disponível, visto que para consegui-lo, o desempenho muitas vezes precisa ser comprometido.

Para solucionar a comunicação entre os nós presente nesse paradigma, uma nova família de protocolos foi projetada. Esse conjunto de protocolos faz parte do Roteamento Oportunístico (*Opportunistic Routing*). A Figura 2.7 exhibe a taxonomia desta família de roteamento.



Figura 2.7: Taxonomia do Roteamento Oportunístico.

A primeira classificação refere-se quanto à presença ou não de infraestrutura na rede. No caso onde não existe infraestrutura, as abordagens podem ainda ser divididas em Roteamento Baseado em Disseminação (*Dissemination Based*) e Roteamento Baseado em Contexto (*Context Based*). No caso onde há infraestrutura, as abordagens podem ser divididas em Roteamento Baseado em Infraestrutura Fixa (*Fixed Infrastructure*) e Roteamento Baseado em Infraestrutura Móvel (*Mobile Infrastructure*).

2.4.1.1 Roteamento Baseado em Disseminação

As técnicas de roteamento baseado em disseminação realizam a entrega de uma mensagem para o destino através de sua disseminação através de toda a rede. A heurística por trás da ideia é que uma vez que não há conhecimento de um caminho possível para o destino, uma mensagem deve ser enviada em todos os lugares. Ele acabará por chegar ao destino já que está passando por todos os nós [26].

O *Epidemic* [29] e o MV (*Meetings and Visits*) [30] são alguns dos principais protocolos desta classe.

2.4.1.2 Roteamento Baseado em Contexto

Roteamento baseado em contexto explora mais informações sobre o contexto em que nós estão operando para identificar o modo adequado para os próximos saltos (*next hops*) em direção aos eventuais destinos. Por exemplo, o endereço residencial de um usuário é uma parte valiosa de informações de contexto para decidir o seu próximo salto [26].

O CAR (*Context-Aware Routing*) [31] e o *MobySpace* [32] são alguns dos principais protocolos desta classe.

2.4.1.3 Roteamento Baseado em Infraestrutura Fixa

Um nó origem que deseje enviar uma mensagem geralmente a mantém até que esta alcance a estação base pertencente à infraestrutura, em seguida, encaminha a mensagem para ele. Estações base são geralmente *gateways* que podem proporcionar acesso à Internet ou podem ser conectados a uma LAN. Assim, o objetivo de um algoritmo de roteamento oportunístico é entregar mensagens para os *gateways*, que é suposto ser capaz de encontrar o destino final mais facilmente [26].

Duas conhecidas variações do protocolo são o *Infostation* e o SWIM (*Shared Wireless Infostation Model*) [33].

2.4.1.4 Roteamento Baseado em Infraestrutura Móvel

Os nós da infraestrutura são coletores de dados móveis. Eles se movem em torno da área de rede, recolhendo as mensagens dos nós que passam por eles. Esses nós especiais são referidos como transportadores (*carriers*), suportes (*supports*), encaminhadores (*forwarders*), *Mobile Ubiquitous LAN Extensions (MULEs)*, ou mesmo *ferries* [26].

Eles podem ser as únicas entidades responsáveis pela entrega de mensagens quando apenas a comunicação nó para transportador (*node-to-carriers*) é permitida, ou podem simplesmente ajudar a aumentar a conectividade em redes esparsas, dessa forma garantindo também que os nós isolados possam ser alcançados. Neste último caso, a entrega de mensagens é realizado tanto por nós transportadores e nós ordinários, e ambos os tipos de comunicação de nó para nó e nó para transportador são permitidos [26].

Duas conhecidas variações do protocolo são o *Data-MULE system* [34] e o *Message Ferrying Approach* [35].

2.5 Considerações Finais

Neste capítulo, foi apresentada a fundamentação teórica a respeito de Sistemas de Sensores Sem Fio. Conceitos básicos sobre sensores e RSSFs foram definidos bem como suas principais características e o roteamento de informações.

É importante salientar que, em todos os trabalhos referenciados neste capítulo, a questão energética é colocada como o ponto chave dentro da área de sensores. Produzir uma fonte energética capaz de aumentar o tempo de vida de um sensor ainda é um problema longe de uma solução, tanto que até mesmo em trabalhos mais contemporâneos (como o de Agrawal [36]), a melhor forma de driblar essa questão continua sendo a elaboração de técnicas que economizem o uso da energia.

Capítulo 3

Trajetórias Semânticas

Neste capítulo será formalizado o conceito de Trajetórias Semânticas usado no projeto. Mas antes disso, o conceito básico da Mineração de Dados será abordado, uma vez que as técnicas de Semântica utilizam muitas ideias provindas desta sub-área da Inteligência Artificial.

3.1 Mineração de Dados

Os recentes avanços na tecnologia de monitoramento global e a redução de custo de dispositivos móveis, como aparelhos de GPS, telefones celulares e redes sensores, tem aumentado significativamente o uso desses equipamentos de monitoramento. Esses dispositivos registram a mobilidade dos indivíduos ou dos equipamentos que os carregam [6].

O aumento do uso desses dispositivos ocasionou também no aumento do armazenamento de dados espaço-temporais produzidos por essas ferramentas. Na forma em que são coletados, esse conjunto de dados não possuem muita utilidade. Entretanto, aplicando-se um conjunto de técnicas especiais, é possível extrair conhecimento relevante e interessante.

Importantes publicações em tecnologia, como os relatórios da *Gartner Group*, apontam a Mineração de Dados (*Data Mining*) como uma das tecnologias mais promissoras para o futuro [37]. Sendo um termo cunhado na década de 90, a Mineração de Dados é uma tecnologia que emergiu da intersecção entre as áreas da Estatística Clássica, Inteligência Artificial e do Aprendizado de Máquina.

A literatura fornece várias definições para Mineração de Dados. Segundo Elmasri e Navathe [37], refere-se à descoberta de novas informações em função de padrões ou regras aplicadas em grandes quantidades de dados. De acordo com Sferra e Correa [38], pode ser entendido como o processo de extração de informações, sem conhecimento prévio, de um grande banco de dados e seu uso para tomada de decisões. Ela ainda define o processo automatizado de captura e análise de grandes conjuntos de dados para extrair um significado, sendo usado tanto para descrever características do passado como para prever tendências para o futuro.

Por fim, para Weis e Indurkha [39] a Mineração de dados é a busca de informações valiosas em grandes bancos de dados. É um esforço de cooperação entre homens e computadores. Os homens projetam bancos de dados, descrevem problemas e definem seus objetivos. Os computadores verificam dados e procuram padrões que casem com as metas estabelecidas pelos homens.

A técnica consiste também no passo fundamental em um processo conhecido como KDD (*Knowledge Discovery in Databases*) – em português, Descoberta de Conhecimento em Bancos de Dados [38], o qual será abordado na seção a seguir.

3.1.1 Processo de Descoberta de Conhecimento

É o processo global de transformar um conjunto de dados (baixo nível) em conhecimento (alto nível). De acordo com Fayyad *et al.* [40], o KDD é um processo não trivial de identificação de padrões válidos, potencialmente úteis e compreensíveis em um conjunto de dados. Ele preocupa-se com o desenvolvimento de métodos e técnicas que visam conferir sentido aos dados.

O KDD é composto por cinco etapas: seleção de dados, pré-processamento, transformação, mineração de dados e interpretação/avaliação. As cinco etapas são executadas de forma interativa e iterativa. Elas são interativas porque envolvem a cooperação da pessoa responsável pela análise de dados, cujo conhecimento sobre o domínio orientará a execução do processo. Por sua vez, a iteração deve-se ao fato de que, com frequência, esse processo não é executado de forma sequencial, mas envolve repetidas seleções de parâmetros e conjunto de dados, aplicações das técnicas de *Data Mining* e posterior análise dos resultados obtidos, a fim de refinar os conhecimentos extraídos [37, 38, 41]. A Figura 3.1 ilustra todo processo de descoberta de conhecimento.

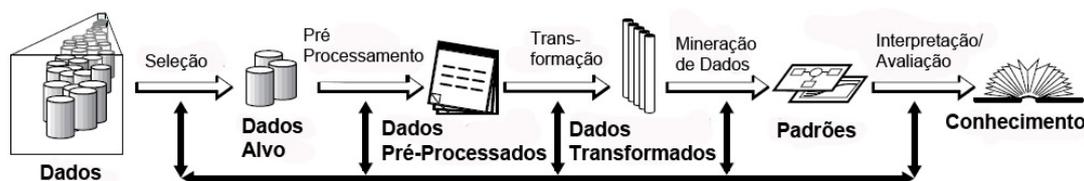


Figura 3.1: Fases do *Knowledge Discovery in Databases* (KDD) [38].

Seguindo as fases criadas por Fayyad *et al.* [40], as três primeiras etapas correspondem a fase de preparação dos dados nas quais problemas como ruídos, inconsistências e incompletude são tratados. Essa fase pode levar até 80% do tempo necessário para a realização de todo o processo, e visa sobretudo preparar os dados para que a próxima etapa possa ser realizada de forma efetiva [42].

O processo inicia com a seleção dos dados, na qual o entendimento do domínio da aplicação e dos objetivos a serem atingidos ao final dos procedimentos são escolhidos. Feito os procedimentos iniciais, os dados relevantes são selecionados e organizados na forma de agrupamentos.

Em seguida é a vez da etapa de pré-processamento, onde é realizada uma limpeza nos dados visando eliminar as informações consideradas desnecessárias. Estratégias para manusear dados faltantes ou inconsistentes são utilizadas, pois caso os erros não sejam encontrados e corrigidos nesta etapa, a qualidade do resultado final pode ser comprometida.

Com os dados pré-processados, chega o momento de executar a etapa de transformação, onde é realizada uma busca por características úteis para representar os dados dependendo do objetivo da tarefa, visando assim a redução no número de variáveis bem como o enriquecimento semântico das informações. Todas as etapas da fase de preparação de dados são ilustradas na Figura 3.2.

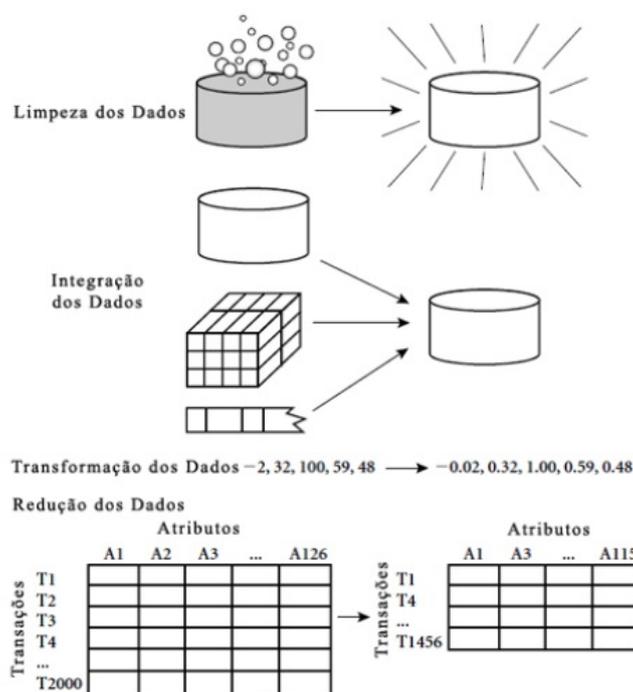


Figura 3.2: Fase de Preparação dos Dados [43].

Com o término da preparação dos dados, o próximo passo é a Mineração de Dados. É a etapa mais importante do processo, sendo considerada como o cerne do processo de descoberta de conhecimento. No entanto, a etapa geralmente leva apenas uma pequena parte (estimada em 15% a 25%) do tempo de processo [42].

A última fase do processo consiste no pós-processamento dos dados. Nesta fase acontece a etapa de interpretação e avaliação dos padrões descobertos e a possibilidade de retorno a qualquer uma das etapas anteriores. Dessa forma, a informação obtida é analisada em relação ao objetivo proposto no início do processo, sendo identificadas e apresentadas as melhores informações. Assim, o grande propósito do resultado não consiste apenas em visualizar o rendimento da Mineração, mas, também, em filtrar a informação que será apresentada, eliminando possíveis padrões redundantes ou irrelevantes que podem aparecer [38].

3.1.2 Mineração de Dados Espaço-Temporais

A elaboração de técnicas para a coleta de informações georreferenciáveis acabou resultando no armazenamento de uma grande quantidade de dados espaço-temporais. No entanto, devido à complexidade dos dados e de seus relacionamentos, o uso de técnicas convencionais de Mineração tornou-se inviável. Logo, novas técnicas foram elaboradas para lidar com esse tipo de dados, criando uma extensão do conceito da Mineração Padrão, conhecida como Mineração de Dados Espaço-Temporais [44, 45].

Mineração de Dados Espaço-Temporal (*Spatial-Temporal Data Mining*) refere-se à extração de conhecimento implícito, relacionamento espacial e temporal ou outros padrões de dados espaço-temporal. Ele envolve a extração e análise de informação útil armazenada em bancos de dados espaço temporais. Como o próprio nome diz, o conceito integra os conceitos espaciais e os temporais em um só [46].

O campo da Mineração de Dados Espacial é onde o aspecto espacial dos dados define uma relação entre cada objeto espacial. Segundo Ester *et al.* [47], há três tipos básicos de relações espaciais, onde todos são consideradas binárias entre pares de objetos:

- **Relações Topológicas:** uma relação entre dois objetos A e B são derivados das intersecções de seus interiores, dos seus limites e de seus complementos. As relações são: A disjuntos B, A satisfaz B, A sobrepõe B, A é igual a B, A inclui B, A coberto por B, A contém B, A dentro B.
- **Relações de Distância:** são aquelas relações que comparam a distância de dois objetos com uma determinada constante utilizando um dos operadores aritméticos ($<$, $>$ e $=$). Em outras palavras, em uma relação de distância $A \text{ distance}_{\sigma c} B$, $A \in B$ se $dist(O_1, O_2) \sigma c$, onde c é a constante, σ é um operador aritmético, $dist$ é a função de distância e O_1, O_2 são objetos espaciais.
- **Relações de Direção:** lidam com a ordem no espaço. Pode-se definir a relação entre dois objetos espaciais através de um ponto representativo $rep(O_1)$ do objeto origem O_1 e todos os pontos do objeto destino O_2 . O ponto representativo de um objeto espacial de origem pode, por exemplo, ser o centro do objeto. Este ponto representativo é utilizado como a origem do sistema de coordenadas virtual e seus quadrantes definem as direções.

Para melhor compreensão do conceito, a Figura 3.3 apresenta as três Relações Espaciais descritas anteriormente. Os exemplos presentes no conjunto 1 ilustram as Relações Topológicas, os presentes no conjunto 2 ilustram as Relações de Distância e os exemplos contidas no conjunto 3 ilustram as Relações de Direção.

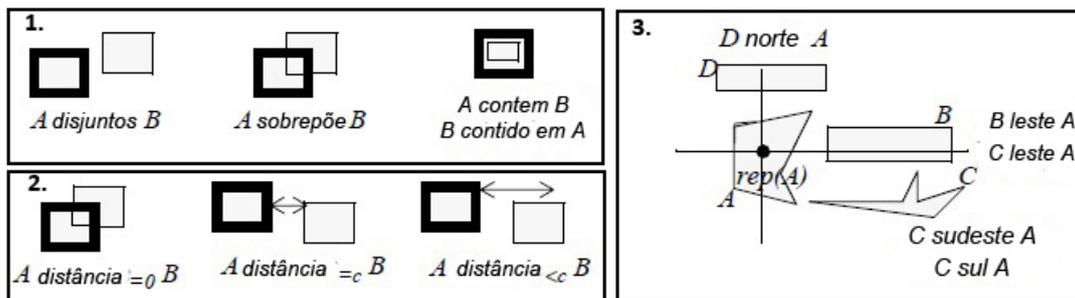


Figura 3.3: Todos os tipos relações espaciais ilustrados [47].

Já o campo da Mineração Temporal é onde o aspecto temporal dos dados define uma relação entre cada ponto de dados. Técnicas de *Temporal Data Mining* extraem as relações ou padrões de conjuntos de dados históricos, colocando maior ênfase no elemento temporal dos dados. Como exemplos de relações temporais, vamos considerar dois objetos temporais A e B onde podemos atribuir as seguintes relações: A antes B , B depois A , A durante B , entre outros [46].

3.2 Trajetórias

Como já abordado no início do capítulo, os avanços na tecnologia de satélites e dispositivos móveis, facilitaram a coleta de grandes quantidades de dados espaço-temporais. Ao dispor esses dados - normalmente definidos com um objeto dotado de coordenadas espaciais (x,y) e temporal (t) - em uma sequência temporal crescente, o resultado era uma trajetória.

Segundo Spaccapietra *et al.* [48], uma trajetória consiste em um registro de evolução do posicionamento de um objeto que está se movendo no espaço durante um determinado intervalo de tempo a fim de alcançar um dado objetivo. O movimento de um objeto qualquer não é necessariamente classificado como uma trajetória. Para isso, o objeto precisa deslocar-se de um ponto inicial a um ponto final com um objetivo definido. Também é preciso que os horários de início e fim da trajetória sejam demarcados, de modo a delimitar o trecho de interesse. Em outras palavras, uma trajetória é o segmento espaço-temporal do caminho percorrido por um objeto.

Dados espaço-temporais coletados pelos dispositivos eletrônicos são conhecidos como **trajetórias brutas** (*raw trajectory*). O formato de uma trajetória bruta resume-se a uma seqüência de pontos de amostra, definidos em termos de coordenadas (x, y, t) , que definem o comportamento de uma entidade durante certo intervalo de tempo. As enormes quantidades de dados brutos geralmente escondem conhecimento novo e interessante, que pode ser muito útil em processo de tomadas de decisão nas mais diversas áreas. Sistemas Gerenciadores de Bancos de Dados (SGBD) convencionais e Sistemas de Informação (SI) tradicionais, em geral, não oferecem suporte para a análise e a mineração de tais dados. Consultas sobre dados brutos são custosas e complexas. A mineração de trajetória brutas busca descobrir padrões geométricos, mas é totalmente ineficiente em descobrir padrões semânticos [6, 45].

Logo, houve uma necessidade de uma elaboração de novas técnicas, métodos e algoritmos de forma a interpretar esses dados e assim extrair todo conhecimento útil e interessante pelo qual esse conjunto de dados espaço-temporal apresenta.

3.2.1 Trajetória Semântica

O grande pioneiro na pesquisa na análise de comportamento das trajetórias brutas (objetos espaço-temporais) foi o trabalho de Laube *et al.* [49] que serviu como molde para várias pesquisas da área. Nele define-se um padrão para trajetórias que possuem comportamento similar, avaliando a mudança de direção. Um padrão deve conter um número mínimo de trajetórias que se movimentam na mesma direção. Entretanto, os aspectos temporais não são considerados.

Anos mais tarde, Laube e Imfeld [50] propuseram um novo tipo de padrão de trajetórias, analisando a direção do movimento e a região onde o determinado movimento ocorreu. Foram definidos quatro tipos de padrões de trajetórias: *convergence*, *encounter*, *flock* e *leadership*. Os padrões são definidos logo abaixo e ilustrados na Figura 3.4.

- **Convergence:** caracteriza um grupo de objetos que se movem em direção ao mesmo local.
- **Flock:** caracteriza objetos que se movem na mesma direção e suas trajetórias estão próximas umas das outras.
- **Encounter:** caracteriza um grupo de objetos móveis que se deslocam para o mesmo local ao mesmo tempo.
- **Leadership:** caracteriza um conjunto de objetos que se movem na mesma direção, suas trajetórias estão próximas umas das outras e um dos objetos lidera o grupo, estando à frente dos demais.

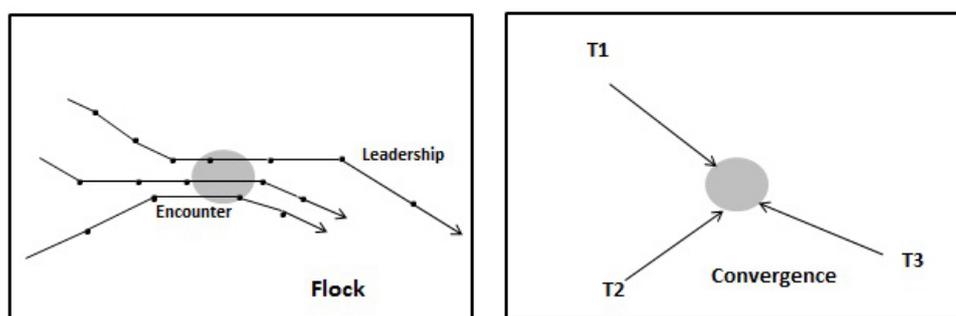


Figura 3.4: Exemplos de padrões de comportamento em grupos de trajetórias.

As pesquisas sobre Trajetórias de Objetos Móveis, até o ano de 2007, consideravam essencialmente o aspecto geométrico da trajetória. Nesse mesmo ano, o conceito de Trajetória Semântica foi originalmente introduzido por Alvares *et al.* [51]. Segundo o pesquisador, uma trajetória semântica consiste em adicionar uma contextualização/significado

às trajetórias brutas, a fim de facilitar e otimizar consultas espaço-temporais. Com base no modelo conceitual de *Stops* e *Moves* proposto por Spaccapietra *et al.* [48], Alvares introduziu o algoritmo **SMoT** (*Stops and Moves of Trajectories*).

Uma Trajetória bruta pode ser enriquecida com diferentes informações semânticas, de acordo com o contexto da aplicação e o problema que o usuário pretende resolver. Um grande exemplo disso pode ser visto na Figura 3.5, baseada em uma imagem do trabalho de Silva [52].

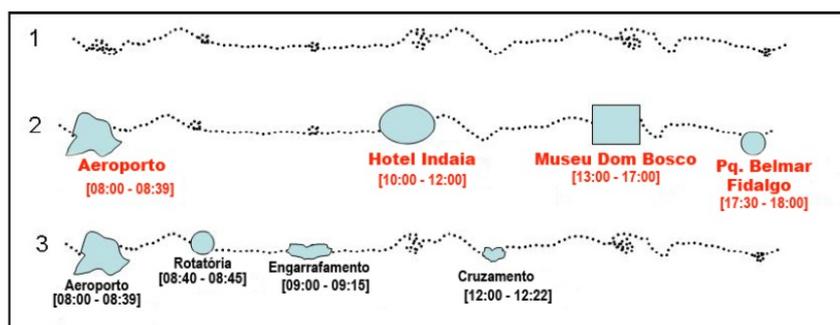


Figura 3.5: Exemplos de Trajetórias sem e com semântica.

A trajetória 1 é apenas um conjunto de vários pontos que compõem uma trajetória no estado bruto. Na trajetória 2, a semântica se insere no contexto de uma aplicação de turismo, na qual os aspectos importantes são locais como aeroportos, hotéis e pontos turísticos. Já trajetória 3, a semântica está relacionada a uma aplicação de trânsito. Logo, um conjunto de dados brutos, após a inserção de semântica, podem ser utilizados em diferentes domínios de aplicação [52].

3.2.2 Algoritmos de Trajetórias Semânticas

Nesta seção será apresentado os principais algoritmos da área de Trajetórias Semânticas utilizados no projeto de pesquisa. Uma parte dos algoritmos empregados utilizam como base o conceito de *Moves* e *Stops*, que podem ser entendidos como:

- **Stops:** são as partes de uma trajetória que interceptam um objeto geográfico de interesse da aplicação por um tempo mínimo.
- **Moves:** são as demais partes da trajetória.

Os principais algoritmos utilizados para a extração de *Stops* e *Moves* de um conjunto de trajetórias foram o **SMoT** ou **IB-SMoT** (*Intersection-Based Stops and Moves of Trajectories*) e o **CB-SMoT** (*Clustering-Based Stops and Moves of Trajectories*). Outra metodologia implementada no projeto foi o **Padrão Avoidance**, um algoritmo de análise comportamental para trajetórias.

3.2.2.1 SMOt ou IB-SMOt

O SMOt ou IB-SMOt [51] possui o princípio de adicionar informação geográfica relevante para a aplicação que está sendo considerada. O usuário deve fornecer informações sobre os tipos de objetos geográficos de interesse para a aplicação. Por exemplo, se a aplicação é na área de varejo, o usuário deve informar onde estão localizados os caixas, seção de limpeza, hortifruti, açougue entre outros. Esses tipos de objetos fornecidos são denominados candidatos a parada (*candidate stops*). O usuário também precisa informar o tempo mínimo de permanência da trajetória em cada tipo de objeto geográfico para que seja considerado um *stop*. Por exemplo, no contexto de uma aplicação turística, uma parada de 30 minutos no hotel pode ser entendida como irrelevante enquanto o mesmo período em um restaurante não.

O algoritmo funciona da seguinte maneira: para cada trajetória existente no conjunto de dados, verifica-se a sua intersecção com os objetos geográficos relevantes para a aplicação (candidatos a *stop*). Para os candidatos a *stop* com intersecção com a trajetória, verifica-se o período de tempo dessa intersecção com a trajetória. Se a duração for maior que o tempo mínimo estabelecido para aquele tipo de objeto geográfico, o objeto será considerado um *stop*. Se for inferior, ele não será considerado um *stop*. Em outras palavras, a principal idéia do SMOt é juntar um grupo de trajetórias com um grupo de objetos geográficos fornecido para o usuário, tendo como resultado a intersecção válida de ambos.

Baseada em uma figura do trabalho de Alvares *et al.* [51], a Figura 3.6 mostra um conjunto de trajetórias antes e depois da aplicação do algoritmo. No lado esquerdo, um conjunto de trajetórias é representado na forma de pontos de amostragem, sem qualquer semântica empregada. No outro lado, há informação geográfica é integrada para trajetórias. Enquanto o conjunto de pontos do lado esquerdo não possui padrões a serem visualizados, no lado direito há trajetórias com semântica adicionada.

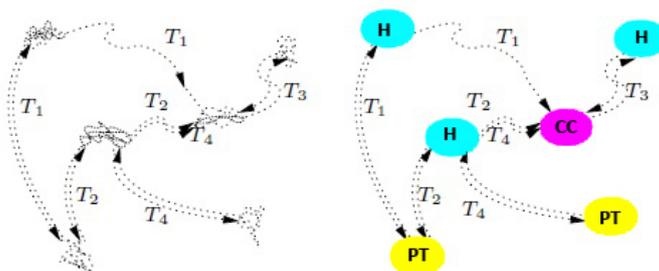


Figura 3.6: Esquerda: um conjunto de pontos sem semântica. Direita: o mesmo conjunto após cruzar com os dados geográficos fornecidos.

Três padrões semânticos puderam ser inferidos ao longo das quatro trajetórias presentes na Figura 3.6. (1) Três objetos móveis se moveram do Hotel (H) para um Ponto Turístico (PT); Três objetos apresentaram trajetórias retornando do Ponto Turístico para o Hotel de onde eles foram para um Centro de Conferência; e (3) todos os objetos móveis moveram do Hotel para o mesmo Centro de Conferência.

3.2.2.2 CB-SMoT

O CB-SMoT [53] é um algoritmo onde a principal idéia é encontrar *Stops* desconhecidos (*Unknown Stops*), isto é, possíveis *stops* que não estão presentes entre os objetos geográficos fornecidos pelo usuário. O procedimento é realizado em duas fases distintas. A primeira foca-se em identificar os trechos de maior densidade dos pontos tendo como base as velocidades de cada ponto. Se um conjunto de pontos apresentar uma velocidade média inferior da estipulada pelo usuário, esse conjunto será identificado como uma região de densidade.

Essa primeira fase utiliza uma ideia semelhante ao algoritmo de Agrupamento (*Clustering*) DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [54]. Enquanto no CB-SMoT um conjunto de densidade é formado por pontos de baixa velocidade, no DBSCAN um conjunto de densidade é uma região em que todos os pontos têm pelo menos um número mínimo de pontos em um raio ao seu redor. O número mínimo e o raio são argumentos fornecidos pelo usuário. O DBSCAN ainda apresenta uma propriedade de tratamento de ruídos (*noises*), isto é, a eliminação dos pontos que não se encaixam nos argumentos recebidos pelo algoritmo.

Com todas as regiões identificadas, o segundo passo é aplicar o IB-SMoT sobre a trajetória e os conjuntos de densidades. Os conjuntos de densidade e os pontos da trajetória que possuem intersecção com os candidatos a *stop* (além de respeitar o limite de duração), serão considerados *stops*. Caso ao final haja conjuntos de densidade que não se transformaram em *stops*, isso significa que eles são um ponto de parada da trajetória que é desconhecido ao usuário, recebendo a alcunha de *Unknown Stop*.

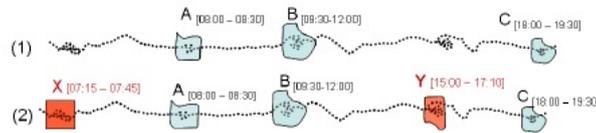


Figura 3.7: (1) é resultado do algoritmo IB-SMoT e a (2) é resultado do CB-SMoT [53].

A Figura 3.7 mostra o resultado final de uma trajetória bruta ao se aplicar o IB-SMoT e o CB-SMoT. Na Trajetória 1, o algoritmo IB-SMoT gerou um conjunto de *Stops* (A,B,C) formado por pontos da trajetória que tiveram intersecção com os candidatos à *Stop*. Na Trajetória 2, além de gerar os *Stops* (A,B,C), o CB-SMoT gerou o conjunto de *Unknown Stops* (X,Y), formado por trechos da trajetória onde a densidade dos pontos era elevada. Logo, graças a execução do CB-SMoT, foi possível descobrir duas novas áreas de parada (*Stops*) que não se encontravam presentes no conjunto de candidatos a parada.

3.2.2.3 Padrão *Avoidance*

Proposto inicialmente no trabalho de Loy [55], o padrão comportamental *Avoidance* ou Desvio caracteriza um comportamento em que um objeto móvel evita determinados objetos ou regiões espaciais. Esse tipo de padrão ocorre, por exemplo, quando pessoas evitam

transitar próximo a alguma região enquanto caminham em um parque.

A Figura 3.8 exemplifica um cenário onde há existência deste padrão. Seja as trajetórias t_1 , t_2 , e t_3 pertencentes a pessoas passeando em um parque. E seja a área cinza uma região cheia de poças de água. Enquanto o objeto móvel da trajetória t_2 atravessa a região sem exitar, os objetos de t_1 e t_3 mudam de direção, evitando desta forma a área alagada.

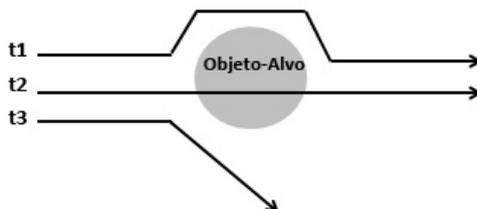


Figura 3.8: Exemplo do Padrão *Avoidance*.

Tendo em vista o exemplo acima, a primeira condição para existência de desvio é quando a trajetória de um objeto deve ir em direção ao objeto alvo e mudar de direção de forma a evitá-lo, sem que ocorra uma intersecção da trajetória com esse objeto-alvo. É importante salientar que deve haver uma distância máxima onde a mudança de direção ocorre. No exemplo do parque, uma pessoa que muda de direção à vários metros de distância da região alagada certamente não mudou seu caminho a fim de evitar cruzar as poças de água. Logo, define-se uma região específica onde a procura pelo padrão será realizada. Tal região recebe o nome de região de interesse e é formada por um *buffer* (no formato de uma circunferência) de tamanho de raio d que envolve a região alvo [6]. A Figura 3.9 mostra um exemplo desta ideia.

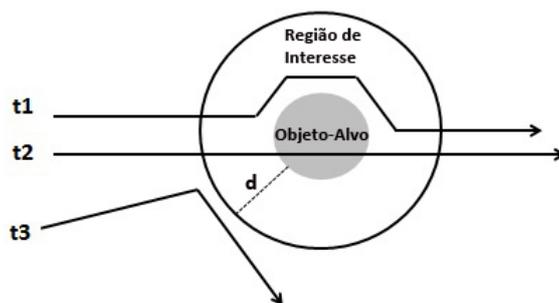


Figura 3.9: Um exemplo com uma região de interesse sobre objeto-alvo.

A Figura 3.9 mostra uma área de interesse de raio d . Se antes, eram considerados as trajetórias t_1 e t_3 como as que apresentavam desvios, agora, t_3 não se encontra dentro da área de interesse e t_1 é a única com presença de desvio. Portanto, a condição de *avoidance* agora deve considerar que uma trajetória deve ir em direção ao objeto-alvo e mudar de direção de forma a evitá-lo, dentro da região de interesse. Quanto maior a área de interesse, mais trajetórias poderão ser consideradas pela condição.

Para tornar mais robusta a identificação do padrão e menos susceptível a eventuais erros nos dados da trajetória, definiu-se que o objeto móvel deve se deslocar por um percurso mínimo dentro da região de interesse em direção ao objeto-alvo antes de mudar de direção, para só então se caracterizar um *avoidance*. O tamanho mínimo dessa porção da trajetória, denominada subtrajetória direcionada ao alvo, cujo símbolo é l , depende do contexto da aplicação e do tipo de dado de trajetória a ser analisado [6]. A Figura 3.10 exemplifica o conceito de subtrajetórias direcionadas ao alvo.

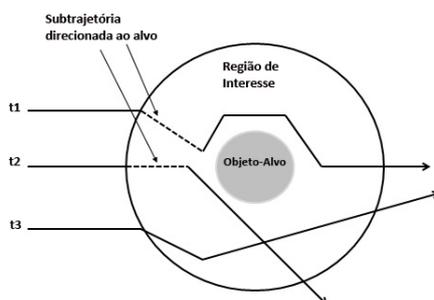


Figura 3.10: Exemplo de subtrajetórias direcionadas ao alvo.

Na Figura 3.10, todas as trajetórias interceptam a região de interesse, mas apenas a trajetória $t1$ e $t2$ apresentam uma subtrajetória (representadas pela linha tracejada) de tamanho maior ou igual l e direcionada ao objeto alvo. A trajetória $t3$ até apresenta uma subtrajetória equivalente, mas nenhuma que seja direcionada à região em questão. Se a região de interesse for grande o suficiente, podem ocorrer situações em que o objeto móvel se desloca mais de uma vez em direção ao objeto-alvo antes de evitá-lo definitivamente. A subtrajetória direcionada ao alvo deve ser a maior porção da trajetória, dentro da região de interesse, que vai na direção do alvo [55].

Por fim, ao se observar algumas trajetórias, é possível notar que alguns objetos móveis retomam o seu caminho original após evitar a região alvo. O fato de uma trajetória retomar o seu caminho mostra que o desvio foi totalmente intencional. Logo, a fim de determinar o grau de *avoidance* realizado pela trajetória, um teste de intersecção foi elaborado utilizando uma operação de intersecção entre a trajetória e uma área denominada região de incremento de confiança [6]. A Figura 3.11 exemplifica a região em questão.

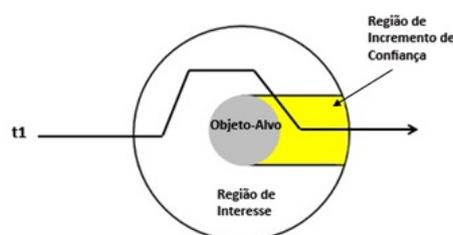


Figura 3.11: Exemplo de uma trajetória que passa pela região de incremento de confiança.

Uma região de incremento de confiança é uma área contida em região de interesse, que não inclui a área do objeto-alvo. Ela está localizada entre o objeto-alvo e o limite

da região de interesse, ao lado oposto ao primeiro ponto da subtrajetória direcionada ao alvo. A largura da região é delimitada por duas retas paralelas e tangentes ao objeto-alvo, com uma inclinação dada pelo ângulo formado por uma reta entre centro geométrico do objeto-alvo e primeiro ponto da subtrajetória direcionada ao alvo [55].

Quando uma trajetória, ao realizar o desvio, realiza uma intersecção na área de incremento de confiança, o grau de *avoidance* realizado pela trajetória pode ser considerado como forte, uma vez que ele evitou a região e retomou seu caminho. Caso ele desvie do alvo mas não retome o caminho, o grau de *avoidance* é considerado fraco.

3.3 Considerações Finais

Neste capítulo foi apresentada a contextualização teórica de Trajetórias Semânticas, uma subárea da Inteligência Artificial. Além de mostrar conceitos básicos sobre trajetórias e alguns de seus principais algoritmos, o capítulo apresentou informações a respeito de Mineração de Dados e Mineração de Dados Espaço Temporais.

Trajetoária Semântica é um conceito de extrema importância para o projeto, uma vez que permitirá a extração de conhecimento relevante de dados espaço temporais coletados do rebanho. Nos capítulos posteriores, será descrito como cada algoritmo foi implementado e utilizado no desenvolvimento da ferramenta.

Capítulo 4

Trabalhos Correlatos

Esta pesquisa, na área de Pecuária de Precisão, foca na construção de um sistema computacional que seja capaz de identificar padrões comportamentais e informações relevantes da interação dos animais entre si e com o ambiente. Para atingir os objetivos deste projeto, diversos trabalhos correlatos em várias áreas de pesquisa foram explorados, em especial aqueles que tratam de sistemas computacionais elaborados para área agropecuária e textos que tratam sobre monitoramento de animais e do ambiente.

Neste capítulo são apresentados os principais trabalhos que motivaram e possibilitaram este estudo. Além disso, são apresentados alguns dos trabalhos na área de Pecuária de Precisão realizados entre a Embrapa - Gado de Corte e a FACOM/UFMS.

4.1 Monitoramento Animal

Como visto no primeiro capítulo, a Pecuária de Precisão é responsável pelo fornecimento de tecnologias que monitoram o animal no seu ambiente, sendo uma forma eficiente de gerenciar a produção animal, bem como conhecer mais a respeito do comportamento do animal inserido naquela região.

O projeto ZebraNet [56], desenvolvido em parceria pela Universidade de Princeton (Estados Unidos) e pelo Centro de Pesquisa de Mpala (Quênia), foi um dos grandes precursores do monitoramento animal. O sistema foi criado com o intuito de rastrear as migrações de Zebras no continente Africano. O estudo fez uso de colares com sensor GPS embutido, colocados no pescoço dos animais (Figura 4.1).

Na campo biológico, a pesquisa permitiu aos pesquisadores conhecer questões importantes referentes ao processo migratório das zebras, suas relações interespecíes e o seu comportamento noturno. No campo tecnológico, incentivou o desenvolvimento de dispositivos eletrônicos capazes de realizar o monitoramento de animais em um ambiente selvagem e sem infraestrutura.

A semelhança entre o projeto e esta proposta de estudo reside no uso de colares com sensores GPS embutidos para o monitoramento de um certo grupo de animais, a



Figura 4.1: Um colar do programa Zebranet sendo colocado em uma zebra.

fim de obter informações a respeito da relação dos animais com o ambiente. A grande diferença entre ambos fica na questão do ambiente: enquanto o ZebraNet foi testado em um ambiente aberto (Savana Africana), este projeto será aplicado em uma pastagem cercada.

O sistema consiste em colares especiais que são transportados por zebras para a coleta de dados, uma estação base que, periodicamente, se move em torno do cerrado para coletar informações dos colares, e um protocolo de rede para enviar os dados dos colares para a estação base móvel. Os colares formavam uma rede *ad-hoc* e se comunicavam entre eles através do método *peer-to-peer*. A estação base era um veículo móvel, guiado pelos pesquisadores.

Outra pesquisa a contribuir com este trabalho foi desenvolvida por Handcock *et al.* [57]. Eles realizaram um estudo de monitoramento bovino na *Belmont Research Station*, a 26 Km da cidade de Rockhampton, na região nordeste da Austrália. Os pesquisadores australianos, assim como propõe este projeto, utilizaram colares com sensores GPS embutido em bovinos (Figura 4.2), confinados em um ambiente controlado. A principal diferença fica por conta do uso de imagens via satélite no monitoramento do rebanho. Combinando o sensoriamento terrestre (feito pelos colares) com o sensoriamento remoto (realizado pelos satélites), os pesquisadores obtiveram informações ainda mais precisas a respeito do posicionamento geográfico dos animais.



Figura 4.2: Bovinos com os colares no experimento do *Belmont Research Station* [57].

No Brasil, também existem pesquisas do gênero como a realizada por Visoli *et al.* [58]. O projeto apresentou uma solução para a identificação de ocorrência de contatos entre animais, a partir do uso de trajetórias vindas de dados georreferenciados de movimentação animal. A base de dados de contatos entre animais viabiliza a identificação de possíveis animais que contaminaram ou podem ser contaminados a partir de uma determinada ocorrência da febre aftosa, uma das doenças que podem atingir os animais e prejudicar a produção de carne bovina.

Utilizando equipamentos elaborados pelo projeto *Operational Management and Geodecisional Prototype to Track and Trace Agricultural Production* (OTAG), a pesquisa coletou dados de posicionamento referentes à movimentação dos animais utilizados no experimento. O contato entre animais era observado através do cálculo de intersecção entre duas trajetórias num determinado período de tempo.

Ao final do estudo, embora outras fontes de informação sejam necessárias para um procedimento mais preciso, como por exemplo análise da existência de rios e florestas, potenciais disseminadores de doença, ou barreiras naturais, o uso de trajetórias se demonstrou como uma alternativa para a organização de uma base de informações sobre contatos entre animais. Essa base de dados pode ser consultada quando há suspeita ou comprovação de uma doença como a febre aftosa, e rapidamente responder quais animais entraram em contato em prazos determinados como possíveis de contaminação, e onde se encontram esses animais, para uma eventual ação sanitária.

4.2 Mineração de Dados na Agropecuária

Com todos os dados adquiridos pelos sensores, é preciso realizar um estudo sobre esse conjunto de dados de modo a extrair informações e padrões ocultos nessas coleções, algo que as técnicas de Mineração de Dados são capazes de fazer.

O uso de metodologias de *Data Mining* aliadas ao monitoramento animal foi abordado de maneira interessante por Nadimi *et al.* [59]. Realizado em uma fazenda com um rebanho composto por caprinos, o estudo disponibilizou colares equipados com acelerômetros para coletar em determinado intervalo de tempo a angulação das cabeças dos animais. Utilizando um sistema de Rede Neural, através dos valores coletados, o sistema era capaz de prever quando o animal estava pastando, deitado, caminhando, em pé entre outros.

Outra aplicação da mineração de dados no meio rural pode ser encontrada na pesquisa de Moura *et al.* [60]. Para uma detecção acurada do cio para vacas da raça Holandesa, os pesquisadores utilizarão métodos quantitativos preditivos, por meio da interposição dos pontos intermediários provenientes de uma série histórica do rebanho. Uma base de regras foi formulada sendo que os valores dos pesos de cada afirmação pertenciam a um intervalo de zero a um, e esses limites foram utilizados para gerar a função de pertinência *Fuzzy*, cuja saída era a predição do cio. Na etapa seguinte, foi aplicada a técnica de árvore de decisão utilizando os parâmetros de movimentação, produção de leite, dias de lactação e comportamento de monta, sendo gerada uma árvore de decisão para analisar os parâmetros mais significativos na previsão de cio em vacas leiteiras.

Por fim, uma maneira para lidar com dados georreferenciados através da Mineração de Dados foi abordada no trabalho de Guimarães *et al.* [61]. A pesquisa trata sobre uma ferramenta brasileira conhecida como DAGER, criada para realizar mineração de dados em contextos agrícolas. Os algoritmos de mineração da ferramenta tratam atributos de latitude e longitude, fornecendo inclusive uma visualização dos agrupamentos gerados. Para validar a ferramenta, foi levantado um estudo para verificar se em uma lavoura de soja existem diferenças de produtividade entre as áreas de plantio. As diferenças de produção entre as regiões eram mostradas por meio de uma representação gráfica.

4.3 Trabalhos realizados entre Embrapa e FACOM

A Embrapa - Gado de Corte e a Faculdade de Computação da Universidade Federal do Mato Grosso do Sul possuem um histórico de várias parcerias visando à construção de aplicações para Pecuária de Precisão. O fortalecimento desta parceria aconteceu com o desenvolvimento do portal e-SAPI [62]. O projeto e-SAPI é um portal eletrônico do Sistema Agropecuário de Produção Integrada - Carne Bovina, cujo objetivo é prover softwares de controle e manejo de animais nas propriedades rurais. O e-SAPI é dividido em quatro módulos:

- **e-GTA:** Sistema de Guia de Trânsito de Animais Eletrônica é responsável pela automação do processo de trânsito de animais, que no e-SAPI é feita inteiramente *online*. Tendo as coordenadas de cada propriedade, o sistema possibilita recuperar, no histórico de cada animal, o seu trânsito pelo país e exibi-lo de forma geoprocessada.
- **Sistema de Gestão de Rastreabilidade e de Análise de Riscos:** realiza a propagação de alertas de focos de doenças.
- **e-SAPI Client:** Sistema para a exportação/importação de dados da propriedade (*softwares* das fazendas) para arquivos XML pré-formatados que posteriormente são usados por *Web Services* para alimentar os dados no e-SAPI bovis.
- **Sistema de Gestão de BPA** (Boas Práticas Agropecuárias - Bovinos de Corte): o portal BPA foi construído como parte integrante do e-SAPI bovis, como ferramenta de disseminação de informações relativas, principalmente, às boas práticas agropecuárias definidas através da cartilha do Manual BPA.

Um dos trabalhos mais recentes é o Taurus [3], um sistema de manejo e gestão para bovinos. Tendo em vista o baixo grau de instrução e escolaridade do trabalhador rural, um dos grandes focos do *software* reside justamente na usabilidade de sua interface, ou seja, na facilidade que o usuário tem em interagir com as funcionalidades do sistema. Como exemplo, podemos citar o uso de botões com imagens (semelhante ao padrão usados pelos *smartphones*) que representem as funções a serem selecionadas.

Além da gestão do rebanho, outro objetivo do Taurus é o de fornecer suporte a tecnologias e soluções elaboradas pela Embrapa - Gado de Corte. Atualmente, o sistema

oferece suporte a equipamentos que recebem os dados via portas seriais, a equipamentos que utilizem a estrutura RFID, como a Leitora da Destron (Figura 4.3), e ao *e-SAPI bovis* através do *middleware e-SAPI Client*.



Figura 4.3: Leitora de RFID da Destron

Outro trabalho desenvolvido entre as duas instituições foi o Thermus [5], uma plataforma para aferição remota de temperatura que possibilita o monitoramento das variações de temperatura de bovinos, de forma segura, rápida e precisa, tomando como base os experimentos realizados em animais da raça Brangus. A solução proposta foi encomendar um novo transponder para medição de temperatura e desenvolver um sistema que, através da tecnologia de identificação por rádio frequência, conseguisse monitorar a variação de temperatura nos rebanhos.

O transponder encomendado foi o TXUM00 *Umbilical Implant* (Figura 4.4), um dispositivo que possui um termômetro que, ao mesmo instante da leitura, mede a temperatura do animal. A plataforma implementada precisava converter a temperatura do transponder, situado na prega umbilical do animal, em temperatura retal, possibilitando o monitoramento da temperatura real. A temperatura retal, de acordo com Mota [63], é usada com frequência como índice de adaptação fisiológica ao ambiente quente, pois seu aumento indica que os mecanismos de liberação de calor tornaram-se insuficientes para manter a homeotermia.



Figura 4.4: Alguns transponders TXUM00

Ao final do estudo, a plataforma mostrou-se eficiente na detecção da condição febril dos animais. Entretanto, não foi eficiente em outros testes, como por exemplo, a detecção

de prenhez de uma fêmea. Da mesma maneira que o Taurus, o Thermus também está integrado ao portal e-SAPI bovis.

Por fim, a parceria entre as duas instituições também rendeu o desenvolvimento *middleware* Cerberus [64]. A plataforma abstrai a conexão com dispositivos utilizados no agronegócio, principalmente em conexões com estrutura de rádio frequência, a fim de gerenciar e formatar os dados coletados por estes aparelhos.

Durante o desenvolvimento do Cerberus, para validar a plataforma, foi implementado o Sistema de Balança de Passagem, uma aplicação de registro diário e automático do peso dos animais. O Cerberus envia/recebe dados da leitora RFID e da balança, mostrando assim sua utilidade para a instanciação de aplicações que utilizam diferentes dispositivos para captura de dados, não somente centrado na estrutura RFID. Outra grande contribuição do *middleware* foi padronizar as informações por ele coletadas através do padrão XML, com intuito de alimentar o portal e-SAPI bovis.

4.4 Considerações Finais

Este capítulo apresentou alguns dos principais trabalhos que de alguma forma contribuirão com este projeto. Os trabalhos [56, 57, 58] mostraram como os sensores podem ser empregados no monitoramento animal. Já os trabalhos [59, 60, 61] apresentam maneiras de utilizar a mineração de dados para a extração de conhecimento dos dados produzidos pelo monitoramento do rebanho.

Alguns trabalhos desenvolvidos em parceria entre a Embrapa - Gado de Corte e a FACOM/UFMS também foram abordados, mostrando um pouco do histórico das duas instituições frente a pecuária de precisão.

Capítulo 5

Sistema de Análise de Trajetórias Bovinas

Neste capítulo é descrito a elaboração do sistema, batizado como Sistema de Análise de Trajetórias Bovinas (SATB), assim como as tecnologias envolvidas, as funcionalidades construídas e a forma como a teoria foi utilizada ao *software*.

5.1 Motivação

Conforme apresentado na Seção 1.2, a proposta deste trabalho é a utilização de metodologias e tecnologias de manipulação e análise comportamental de dados espaço-temporais de bovinos a fim de extrair desses conjuntos um conhecimento apurado sobre os animais monitorados pelo produtor/pesquisador.

A proposta inclui o desenvolvimento de um sistema capaz de fornecer as metodologias e tecnologias necessárias para ajudar o usuário a realizar o estudo sobre o conjunto de dados georreferenciados. Além disso, é necessário um Sistema Gerenciador de Banco de Dados (SGBD) capaz de armazenar os objetos espaciais e outros dados utilizados pelo sistema. Por fim, o sistema devia apresentar um modo de visualização para os resultados dos algoritmos.

Levando em conta que o principal objetivo deste estudo é a elaboração de um sistema para a agropecuária, uma aplicação *web* foi descartada pelo fato de poucos domicílios rurais brasileiros possuírem conexão a *internet*. Só para ilustrar esse quadro, atualmente apenas 10% dos brasileiros vivendo na zona rural possuem qualquer tipo de acesso [65]. Logo, optou-se na elaboração de um sistema *desktop*, isto é, um programa que pudesse ser utilizado em uma máquina sem a necessidade de conexão com a *internet*.

5.2 Arquitetura do Sistema

O projeto do SATB seguiu o princípio da Arquitetura de Três Camadas (*Three-Tier Architecture*). Esse paradigma envolve a separação das funcionalidades visuais, lógicas e de conexão com o banco de dados através de camadas. A separação em três camadas torna o sistema mais flexível, de modo que partes podem ser alteradas independentemente. Uma alteração em uma determinada camada não influi nas demais, desde que o mecanismo de comunicação entre elas permaneça inalterado [66].

Além da flexibilidade, o emprego desse modelo privilegia características como modularidade, manutenibilidade, extensibilidade e reusabilidade [67]. A Figura 5.1 ilustra a arquitetura da plataforma.

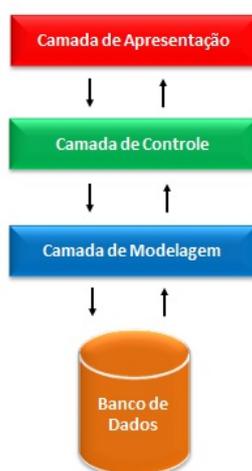


Figura 5.1: Arquitetura da Plataforma.

Na Arquitetura de Três Camadas, cada camada é considerada como um subsistema, oferecendo serviços à camada imediatamente superior. No sistema, as três camadas foram definidas como: Camada de Apresentação, Camada de Controle e Camada de Modelagem, detalhadas a seguir nas seções subsequentes.

5.2.1 Camada de Apresentação

É responsável por exibir as informações do sistema. É usada para receber as entradas e apresentar o resultado. Em outras palavras, a camada é a responsável pela interface com o usuário. É a única camada visível ao usuário final, comunicando-se apenas com a Camada de Controle. As principais funcionalidades implementadas pela camada são apresentadas a seguir.

5.2.1.1 Alimentação do Sistema com Trajetórias dos Animais

Os dados geográficos que compõem as trajetórias devem ser adicionados pelo usuário através da submissão ou de um arquivo texto gerado pela tecnologia da OTAG (*Operational*

Management and Geodecisional Prototype to Track and Trace Agricultural Production) ou de um arquivo ARRF (*Attribute-Relation File Format*).

A tecnologia desenvolvida pelo projeto OTAG, também conhecida como *Electronic Collars to Track Cattle* (ECTC), é uma solução para rastreabilidade bovina constituída de um colar eletrônico com sensor GPS embutido e uma estação base com a função de armazenar os dados espaço-temporais produzidos pelos sensores em campo [68]. A Figura 5.2 mostra como são os equipamentos.



Figura 5.2: O Colar GPS (A) e a Estação Base que armazena os dados dos colares (B).

Esses equipamentos foram adquiridos pela Embrapa - Gado de Corte através do Projeto GeoRastro [69], coordenado pela Embrapa Monitoramento por Satélite. Em campo, os colares coletam a posição geográfica dos animais de acordo com período de tempo determinado na configuração dos dispositivos. Ao se aproximar da estação base, o colar transmite todos os dados armazenados em sua memória para o disco rígido da estação. Todos os dados referentes ao dia da coleta em questão são transcritos em um arquivo texto, armazenado na Estação Base. O Roteamento realizado pela tecnologia OTAG é o Roteamento Oportunístico, descrito na Seção 2.4.1.

Já um arquivo ARFF é um arquivo texto ASCII (*American Standard Code for Information Interchange*) que descreve uma lista de instâncias que compartilham um conjunto de atributos. Este tipo de arquivo foi desenvolvido pelo *Machine Learning Project* da Universidade de Waikato para o uso do sistema Weka de aprendizado em máquina [70].

5.2.1.2 Alimentação do Sistema com Dados Referentes ao Ambiente

Ao contrário dos dados espaço-temporais de movimentação animal, os dados espaciais referentes aos ambientes apresentam seções do tipo CRUD (*Create Read Update Delete*) para que o próprio usuário possa gerenciar essas informações. Sendo cruciais para os

algoritmos de Trajetórias Semânticas, as informações ambientais a serem cadastradas no SATB correspondem às informações sobre a área de pastagem e seus elementos espaciais, tais como áreas de interesse (área do bebedouro, área do cocho), árvores e as suas respectivas áreas de sombra. No contexto da aplicação, as Áreas de Sombra correspondem a regiões de sombreamento (onde não há sol) criadas por uma árvore. Conforme o deslocamento do sol durante o dia e a época do ano, aquela área de sombra deixa de existir e outras áreas são criadas em localizações diferentes, mas sempre próximas da árvore. Logo, uma árvore pode produzir várias áreas de sombra.

Uma região de sombreamento depende exclusivamente de três fatores: a localização da árvore, a altura da árvore e o posicionamento do sol. Como nesta pesquisa, a altura das árvores presentes no ambiente não foi informada, uma solução alternativa para gerar as sombras foi implementada. Como a sombra de uma árvore é dependente da posição do sol, é possível afirmar que durante um dado intervalo de tempo, a sombra existirá em uma região específica. Quanto mais o tempo se aproxima do final do intervalo, a área da sombra tende a diminuir. Mas para fins da aplicação, será considerada que a área continuará sombreada durante todo o intervalo.

5.2.1.3 Visualização de Objetos Espaciais em Mapa

Para a observação dos dados espaciais em um mapa, o sistema faz uso do serviço *Google Earth* [71]. Entretanto, como o serviço não está integrado com o software, é necessário que o serviço esteja instalado na máquina para que possa ser utilizado. Como o *Earth* realiza a visualização dos objetos espaciais tratados pelo SATB, mesmo não estando integrado a plataforma, o serviço é tratado pelo sistema como parte da Camada de Apresentação.

O *Google Maps* chegou a ser cotado para a tarefa, no entanto, o que determinou a escolha do *Google Earth* em relação ao *Maps* foi a necessidade de conexão com a *internet*. Enquanto o *Maps* precisa estar conectado na rede para funcionar, o *Earth* pode ser executado sem a necessidade de estar *online*.

5.2.2 Camada de Controle

Essa camada é responsável pelo controle de todo o fluxo de informação. É onde as funções de negócio, como a análise comportamental através de trajetórias de animais, as operações lógicas, as operações sobre tipos espaço-temporais e manipulação de arquivos são realizadas. Sendo intermediária à Camada de Apresentação e à Camada de Modelagem, ela realiza a comunicação entre as duas camadas. As principais funcionalidades implementadas pela camada são apresentadas a seguir.

5.2.2.1 Cálculos com Tipos Espaciais

De posse das coordenadas geográficas que formam uma trajetória, a Camada de Controle possui funções com a capacidade de calcular o deslocamento total registrado pelos pontos.

Através da **Fórmula de Harvesine** (Figura 5.3) [72], é possível calcular as distâncias entre dois pontos de uma esfera (no caso, o globo terrestre) a partir de suas longitudes e latitudes.

$$\begin{aligned} dlon &= lon2 - lon1; \\ dlat &= lat2 - lat1; \\ a &= (\sin(dlat/2))^2 + \cos(lat1) * \cos(lat2) * (\sin(dlon/2))^2; \\ c &= 2 * \text{atan2}(\text{sqrt}(a), \text{sqrt}(1 - a)); \\ distance &= \text{EarthRadius} * c; \end{aligned}$$

Figura 5.3: Formula de Haversine.

Somando as distâncias entre os pontos da trajetória, o total é distância percorrida pelo objeto móvel em questão.

5.2.2.2 Incorporação de Semântica às Trajetórias

Mais do que conhecer dia-a-dia de seu rebanho, ao monitorar seus animais, as informações que o produtor deseja descobrir **onde, quando e por quanto tempo** um animal esteve em um ambiente de pastagem de uma propriedade. Conhecer as preferências de seus animais pode auxiliar o mesmo na elaboração e manutenção de ambientes de pastagens de qualidade a fim de melhorar o bem-estar do bovino.

A adição de informação geográfica relevante às trajetórias é uma das principais funcionalidades do sistema. Partindo do princípio apresentado no algoritmo IB-SMoT (seção 3.2.2.1), o algoritmo busca encontrar coordenadas das trajetórias que possuam intersecção com os principais tipos geográficos de interesse para a aplicação. Por exemplo, tendo uma trajetória de um bovino em um ambiente de pastagem, dois tipos de objetos geográficos seriam a área do cocho de sal e a área do bebedouro.

Assim como está na definição do algoritmo, o usuário deve fornecer ao sistema as informações sobre os tipos de objetos geográficos de interesse. No caso da plataforma em questão, os objetos geográficos a serem fornecidos são acrescentados na pastagem, que também é informada pelo usuário. Em uma área de pastagem, os objetos geográficos de interesse podem corresponder a áreas específicas do ambiente, a árvores e a regiões de sombra. Outra informação que deve ser informada pelo usuário é o **período de tempo mínimo** que uma trajetória móvel precisa interceptar um objeto geográfico para ser considerada como relevante para a plataforma.

5.2.2.3 Localização de Regiões de Alta Concentração

Em um ambiente de pastagem, existem lugares onde os animais possuem uma maior preferência em ficar do que outros, seja pelo tipo de capim, pela quantidade de árvores, distância do bebedouro, sempre haverá na área de pastagem regiões onde gado se concentra com mais frequência. Para descobrir essas regiões, a plataforma utiliza as metodologias CB-SMoT e o algoritmo de Mineração de Dados DBSCAN.

Como já foi descrito anteriormente, o CB-SMoT (Seção 3.2.2.2) consiste em duas etapas, onde a primeira localiza os conjuntos de pontos de menor velocidade de uma trajetória e a segunda aplica o IB-SMoT (Seção 3.2.2.1) na trajetória com a finalidade de verificar se esses conjuntos estão em alguma região importante. Para a identificação dos lugares onde os animais se concentram, a aplicação do CB-SMoT no sistema fica restrita apenas a primeira fase do algoritmo, pois ao obter os conjuntos de baixa velocidade de uma trajetória, esses são enviados para a Camada de Modelagem e armazenados no Banco Dados.

Ao aplicar o DBSCAN (Seção 3.2.2.2) para esses conjuntos de uma ou mais trajetórias, o resultado final será um ou mais conjuntos de coordenadas bastante densos que, ao serem enviados para a Camada de Apresentação e dispostos em um mapa, revelam os locais onde os animais mais se concentraram no ambiente de pastagem. Conforme explanado na Seção 3.2.2.2, os pontos que não se encaixarem nos argumentos recebidos pelo algoritmo, serão considerados ruídos e, conseqüentemente, eliminados do resultado final.

5.2.2.4 Identificação de Áreas Evitadas pelos Animais

Da mesma maneira que há lugares em um ambiente de pastagem onde os animais possuem maior preferência, existem também áreas onde os bovinos possuem uma tendência a evitar. Seja pela distância da região de uma fonte aquífera, pela falta de árvores ou até mesmo por motivos de superpastejo, esses locais dificilmente são visitados pelo rebanho. Logo, a identificação dessas regiões é vital para que o produtor possa tomar as devidas providências a fim de aperfeiçoar o uso de sua área de pastagem.

A metodologia utilizada para a identificação dessas regiões é o padrão *Avoidance* (Seção 3.2.2.3). Para a aplicação da técnica, o ambiente onde a trajetória foi realizada será dividido em várias áreas circulares. A escolha da circunferência baseou-se no argumento de Loy [55], o responsável pelo padrão *avoidance*. Segundo ele, áreas circulares são mais triviais de se trabalhar devido ao formato do objeto geométrico.

O padrão a ser utilizado neste estudo possui algumas diferenças com o descrito em [55]. Originalmente, a metodologia havia sido testada em uma aplicação na área de segurança no trânsito, onde os caminhos são pré-determinados (as ruas) e as áreas objeto-alvo eram locais da rua onde havia presença de câmeras. Na aplicação deste projeto, as áreas objeto-alvo são todas as regiões obtidas pelo processo de divisão da área de pastagem e, ao contrário do trânsito de uma cidade, por ser um ambiente bem mais aberto (apesar das cercas que delimitam o espaço), há vários caminhos para se chegar ou aproximar de um objeto-alvo.

Para a identificação das regiões que os bovinos tendem a não ir, cada área presente no ambiente de pastagem receberá **uma taxa de desvios** após a execução do padrão, onde a variação desse valor irá determinar a ocorrência de desvios. Quando esta taxa é 0%, significa que não houve presença de desvios. Quando varia de 1% até 49%, significa que houve uma baixa ocorrência de desvios. De 50% à 99% indica que houve uma alta ocorrência de desvios. Uma taxa de 100% significa que as trajetórias sempre desviaram seu rumo ao se aproximarem da região. Por fim, caso uma área do ambiente tenha 0%

de desvios e nenhuma trajetória tenha a interceptado, essa região será considerada como evitada pelo animal.

Ainda no projeto original, caso uma área objeto-alvo obtivesse uma taxa de desvio 100%, era bem provável que houvesse algo impedindo o trânsito naquela região espacial, de modo que esses *avoidances* iriam deturpar o resultado final no cálculo do desvio das trajetórias. Logo nesta situação, o objeto-alvo era desconsiderado - como se não existisse. Para a aplicação desta pesquisa, uma área com 100% não será desconsiderada uma vez que há interesse em conhecer toda e qualquer área do ambiente de pastejo evitada pelos animais. Outra adaptação realizada neste estudo é no termo Região de Interesse, explicado na Seção 3.2.2.3. A fim de evitar confusões com termo Área de Interesse, o termo será redefinido como **Região de Desvios**.

5.2.2.5 Manipulação de Arquivos de Entrada sobre Trajetórias Animais

Quando o sistema recebe um arquivo entrada, seja um arquivo texto gerado por uma tecnologia OTAG ou um arquivo ARFF (Seção 5.3.1.1), a camada de apresentação envia os arquivos para a camada de controle, para que os dados sejam lidos, interpretados, formatados e posteriormente enviados a camada que fará uso das informações. A Figura 5.4 mostra como é a disposição e a formatação dos dados de um arquivo texto armazenado em uma Estação Base da OTAG.

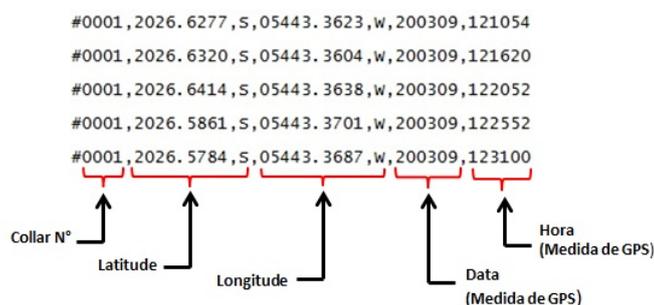


Figura 5.4: Exemplo do formato dos arquivos texto criado por um colar da OTAG.

Como visto acima, todas as linhas com os dados de coleta iniciam com uma *#* (*hashtag*) seguida do código do colar. Os elementos das linhas são separados por vírgulas e entre cada linha com dados, existe uma linha em branco. Os dados do GPS são escritos na forma WGS84 [73], onde suas coordenadas assumem o formato DDMM.MMMM, conhecido como *Degrees and Minutes*. Para entender melhor o formato, seja a coordenada 5529.1355. Os caracteres DD são a parte inteira do grau de uma coordenada, onde no caso do valor seria 55. Os caracteres MM que estão antes do ponto são a porção inteira dos minutos de uma coordenada, onde em nosso exemplo corresponde ao valor 29. Por fim, os caracteres MMMM correspondem a parte fracionária dos minutos, logo, para o exemplo acima o valor deste trecho é 0.1355.

Ao lado de cada coordenada, há um caractere indicando se o objeto está ao Norte da linha do Equador (N de *North*) ou ao Sul (S de *South*), e se está ao Oeste do Meridiano de

Greenwich (W de *West*) ou ao Leste (E de *East*). As informações temporais (data e hora) se encontram respectivamente no formato ddmmyy (onde dd corresponde aos dígitos dos dias, mm aos dígitos do mês e yy aos dois últimos dígitos de um ano) e HHmmss (onde HH corresponde aos dígitos da hora, mm corresponde aos dígitos de minutos e ss aos de segundos).

Enquanto o processo de leitura é realizado, outro procedimento de formatação dos dados é efetuado. Nos dados temporais, são adicionados os caracteres “/” para a data e os caracteres “:” na hora para que o formato desses dois objetos sejam, respectivamente dd/mm/yy e HH:mm:ss. Já em relação as coordenadas geográficas, o formato DDMM.MMMM é convertido para o formato D.D (*Decimal Degrees*). A conversão é bastante simples: basta dividir MM.MMMM por 60 para se obter .D, depois adicione .D ao D para ter o D.D. Por exemplo, seja a mesma coordenada 5529.1255. Para se obter o .D, basta dividir 29.1255 por 60, resultando no valor 0.485425. Juntando o valor de .D com o valor de D (onde o valor é 55), a soma dos valores resultam no valor 55.485425.

Quanto às direções geográficas, essa informação será anexada ao valor das coordenadas, ou seja, se a latitude pertencer ao Sul e a longitude pertencer ao Oeste, os valores de suas coordenadas serão multiplicados por -1 . Caso contrário, as coordenadas permanecerão positivas. Essas mudanças são necessárias para melhor aproveitamento dos algoritmos e disposição visual desses dados.

Em relação ao outro formato de entrada via arquivo ARRF, a Figura 5.5 mostra um exemplo desse tipo um arquivo contendo dados espaço-temporais referentes ao projeto.

```
@RELATION trajectory

@ATTRIBUTE collar STRING
@ATTRIBUTE longitude NUMERIC
@ATTRIBUTE latitude NUMERIC
@ATTRIBUTE day DATE "dd/MM/yy"
@ATTRIBUTE hour DATE "HH:mm:ss"

@DATA
0001,-54.72276166666666,-20.44302,19/03/09,13:56:11
0001,-54.722835,-20.443041666666666,19/03/09,14:01:09
0001,-54.72273864734826,-20.44292410636202,19/03/09,14:06:37
0001,-54.72286,-20.442993333333334,19/03/09,14:11:20
0001,-54.722916666666667,-20.44305,19/03/09,14:16:14
0001,-54.722891666666667,-20.44304,19/03/09,14:21:14
0001,-54.722916666666667,-20.443051666666666,19/03/09,14:26:14
```

Figura 5.5: Exemplo do formato dos dados em um ARRF.

O ARRF é dividido em duas seções bem distintas, onde a primeira corresponde às informações de cabeçalho e a segunda corresponde às informações dos dados. O cabeçalho contém o nome da relação, a lista de seus atributos e os seus tipos. Em alguns tipos, como o DATE, é possível especificar qual é o formato do atributo em questão. Já a seção de dados é formada por uma linha de declaração de dados (@DATA) e as linhas de instâncias. Como pode ser visto na figura anterior, os dados referentes à latitude, à longitude, à data e à hora já estão na formatação exigida pelo sistema. Logo, quando o sistema receber um

arquivo deste tipo, o único trabalho da camada de controle será ler os dados armazenados no arquivo e enviá-los para a camada que os solicita.

A simplicidade do arquivo ARFF acabou por popularizar o formato para os *softwares* de Mineração de Dados. Tanto que atualmente, a grande maioria dos sistemas da área possui suporte a esse formato.

5.2.2.6 Manipulação de Arquivos KML

Como visto anteriormente, o SATB utiliza o serviço *Google Earth* para a visualização dos objetos espaciais gerados e/ou armazenados em sua estrutura. Para que o serviço possa gerar a visualização desejada pelo usuário, é necessário o envio de um arquivo KML (*Keyhole Markup Language*) como parâmetro de entrada. O KML possui uma formatação bastante similar a de um arquivo XML, como pode ser observado na Figura 5.6.

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Placemark>
    <name>Mangueiro Digital</name>
    <description>Um ponto posicionado
nas imediações do mangueiro digital
da Embrapa - Gado de Corte.</description>
    <Point>
      <coordinates>-54.72276847915689,-20.44296329610641,0</coordinates>
    </Point>
  </Placemark>
</kml>
```

Figura 5.6: Exemplo da formatação de código de um KML.

Um arquivo KML é gerado através dos dados espaciais recebidos da Camada de Modelagem. Como o *Google Earth* é considerado pelo sistema como uma parte da Camada de Apresentação, o KML criado na Camada de Controle é enviado diretamente para esse serviço.

5.2.3 Camada de Modelagem

É responsável pelo armazenamento e a modelagem dos dados da aplicação. Entre as três camadas, a de Modelagem é a única com acesso ao Banco de Dados. Ela se comunica apenas com a Camada de Controle. As principais funcionalidades implementadas pela camada são apresentadas a seguir.

5.2.3.1 Operações no Banco de Dados

Por ser a única camada com acesso ao Banco de Dados, a Camada de Modelagem é a única capaz de realizar as operações básicas (inserção, remoção, edição e seleção) na base de dados. Quando uma operação é solicitada pela Camada de Apresentação, o pedido é enviado para a Camada de Controle que encaminha esse pedido para a Camada de

Modelagem. O caminho inverso é realizado quando a Camada de Modelagem precisa enviar dados para uma das camadas superiores.

5.2.3.2 Operações sobre Dados Espaciais

Quando um algoritmo da camada de controle necessita realizar uma função de ordem espacial, ela se comunica com a camada de modelagem solicitando a execução da funcionalidade. Ao concluir a operação, a camada retorna o resultado.

5.3 Implementação

Seguindo os objetivos estabelecidos e o projeto de arquitetura mostrado, as próximas seções detalham a implementação do Sistema de Análise de Trajetórias Bovinas (SATB).

5.3.1 Tecnologias Envolvidas

A linguagem escolhida para a implementação do sistema computacional foi o **Java**. Essa escolha envolveu vários aspectos: suporte a Orientação a Objeto, APIs (*Applications Programming Interface*) e ferramentas de fácil integração, elaboração de ricas *interfaces* gráficas e principalmente, o fato de ser a linguagem padrão da Embrapa - Gado de Corte para construção de seus sistemas.

O SATB foi codificado na IDE Netbeans, versão 7.2.1. Para o controle de versões, o Tortoise SVN (versão 1.7.9) foi utilizado juntamente de um repositório fornecido pelo *Google Code*. Uma ferramenta auxiliar para esse controle foi o serviço de armazenamento *Dropbox*. A versão Java usada na implementação foi a **JDK 1.7**, para máquinas de 64 bits (x64) de processador. A Interface Gráfica foi realizada através **JavaFX 2.2** [74], uma plataforma multimídia desenvolvida e distribuída pela Oracle. Em relação às APIs Java, foram integradas ao sistema:

- **JDOM:** é uma API que fornece uma solução robusta e leve para leitura e a escrita de dados XML sem opções complexas e sem o alto consumo de memória oferecida por outras APIs do gênero. No projeto, o JDOM (*Java Document Object Model*) fornece funções para a escrita de arquivos KML, que possuem um formato muito parecido com o XML [75].
- **Joda Time:** é uma API que lida com objetos temporais, como data e hora. Ela permite múltiplos sistemas de calendário como o Gregoriano, Juliano, Budista, Etíope, entre outros. O padrão *default* da ferramenta é a ISO8601. No projeto, ela lida com as operações entre os atributos temporais [76].
- **PostgreSQL:** um JDBC (*Java Database Connectivity*) que trata a conectividade entre o código Java com um Banco de Dados PostgreSQL [77].

- **PostGIS:** um JDBC que trata a conectividade entre o código Java com a extensão do PostgreSQL, o PostGIS [78].
- **Weka:** é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou serem invocados a partir de um código Java. Weka contém ferramentas para pré-processamento de dados, classificação, regressão, *clustering*, regras de associação e visualização. No projeto, a API fornece funções para leitura e criação de arquivos ARRF [79].

O Sistema de Gerenciamento de Banco de Dados (SGBD) selecionado para armazenamento dos dados foi o PostgreSQL. Além de possibilitar uma gama de tipos de armazenamento e múltiplas funções, ele ainda possui o PostGIS, uma extensão capaz de fornecer armazenamento para objetos espaciais além de inúmeras funções a serem executadas através de comandos via SQL (*Structured Query Language*), como por exemplo, verificar se uma coordenada qualquer está contida em uma determinada área.

Para a visualização dos objetos espaciais (trajetórias, áreas, coordenadas), o SATB utiliza as funcionalidades do programa *Google Earth*. Apesar de fornecer suporte a conexão *web*, o *software* pode ser utilizado sem necessitar de qualquer tipo de conexão com a *internet*.

5.3.2 Funcionalidades Implementadas

No menu principal do SATB, as funcionalidades implementadas foram divididas em três grupos. O primeiro corresponde a uma única seção onde as trajetórias são adicionadas ao *software*. O segundo grupo corresponde as seções onde os dados referentes ao ambiente são cadastrados. Por fim, o terceiro e último grupo corresponde as seções onde as trajetórias são analisadas. As principais implementações realizadas serão apresentadas nas próximas subseções.

5.3.2.1 Submissão das Trajetórias Bovinas

Como descrita na Arquitetura (Seção 5.2), a adição dos dados referentes às trajetórias animais são realizadas através da submissão de dois tipos de arquivo: um arquivo texto criado pela tecnologia OTAG e um arquivo ARFF. A Figura 5.7 apresenta a janela principal desta funcionalidade.



Figura 5.7: Janela referente a submissão de arquivos de trajetórias.

A submissão é realizada de maneira simples. Selecionando o tipo de arquivo entrada e em seguida apertando o botão abrir, basta selecionar o documento onde se encontra as trajetórias. O *software* realizará leitura, formatação dos dados e, por fim, a adição desses dados no repositório.

5.3.2.2 Gerenciamento dos Dados Referentes ao Ambiente

O gerenciamento dos dados referentes ao ambiente corresponde ao segundo grupo de funções do SATB, onde o sistema possibilita ao usuário opções do tipo CRUD (*Create Read Update Delete*). Os elementos ambientais com seções neste grupo são: áreas de pastagem, áreas de interesse (regiões especiais que estão na área de pastejo), árvores e áreas de sombreamento.

Vale ressaltar que os elementos áreas de interesse, árvores e áreas de sombreamento estão todas associadas a uma área de pastagem e os elementos de sombra estão associados a uma árvore. Todas as seções deste grupo possuem o mesmo padrão visual para suas telas. Exemplificando essa afirmação, a Figura 5.8 apresenta a janela principal da seção Área de Pastagem (Figura 5.8a) e a janela principal da seção Áreas de Sombra (Figura 5.8b).

(a) Janela Principal da seção Área de Pastagem

Opções: Novo Remover Editar

Áreas de Pastagem Cadastradas

Área de Pastagem: Mangueiro Digital

Hectares: 4.090000152587891

Coordenadas da Área de Pastagem

Longitude	Latitude
-54.7240306968837	-20.443957813812
-54.7227924450442	-20.4427089381523
-54.7227244141028	-20.4428275214557
-54.7227476624494	-20.4439954732498
-54.7240306968837	-20.443957813812

Ver Área

(b) Janela principal da seção Áreas de Sombra

Opções: Novo Remover Editar

Áreas Sombreadas Cadastradas

Área de Pastagem: Mangueiro Digital Árvore: arv001

Área Sombreada: sombraarv001

Intervalo Horário: 12:00:00 à 13:00:00

Coordenadas da Área Sombreada

Longitude	Latitude
-54.7231654597465	-20.4436193974798
-54.7231423130111	-20.4436516251696
-54.7230743308849	-20.4436872276963
-54.7229840668006	-20.4437132221581
-54.7229021578881	-20.4436563377035
-54.7228393251572	-20.443558763674
-54.7228695068903	-20.443510319498

Ver Área

(a) Janela Principal da seção Área de Pastagem. (b) Janela principal da seção Área de Sombra.

Figura 5.8: Janela Principal das seções Área de Pastagem e Área de Sombra.

Como observado na figura anterior, às janelas principais de Área de Pastagem e de Área de Sombra apresentam a possibilidade de visualização das informações de cada elemento cadastrado nas seções. No topo das duas janelas, as seções apresentam as outras opções de gerenciamento: adicionar novo elemento, remover elemento e editar as informações de um elemento cadastrado. Como já afirmado anteriormente, todas as seções deste grupo possuem essas opções de gerenciamento.

Observando a figura referente à janela principal da seção Área de Sombra (Figura 5.8b), é possível ver um exemplo da associação dos objetos espaciais. Para se visualizar os dados referentes a uma área de sombra, primeiro é necessário selecionar em qual área de pastagem essa sombra está. Com a área de pastagem selecionada, o segundo passo é selecionar a árvore no qual a sombra está associada. Só depois disso será possível escolher a área de sombra.

Outra particularidade dessas seções pode ser observada na Figura 5.9, onde há exemplos de uma janela de adição de uma nova Área de Pastagem (Figura 5.9a) e de uma janela de adição de uma nova Área Sombreada (Figura 5.9b).

The figure shows two screenshots of a web application interface. The left screenshot, titled 'Adicionar Área de Pastagem', shows a form where the user has selected 'Mangueiro Digital' as the pasture area and '4.09' hectares. Below this is a table with two columns: 'Longitude' and 'Latitude'. The table contains two rows of coordinates: (-54.7220306968, -20.44395781) and (-54.722792445, -20.442708938). At the bottom, there are input fields for longitude and latitude, and buttons for 'Adiciona Dados', 'Abrir ARFF', 'Adicionar', and 'Cancelar'.

The right screenshot, titled 'Adicionar Área Sombreada', shows a form where the user has selected 'Mangueiro Digital' as the pasture area and 'arv001' as the tree. Below this is a table with two columns: 'Longitude' and 'Latitude'. The table contains five rows of coordinates: (-54.7231175256001, -20.44303061047191), (-54.72272971762857, -20.44307348616004), (-54.72272271651603, -20.44283141830334), (-54.72279411373, -20.44270778915093), and (-54.7231175256001, -20.44303061047191). At the bottom, there are input fields for longitude and latitude, and buttons for 'Adicionar', 'Abrir ARFF', 'Adicionar', and 'Cancelar'.

(a) Janela de Adição de uma nova Área de Pastagem.

(b) Janela de Adição de uma nova Área de Sombra.

Figura 5.9: Janela de Adição de Área de Pastagem e Área de Sombra.

Como pode ser observado nas duas figuras, a tela adição apresenta uma tabela onde as coordenadas que formam a área daquela determinada região são adicionadas. A criação do polígono em questão pode ser realizado de duas formas: escrevendo e adicionando uma coordenada por vez através dos campos latitude e longitude ou através da submissão de um arquivo ARFF contendo todos os pontos que formam o espaço em questão.

Em comparação aos ARFFs de trajetórias, os ARFFs com os pontos de uma área são bem mais simples uma vez que só há a necessidade de duas relações (latitude e longitude) para representar as coordenadas. A única seção de gerenciamento que não apresenta tabelas para representação de áreas é a de árvores, já que uma árvore pode ser representada apenas por um ponto. A Figura 5.10 apresenta o formato do ARFF para pontos de uma área.

```

@RELATION area

@ATTRIBUTE longitude NUMERIC
@ATTRIBUTE latitude NUMERIC

@DATA
-54.72402618788541,-20.44395891119062
-54.7227469485695,-20.44399454187042
-54.72274300341695,-20.44372013106546
-54.7237648467254,-20.44368672607686
-54.72402618788541,-20.44395891119062

```

Figura 5.10: ARFF com as coordenadas de uma área.

As janelas de gerenciamento de áreas de interesse e áreas de sombra (Figura 5.9b) apresentam um atributo em comum, conhecido como **tempo mínimo de permanência**. O valor deste atributo temporal corresponde à duração mínima de tempo em que um animal precisa estar em uma sombra ou em um bebedouro para que aquela “parada” seja reconhecida como relevante ao sistema. O atributo será importante no momento em que as trajetórias bovinas passarem pelo processo de semântica.

A janela de edição apresenta o mesmo formato das janelas de adição, onde a diferença reside no fato de existir um espaço para a escolha do objeto a ser editado. Já a janela de remoção apenas apresenta um espaço de seleção do objeto a ser removido do banco.

5.3.2.3 Adicionar Semântica para Trajetórias

Partindo para o grupo de seções referentes a trajetórias, a seção de Adição de Semântica é onde o usuário poderá escolher pela aplicação de dois algoritmos de trajetórias semânticas, o IB-SMoT (Seção 3.2.2.2) e o CB-SMoT (Seção 3.2.2.2). Enquanto o primeiro resultará nos lugares em que a trajetória passou e o tempo que ela permaneceu em cada lugar, o segundo resultará em um ou mais conjuntos de pontos de baixa velocidade. A Figura 5.11 mostra duas imagens com a mesma janela, mas cada uma com um dos algoritmos selecionados.

(a) Aplicação do algoritmo IB-SMoT.

(b) Aplicação do algoritmo CB-SMoT.

Figura 5.11: Tela de adição de semântica para trajetórias.

Enquanto o IB-SMoT só precisa da trajetória e da área de pastagem como parâmetros

para o funcionamento do algoritmo, o CB-SMoT necessita de três parâmetros: tempo mínimo, velocidade limite e velocidade média do conjunto. Os valores passados para os atributos de velocidade representam porcentagens da velocidade média da trajetória selecionada. Por exemplo, a velocidade limite é 1 e a velocidade média do conjunto é 0.8. Quando o algoritmo for executado, ele irá atribuir 100% do valor da velocidade média da trajetória ao atributo velocidade limite e 80% do valor para velocidade média do conjunto. A maneira como esses atributos são usados no algoritmo é explicado na Seção 3.2.2.2.

A adição de semântica permite a aplicação do algoritmo em trajetórias presentes no Banco de Dados e em trajetórias presentes em um arquivo ARFF. A diferença entre ambos é simples, na primeira os resultados do algoritmo são armazenados na base de dados e no segundo os dados não são adicionados. Durante o desenvolvimento do sistema, o uso dos ARRFs nesta seção tinha como intuito fazer testes sem que houvesse o armazenamento dos resultados.

Os resultados dos dois algoritmos serão utilizados para verificar a presença do animal em uma área de interesse de um ambiente de pastagem, averiguar o período em que permaneceu na área de interesse, descobrir se o animal esteve em uma área de sombra e identificar as regiões da área de pastagem onde os bovinos tem se concentrado.

5.3.2.4 Verificação da Presença Animal em uma Área de Interesse

Com a finalidade de explorar as saídas dos algoritmos IB-SMoT (Figura 5.11a) e CB-SMOT (Figura 5.11b), além de apresentar informações extraídas das trajetórias sem a aplicação de um método de semântica, foi implementada a seção Perguntas e Respostas. A Figura 5.12 mostra a janela principal desta seção, juntamente com as informações que se pode obter.

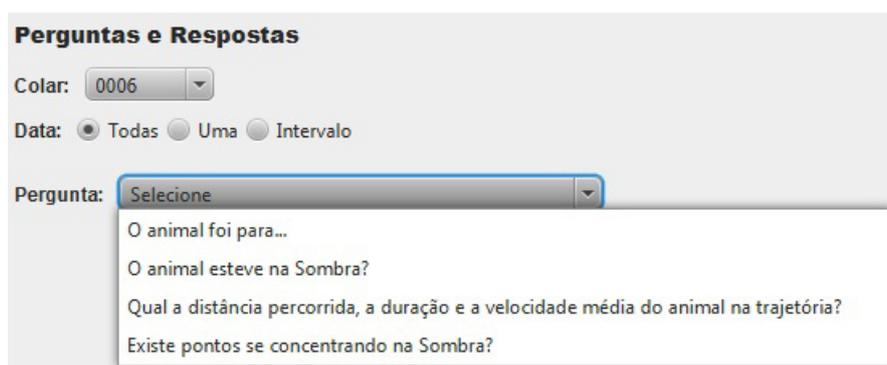


Figura 5.12: As principais perguntas da seção Perguntas e Respostas.

Para responder à primeira pergunta, utiliza-se o resultado da aplicação do IB-SMoT. Ao selecionar a questão “O animal foi para...”, duas caixas de combinação (*ComboBoxes*) surgirão logo abaixo. A primeira caixa apresenta a escolha do ambiente de pastagem e a segunda apresenta as áreas de interesse referentes ao ambiente selecionado pelo usuário. Na Figura 5.13, a tela apresenta a resposta para a primeira pergunta, onde o ambiente de pastejo selecionado foi o Mangueiro Digital e a área de interesse foi *Área 4*.

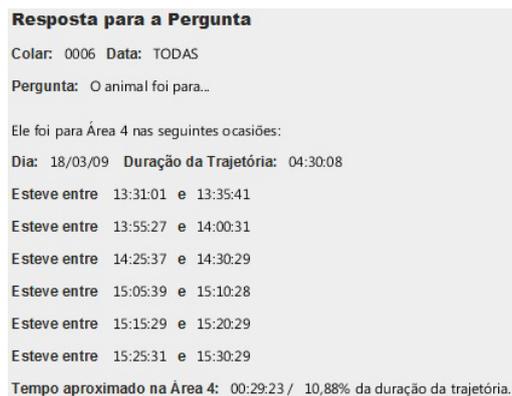


Figura 5.13: Relação dos dias e do momento em que as trajetórias do colar 0006 passaram pela área selecionada.

A Figura 5.13 mostra a tela de resposta detalhando o dia e os intervalos horários em que um determinado animal esteve presente na região escolhida pelo usuário. O sistema fornece como saída o período total que o animal permaneceu na região, apresentando inclusive a porcentagem do tempo em relação à duração total da trajetória.

5.3.2.5 Verificação da Presença Animal em uma Área de Sombra

Quando o IB-SMoT (Figura 5.11a) é aplicado em uma trajetória, logo em seguida uma versão alternativa da técnica é aplicada para verificar se em algum momento o animal esteve em uma área de sombreado. Escolhendo a pergunta “O animal esteve na sombra”, a saída é mostrada em uma tela na Figura 5.14.

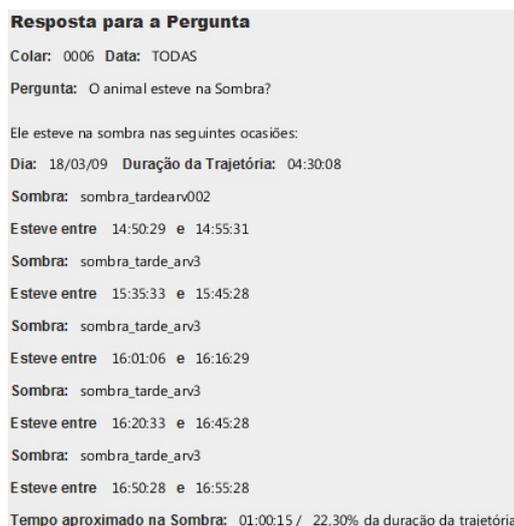


Figura 5.14: Relação dos momentos em que as trajetórias do colar 0006 interceptaram uma área de sombra.

Como pode ser observado na Figura 5.14, o formato da resposta é bem semelhante ao apresentado pela Figura 5.13, com pequenas diferenças como a presença do nome da área

de sombra a qual o animal esteve durante sua interação com o ambiente de pastagem.

5.3.2.6 Distância Percorrida, Duração e Velocidade Média de uma Trajetória

A terceira pergunta (“Qual a distância percorrida, a duração e a velocidade média do animal na trajetória?”) da seção Perguntas e Respostas não necessita da aplicação de semântica, ao contrário das duas perguntas anteriores, pois o sistema utiliza cálculos simples para extrair essas informações. A Figura 5.15 apresenta a janela com as respostas para a pergunta.

Resposta para a Pergunta

Colar: 0006 Data: TODAS

Pergunta: Qual a distância percorrida, a duração e a velocidade média do animal na trajetória?

Distância percorrida (metros), velocidade média (m/s) e duração (horas).

Dia:	18/03/09	Distância:	1766.85	Vel. Média:	0,109	Duração:	04:30:08
Dia:	19/03/09	Distância:	729.37	Vel. Média:	0,081	Duração:	02:29:58

Figura 5.15: Janela principal da seção Padrões em Trajetórias.

Para cada trajetória presente no colar escolhido, é apresentado o dia que a trajetória foi realizada, a distância percorrida pelo animal em metros, a velocidade média do animal em metros por segundo e a duração deste percurso em horas.

5.3.2.7 Identificação de Regiões de Concentração

Com a criação dos conjuntos de baixa velocidade descrita na Seção 5.3.2.3, a segunda etapa do procedimento para a identificação de regiões de concentração consiste na execução de um algoritmo de agrupamento (*clustering*) sobre os conjuntos extraídos das trajetórias. A seção Agrupamento em Trajetórias (Figura 5.16), apresenta os procedimentos necessários para a tarefa.

Agrupamentos em Trajetórias

Selecione o Algoritmo: DBSCAN

Distância: 10 Pts. Min. na Área: 7

Colar: 0009

Data: Todas Uma Intervalo

Intervalo de Dias: 01/08/13 até 03/08/13

Horário: Nenhum Intervalo Antes Depois

Horário Depois: 12:00:00

Figura 5.16: Janela principal da seção Agrupamentos em Trajetórias.

O algoritmo disponível é o DBSCAN, uma técnica de agrupamento de pontos baseada em densidade, conforme explicado na Seção 3.2.2.2. Os parâmetros passados para o algoritmo consistem no valor da distância em metros e no valor mínimo de um conjunto. Lembrando que um conjunto de concentração é uma região onde todos os pontos apresentam pelo menos um número mínimo de pontos a uma distância ao seu redor.

Como visto na Figura 5.16, os conjuntos de baixa velocidade a serem selecionados em um banco de dados ainda podem contar com o uso de filtros de data e hora a fim de criar uma busca ainda mais específica no banco. Para exemplificar a aplicação do algoritmo, aplica-se o DBSCAN para as trajetórias do colar 0009 coletadas entre os dias 01/08/13 e 03/08/13 a partir do meio dia. Setando o valor da distância para 10 metros e o conjunto mínimo para 7, o resultado pode ser conferido na Figura 5.17.

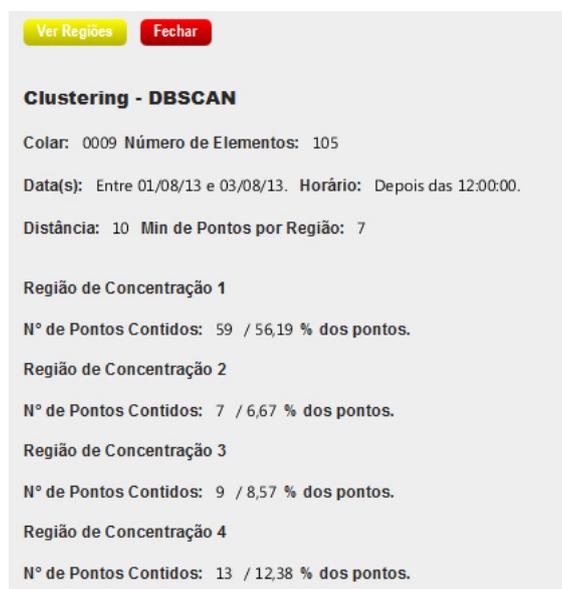


Figura 5.17: Janela de resultado do agrupamento executado.

A Figura 5.17 mostra o resultado do algoritmo, apresentado em uma nova janela. O cabeçalho da tela apresenta as informações de entrada juntamente com a quantidade de dados selecionados que correspondiam com os requisitos escolhidos. Logo abaixo, são enumeradas quantas regiões de concentração foram criadas, juntamente com a quantidade de pontos e a porcentagem que esses pontos correspondem ao total selecionado.

Por fim, a tela apresenta a quantidade de dados total presentes nas regiões de concentração e o número de dados que não satisfizeram os dois parâmetros do DBSCAN. Uma forma mais visual do resultado pode ser conferida ao clicar no botão **Ver Regiões**. O serviço de mapa será invocado e apresentará a disposição das regiões na área de pastagem (Figura 5.18).



Figura 5.18: Regiões de Concentração representadas em um mapa.

Na Figura 5.18, as regiões de concentração são diferenciadas pela cor dos pontos que as compõe. A legenda presente na imagem identifica cada região, de acordo com as informações mostradas pela Figura 5.17. O formato visual, além de um complemento para os resultados em texto, é uma forma mais agradável de se apresentar a informação.

5.3.2.8 Verificar Existência de Pontos de Concentração em Áreas de Sombra

A última pergunta (“Existe pontos se concentrando na Sombra?”) da seção Perguntas e Respostas faz uso dos conjuntos de pontos de baixa velocidade, gerados pela execução do CB-SMoT (Figura 5.11b), com objetivo de verificar se os conjuntos interceptaram áreas de sombra. Além de verificar a intersecção de pontos nas áreas, os intervalos temporais dos conjuntos e das áreas de sombra são avaliados para confirmar a existência da sombra no momento em que supostamente os pontos estariam dentro de seus limites. A Figura 5.19 apresenta um exemplo de uma janela com a resposta para a pergunta.

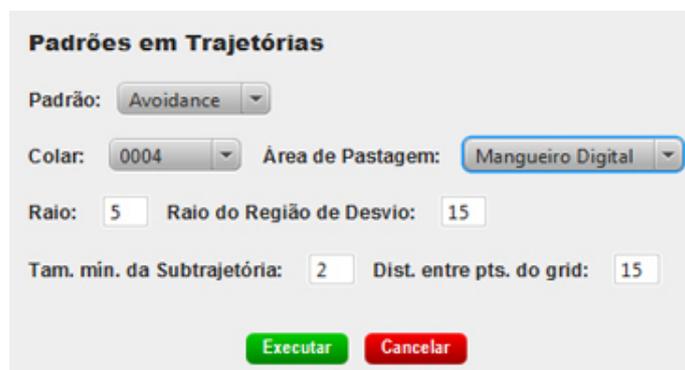
Resposta para a Pergunta	
Colar: 0001	Data: TODAS
Pergunta: Existe pontos se concentrando na Sombra?	
O animal se concentrou na sombra nas seguintes ocasiões	
Dia: 19/03/09	Duração da Trajetória: 02:49:57
Subtraj. de pts. de baixa vel. : 13:56:11 - 14:36:12	Sombra Interceptada: sombra_tardearv002
Subtraj. de pts. de baixa vel. : 15:16:09 - 15:31:07	Sombra Interceptada: sombra_tardearv002
Subtraj. de pts. de baixa vel. : 15:51:08 - 16:31:09	Sombra Interceptada: sombra_tardearv002

Figura 5.19: Tela com a resposta da pergunta “Existe pontos se concentrando na Sombra?” .

Caso um conjunto de pontos de baixa velocidade esteja em uma área de sombra, a resposta será representada com o intervalo de tempo na qual os pontos deste conjunto estão contidos e com o nome da área de sombra interceptada.

5.3.2.9 Identificação Áreas Evitadas pelos Animais

A seção Padrões em Trajetórias é onde o usuário pode aplicar o Padrão *Avoidance* sobre um determinado conjunto de trajetórias, em uma determinada área de pastagem, a fim de encontrar as áreas evitadas pelos bovinos. A Figura 5.20 apresenta a janela principal desta seção.



A janela 'Padrões em Trajetórias' contém os seguintes elementos:

- Padrão: **Avoidance** (menu suspenso)
- Colar: **0004** (menu suspenso) | Área de Pastagem: **Mangueiro Digital** (menu suspenso)
- Raio: **5** (campo de texto) | Raio do Região de Desvio: **15** (campo de texto)
- Tam. min. da Subtrajetória: **2** (campo de texto) | Dist. entre pts. do grid: **15** (campo de texto)
- Botões: **Executar** (verde) e **Cancelar** (vermelho)

Figura 5.20: Janela principal da seção Padrões em Trajetórias.

A aplicação do padrão consiste na escolha de um conjunto de trajetórias relacionadas a um colar e na área de pastagem onde essas trajetórias foram realizadas. O usuário precisa informar o valor de três parâmetros para o algoritmo: o raio em metros da área objeto-alvo, o raio em metros da região de desvio e o tamanho mínimo em metros da subtrajetória. O uso de todos esses atributos é explicado na seção 3.2.2.3.

Com objetivo de exemplificar a aplicação do padrão de desvios, aplica-se o algoritmo para as trajetórias presentes no colar 0004 que foram realizadas na área de pastejo do Mangueiro Digital. Os parâmetros escolhidos foram 5, 15 e 2 para raio, raio da região de desvio e tamanho mínimo da subtrajetória, respectivamente. A Figura 5.21 mostra a tela de resultado do padrão *avoidance*.



A janela de resultados 'Padrão - Avoidance' apresenta:

- Botões: **Ver Áreas** (verde) e **Fechar** (vermelho)
- Colar: **0004** | N° Divisões na Área de Pastagem: **36**
- Raio: **5** | Raio da Reg. de Desvio: **15** | Tam: **2**
- Trajetoária 1 Data: **18/03/09**
Avoidance: **0,000**
- Trajetoária 2 Data: **20/03/09**
Avoidance: **0,011**

Figura 5.21: Janela com o resultado do padrão aplicado.

Na Figura 5.21, com exceção do campo “N° Divisões na Área de Pastagem”, o cabeçalho

da tela corresponde aos parâmetros selecionados na janela da figura anterior (Figura 5.18). Essas divisões são criadas através da função *makegrid* (função criada e adicionada ao PostGIS), onde um conjunto de pontos são dispostos na área de pastejo no formato de um *grid*. Para cada ponto criado, um *buffer* circular é gerado através da função *ST_BUFFER* do próprio PostGIS. É desta maneira que a divisão comentada em 5.2.2.4 é realizada.

O resultado mostra as trajetórias usadas no algoritmo e a taxa de desvios que cada uma apresentou. O índice apresenta uma variação de 0 a 1, onde o aumento da taxa indica que a trajetória realizou desvios. No caso da trajetória 1, o índice foi de 0 indicando que não houve a realização de desvios. Já na segunda trajetória, o valor indica uma baixa ocorrência de desvio na trajetória. Outra forma de resultado pode ser conferida através do botão **Ver Áreas**, onde uma resposta visual em mapa é apresentada (Figura 5.22).

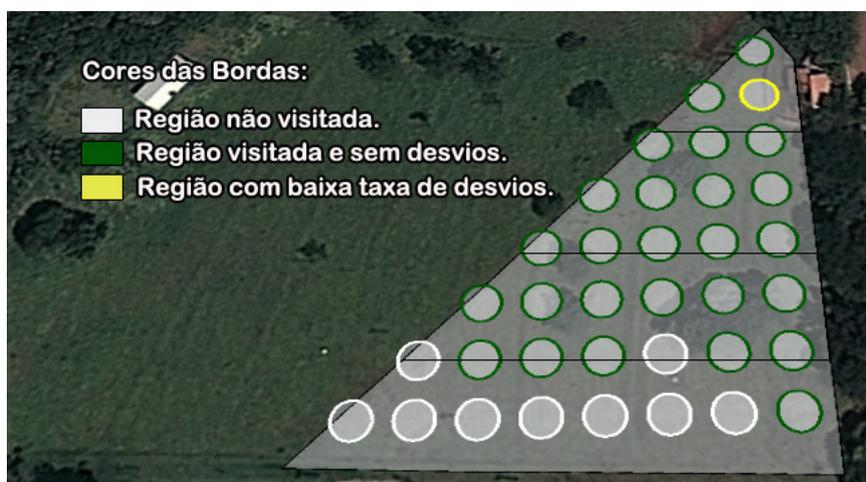


Figura 5.22: Áreas evitadas pelas trajetórias em mapa.

Na Figura 5.22, o ambiente de pastagem coberto pelas áreas circulares que representam uma determinada região da área de pastagem. Assim como as trajetórias apresentam um índice que indica a ocorrência de desvio, cada área possui também um índice próprio, que varia de -1 a 1. As cores das bordas dos círculos indicam o valor da taxa referente a cada área.

Por exemplo, quando o valor é igual a 0 (**cor verde**), isso indica que as trajetórias atravessaram a região sem qualquer desvio realizado. O índice compreendido no intervalo $0 < i \leq 0.5$ (**cor amarela**), onde i é o valor do índice, indica uma baixa presença de desvios. Já o valor compreendido entre $0.5 < i \leq 1$ (**cor vermelha**) indica uma alta presença de desvios. Por fim, o valor -1 (**cor branca**) indica que a área não foi visitada nenhuma vez pelas trajetórias.

É válido ressaltar que a presença de áreas não visitadas não consta no índice de desvios das trajetórias, uma vez que a não visitaç o n o configura um desvio, apesar de ser uma forma de se evitar.

5.3.2.10 Visualização de Trajetórias

Devido a presença de uma ferramenta de manipulação de mapas, usuário pode através do SATB observar a trajetória de um animal em um mapa, contendo a ordem e o período temporal em que cada coordenada foi coletada. A Figura 5.23 apresenta a tela principal da seção Observar Trajetória, onde o usuário pode visualizar a trajetória de um animal através dos dados armazenados no Banco de Dados ou através de um arquivo KML ou KMZ (uma versão compactada do KML).

A interface 'Observar Trajetórias' apresenta os seguintes elementos:

- Dados de Entrada:** Radio buttons para 'Arquivo KML e KMZ' e 'Banco de Dados' (selecionado).
- N° Trajetórias:** Campo de entrada com o valor '1'.
- 1° Colar:** Campo de entrada com o valor '0007'.
- 1° Data:** Campo de entrada com o valor '15/08/13'.
- Horário:** Radio buttons para 'Nenhum' (selecionado), 'Intervalo', 'Antes' e 'Depois'.
- Botões:** 'Mostrar Rastro' (verde) e 'Cancelar' (vermelho).

Figura 5.23: Janela principal da seção Observar Trajetória.

A visualização das trajetórias presentes no banco de dados permite a escolha na geração das trajetórias em mapa. Com a(s) trajetória(s) escolhida(s), um KML é gerado a partir dos dados recuperados da base de dados, como especificado na Seção 5.2.2.6. Com o arquivo criado, ele é enviado ao *Google Earth* que o utiliza como parâmetro de entrada para sua execução. A Figura 5.24 mostra uma trajetória em mapa.

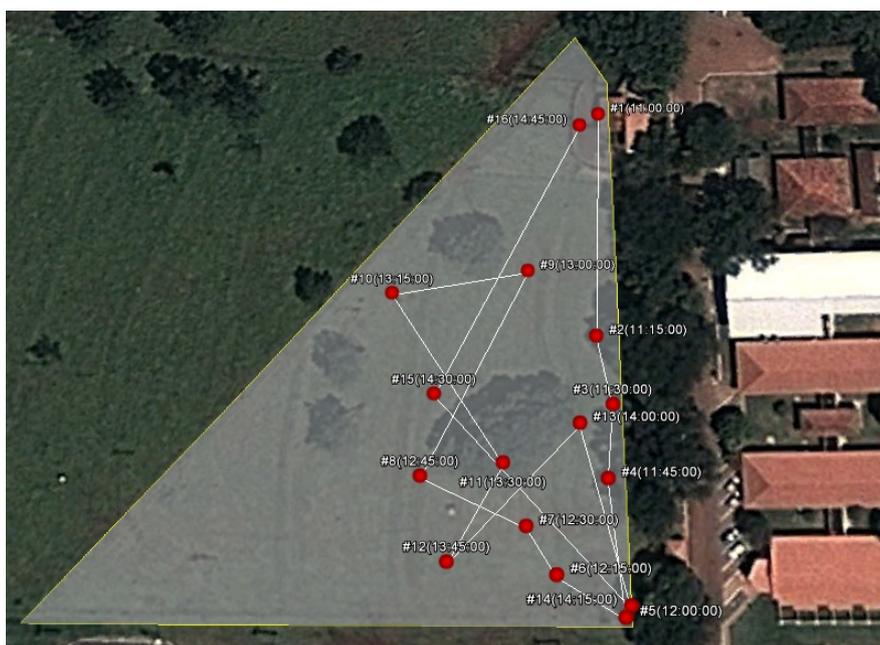


Figura 5.24: Uma trajetória animal representada em um mapa.

Caso a trajetória a ser visualizada seja obtida através de um arquivo KML ou KMZ, o sistema simplesmente envia o arquivo diretamente para o *Google Earth*, sem a necessidade de qualquer leitura ou qualquer acesso ao repositório.

5.4 Considerações Finais

Neste capítulo, o Sistema de Análise de Trajetórias Bovinos (SATB) foi apresentado e detalhado. A elaboração do sistema, as tecnologias envolvidas, as funcionalidades do sistema, bem como as teorias e as tecnologias relacionadas a essas funções, foram descritas e exemplificadas.

Apesar das possibilidades oferecidas pelas metodologias de análise de trajetórias, os dados espaço-temporais sozinhos não são capazes de oferecer uma completa inferência do comportamento animal. Segundo estudos realizados por Ganskopp [80] e Ungar *et al.* [81], os dados de trajetórias sozinhos não são capazes de inferir com precisão atividades bovinas, como por exemplo, o momento em que o animal está “andando”, “pastejando”, “descansando” e até mesmo “ruminando”.

Ainda no estudo de Ungar *et al.*, tais predições são possíveis através da união entre sensores de posicionamento (como o GPS) e sensores de movimento (*motion sensor*), uma vez que os dados de movimento conferem uma precisão muito maior as predições. Posteriormente, Einemo [82] conseguiu inferir o momento em que o animal está ruminando graças a essa união de tecnologias.

Capítulo 6

Estudo de Caso

Neste capítulo é apresentado um estudo de caso com a finalidade de validar este projeto de pesquisa. O estudo consiste na aplicação das metodologias de trajetórias semânticas sobre conjuntos de trajetórias de bovinos com o objetivo de extrair conhecimento relevante dos dados e assim demonstrando a eficiência e a eficácia das técnicas na pecuária. As trajetórias de animais foram obtidas de duas formas: um estudo empírico e um simulado.

6.1 Definição de Estudo

O Estudo de Caso deste projeto baseia-se na aplicação das metodologias de trajetórias semânticas sobre conjuntos de dados de movimentação de bovinos. As técnicas escolhidas para o estudo permitem verificar o período em que cada animal permaneceu em uma determinada região (**IB-SMoT**), analisar a localização das regiões de concentração dos animais em uma área de pastagem (**CB-SMoT** e **DBSCAN**) e, finalmente, encontrar as áreas de um ambiente de pastagem que foram evitadas pelos animais (**Avoidance**). Através do uso das três metodologias, pretende-se avaliar o comportamento dos animais em relação ao ambiente onde as trajetórias ocorreram.

As trajetórias, compostas por pontos georreferenciados, foram obtidas de duas formas: um estudo empírico e um simulado. No estudo empírico, colares da OTAG foram usados para coletar o posicionamento de bovinos em campo. A coleta foi realizada na Embrapa - Gado de Corte, em um ambiente de pastagem pertencente ao Mangueiro Digital. No estudo simulado, o deslocamento animal sobre a área de pastagem foi gerado através de um simulador de trajetórias. O uso de dados gerados por simulação foi necessário, pois o volume dos dados reais não era muito expressivo, além de boa parte deles se encontrar fora do cenário de teste, ou seja, o ambiente de pastejo.

A Tabela 6.1 apresenta quatro sub-tabelas referentes aos dados de posicionamento coletados pelos colares do projeto OTAG (seção 5.2.1.1), identificados como 0001, 0003, 0004 e 0006. Entre os parâmetros destacados nas tabelas, estão a data da coleta das coordenadas, o número de pontos que se encontravam dentro da área de pastagem e os pontos que se encontravam fora do ambiente em questão.

Colar 0001 - 67 pontos		
Data de Coleta	Pontos Dentro	Pontos Fora
19/03/09	30	10
20/03/09	2	3
01/04/09	8	12
13/05/09	0	2
Total	40	27

(a) Dados referentes ao colar 0001.

Colar 0003 - 64 pontos		
Data de Coleta	Pontos Dentro	Pontos Fora
17/03/09	6	15
18/03/09	22	8
19/03/09	3	2
29/04/09	0	5
13/05/09	0	3
Total	31	33

(b) Dados referentes ao colar 0003.

Colar 0004 - 279 pontos		
Data de Coleta	Pontos Dentro	Pontos Fora
16/03/09	11	11
17/03/09	8	12
18/03/09	48	14
19/03/09	1	5
20/03/09	35	9
13/05/09	1	30
14/05/09	0	19
26/05/09	4	19
27/05/09	13	39
Total	121	158

(c) Dados referentes ao colar 0004.

Colar 0006 - 271 pontos		
Data de Coleta	Pontos Dentro	Pontos Fora
17/03/09	1	8
18/03/09	39	26
19/03/09	21	23
20/03/09	2	8
01/04/09	2	37
02/04/09	0	96
13/05/09	0	8
Total	65	206

(d) Dados referentes ao colar 0006.

Tabela 6.1: Relação dos dados de posicionamento coletados em campo.

Conforme visto na tabela acima, as coletas foram realizadas durante os meses de março, abril e maio de 2009. Durante todo o projeto de pesquisa, problemas de natureza técnica - colares sem bateria, demora na compra da bateria específica, problemas de voltagem que atrapalhavam a comunicação com a estação base - inviabilizaram a coleta de novos dados reais. Aparentemente, todos esses problemas foram finalmente sanados. Entretanto, um novo experimento com dados mais atuais só foi realizado no fim de agosto de 2013.

Em relação aos dados de 2009, foram coletados um total de 681 pontos, sendo que apenas 257 pontos (37,7% do total) se encontravam nas dependências da área de pastagem do Mangueiro Digital. Observando ainda a pontuação diária, há dias onde a quantidade de pontos válidos no ambiente é desprezível para qualquer análise. Portanto, devido ao baixo número de dados reais aproveitáveis, decidiu-se pela utilização de dados gerados por simulação para complementar este estudo de caso.

A explicação para esse índice de pontos válidos é simples: todos os pontos foram coletados durante a fase de testes da primeira versão dos colares [83]. Os testes realizados com os colares não se restringiam a animais, mas também a pesquisadores que carregavam o equipamento fora das dependências do ambiente de pastejo. Além disso, a precisão do GPS também contribuiu para que muitos pontos ficassem afastados da área de teste.

Para a realização do estudo empírico e também do simulado, a área de pastagem utilizada foi a do Mangueiro Digital, localizada nas dependências da Embrapa - Gado de Corte de Campo Grande/MS. A Figura 6.1 mostra a área onde os estudos foram realizados, com destaque a seus elementos georreferenciados como as árvores, o mangueiro e o bebedouro.



Figura 6.1: A Área de Pastagem do Mangueiro Digital

É válido ressaltar que durante este capítulo, uma trajetória específica será identificada através da união entre a identificação do colar em que foi coletada e o dia de sua coleta. Por exemplo, “0001 - 19/03/09” é uma trajetória do colar 0001 em que a coleta foi realizada no dia 19/03/09.

6.1.1 Uso de Dados Reais

Como pode ser observado na tabela anterior (6.1), na maioria das trajetórias apresentadas, o número de coordenadas que está dentro dos limites da área de pastagem não é uma quantidade adequada para a realização de uma boa análise. Logo, para o aproveitamento desses pontos no experimento, foi necessário realizar uma seleção das trajetórias com uma quantidade de pontos significativa.

Para isso, um dos critérios estabelecidos da seleção era que uma trajetória deveria ao menos contar com 20 pontos dentro dos limites do Mangueiro Digital. Com isso, seis trajetórias acabaram sendo escolhidas, conforme observado na Tabela 6.2. A tabela apresenta para cada trajetória, características como a distância percorrida em metros, o tempo em que ela se inicia, o tempo em que ela se encerra e a duração de sua movimentação. As trajetórias escolhidas correspondem a 195 pontos (75,8%) dos 257 pontos válidos.

Colar	Data	Dist. Percorrida	Início	Fim	Duração
0001	19/03/09	1090.9 m	13:56:11	16:46:08	02:49:57
0003	18/03/09	969.24 m	11:56:33	14:16:07	02:19:34
0004	18/03/09	1455.9 m	12:20:43	17:20:33	04:59:50
	20/03/09	763.81 m	11:45:27	15:20:22	03:34:55
0006	18/03/09	1766.85 m	12:30:29	17:00:37	04:30:08
	19/03/09	729.37 m	12:30:28	15:00:26	02:29:58

Tabela 6.2: Trajetórias reais selecionadas pelo critério estabelecido.

Ao visualizar as trajetórias escolhidas em um mapa (fornecido pelo *Google Earth*), é possível perceber que alguns pontos desconsiderados estão a uma distância bem próxima dos limites do ambiente. Levando em conta a precisão do aparelho de GPS, algumas condições foram impostas para que esses pontos mais próximos fossem deslocados para dentro dos limites da área de pastagem:

- O ponto precisa estar a uma distância menor ou igual a 5 metros dos limites da pastagem.
- Caso um ou mais pontos pertençam a uma subtrajetória localizada fora do ambiente de pastagem, as coordenadas precisam estar localizadas ao final da subtrajetória.

Ao final desse processo, mais 20 pontos foram adicionados totalizando 215 pontos (83,6% dos pontos válidos). A Tabela 6.3 relaciona os pontos adicionados para cada trajetória.

Colar	Data	Pts. Adic.	Total
0001	19/03/09	5	35
0003	18/03/09	2	24
0004	18/03/09	5	53
	20/03/09	5	40
0006	18/03/09	1	40
	19/03/09	3	24

Tabela 6.3: Relação final dos pontos usados no Estudo de Caso.

É válido ressaltar que todos os pontos que estavam fora da área de pastagem foram desconsiderados na aplicação do estudo.

6.1.2 Ambiente Simulado

Como já explicado anteriormente, o utilização de dados gerados por simulação serviu como complemento ao experimento, uma vez que o número de dados reais aproveitáveis não era muito alto. Mas antes de qualquer coisa, era necessária a elaboração de um simulador de trajetórias a fim fornecer esse tipo de coordenadas.

Para a implementação da simulação, foi utilizado o algoritmo *Random Walk*, utilizado em Simulador de Deslocamento de Rebanho Bovino proposto por Arce *et al.* [84]. O algoritmo descreve a trajetória do animal utilizando um simples procedimento de seleção aleatória de pontos. O *Random Walk* segue basicamente as seguintes regras:

- Existe um ponto de partida;
- A direção de um ponto no caminho ao próximo é escolhida de forma randômica e nenhuma direção é mais provável do que outra.

Uma adaptação foi feita em relação ao algoritmo utilizado em Arce *et al.* [84]. Enquanto na estrutura original a distância entre dois pontos é um valor constante, nesta versão a distância varia dentro de um intervalo numérico passado como parâmetro. O cálculo para descoberta do próximo ponto da trajetória é apresentado na Figura 6.2. As equações seguem as regras apresentadas logo acima.

$$\begin{aligned}x_n &= x_{(n-1)} + d * \text{sen}(\theta); \\y_n &= y_{(n-1)} + d * \text{cos}(\theta);\end{aligned}$$

onde :

(x_n, y_n) = Coordenadas do próximo ponto.

θ = Ângulo gerado de forma randômica (0 – 360 graus).

d = distância percorrida.

Figura 6.2: Equação para a seleção do próximo ponto.

Originalmente a variável “ d ”, ou seja, o tamanho do passo do bovino, era definido na configuração do algoritmo. Como a distância aqui não será um valor constante, o “ d ” possuirá um valor randômico dentro de um intervalo estabelecido pelo usuário. Essa adaptação de um valor variado no lugar de uma constante visa beneficiar a localização de regiões de concentração dos animais no ambiente de pastejo, isso porque um dos passos do processo é a procura pelas regiões de baixa velocidade, algo que não seria possível de se encontrar caso todas as distâncias entre pontos fossem constantes.

O simulador de trajetórias foi incorporado ao SATB, que apresenta uma seção exclusiva, na qual é possível gerar uma trajetória na área de pastagem em que o usuário desejar, desde que a pastagem esteja cadastrada no sistema. A Figura 6.3 apresenta a janela principal do simulador de trajetórias.



Figura 6.3: O Simulador de Trajetórias do SATB.

Para a simulação de uma trajetória bovina, o usuário precisa informar todos os parâmetros apresentados na Figura 6.3. Inicialmente é necessário informar a qual corlar a trajetória será associada e em qual área de pastagem a trajetória ocorrerá. Em seguida, é necessário informar a data e a hora em que o primeiro ponto foi gerado; e o intervalo de tempo em que cada ponto será gerado. Por fim, é necessário informar quantos pontos serão criados, o intervalo de valores de distância que o animal poderá percorrer (a

distância entre cada ponto será um valor do intervalo escolhido de forma randômica) e as coordenadas do primeiro ponto.

Existem duas maneiras de informar as coordenadas do primeiro ponto. Uma delas é preenchendo os campos de latitude e longitude com os valores do ponto em questão. A outra maneira é utilizando o botão “*Ponto Inicial Aleatório*”, onde um ponto presente nos limites da pastagem escolhida é selecionado de forma aleatória. Uma consideração importante a se fazer é de que o simulador possui um método de validação de pontos. Um ponto gerado é válido quando este se encontra localizado dentro dos limites da área de pastagem escolhida. Caso ele esteja fora, ele é descartado e um novo ponto é gerado.

Para o estudo, foram gerados quatro cenários diferentes, representados pelos colares 0009, 0010, 0011 e 0012. Para a criação desses colares, uma seção de gerenciamento para colares foi implementada no sistema. Cada cenário recebeu cinco trajetórias, onde cada deslocamento animal foi gerado em um dia distinto do intervalo 01/08/13 à 05/08/13. Cada trajetória consiste de 100 pontos, totalizando um total de 2000 pontos gerados. Todas as trajetórias criadas possuem pontos de partida distintos. Para colaborar na diferenciação dos caminhos produzidos, os cenários apresentam características próprias, como pode ser visto na Tabela 6.4.

Colar	Duração	Intervalo Dist.	Intervalo Temp.	Movimentação
0009	08:15:00	Entre 2 e 15 metros	5 minutos	Rápida
0010	16:30:00	Entre 2 e 15 metros	10 minutos	Normal
0011	08:15:00	Entre 2 e 7.5 metros	5 minutos	Normal
0012	16:30:00	Entre 2 e 7.5 metros	10 minutos	Lenta

Tabela 6.4: Diferença entre os quatro conjuntos de trajetórias.

A diferenciação nos parâmetros de deslocamento e de tempo não foi por acaso, uma vez que grande parte das variações nos resultados dos algoritmos é devido a esses parâmetros. Uma variação de deslocamento pode, por exemplo, possibilitar ou não o deslocamento de uma trajetória até certos pontos mais afastados do ambiente. Em relação ao tempo, por exemplo, irá influenciar na procura por regiões de baixa velocidade. A variação do tempo ocasionalmente mudará o tipo de conjunto construído pelo algoritmo CB-SMoT (Seção 3.2.2.2).

As variações espaciais e temporais também influenciam no tipo de movimentação animal. Em 0009, o alto deslocamento em um tempo menor gera uma movimentação mais rápida. Em 0010, um alto deslocamento em um tempo maior gera uma movimentação normal. Em 0011, um baixo deslocamento em um tempo menor também gera uma movimentação normal. Por fim, um baixo deslocamento e um tempo maior gera em 0012 uma movimentação lenta. Outra importante observação sobre a Tabela 6.4 é sobre a duração das trajetórias de cada cenário. Como cada cenário possui trajetórias com o mesmo número de pontos e o mesmo intervalo temporal entre cada ponto, logo os caminhos pertencentes aos cenários possuíram a mesma duração de tempo.

6.2 Técnicas de Trajetórias Semânticas em Dados Reais e Simulados

Este estudo de caso consiste na aplicação das técnicas de trajetória semântica sobre os conjuntos de dados reais e gerados pelo simulador. As técnicas utilizadas pretendem verificar o período em que cada animal permaneceu em uma determinada região, analisar a localização das regiões de concentração dos animais na pastagem e, finalmente, encontrar as áreas do ambiente de pastejo que foram evitadas pelos animais. Cada seção apresenta os resultados obtidos a partir da aplicação de uma técnica diferente.

6.2.1 Verificar o período em que cada animal permaneceu em uma determinada região

Através do uso do **IB-SMoT** (Seção 3.2.2.1) sobre as trajetórias, é possível saber **onde**, **quando** e **por quanto tempo** um animal esteve em uma determinada área fornecida pelo usuário. Com isso, a primeira etapa deste estudo de caso com dados reais consiste em averiguar o período de tempo (em relação a duração das trajetórias) em que os animais permaneceram em uma determinada região da área de pastagem.

Para isso, a área de pastejo do Mangueiro Digital foi dividida em quatro áreas de interesse. Nomeadas como *Área Norte*, *Área Centro-Norte*, *Área Centro-Sul* e *Área Sul*, as áreas representaram os objetos espaciais que serão usados pelo algoritmo como os candidatos à parada (*candidates stop*). A Figura 6.4 mostra a disposição dessas áreas de interesse no ambiente do Mangueiro Digital.



Figura 6.4: A área de pastagem do Mangueiro Digital dividida em quatro áreas de interesse.

Cada área gerada possui suas próprias características. A *Área Norte* é a menor área

e a que possui a estrutura do mangueiro em suas dependências. É no mangueiro onde os colares são colocados e retirados dos animais. A *Área Centro-Norte* está a uma distância relativamente pequena do mangueiro e possui um pequeno número de árvores em seus limites. A *Área Centro-Sul* é a área com o maior número de árvores e conseqüentemente, a região com a maior quantidade de áreas de sombreamento. Por fim, na *Área Sul* encontra-se o bebedouro e o cocho de sal. É a maior área das quatro e possui pouquíssimas árvores, que se concentram perto do bebedouro.

6.2.1.1 Resultados Obtidos com Dados Reais

Os resultados da aplicação da metodologia sobre os dados reais de trajetórias bovinas estão expressos na Tabela 6.5. A tabela apresenta a duração de cada trajetória e a porcentagem de tempo em que o animal permaneceu em uma das quatro áreas durante sua movimentação. Para exemplificar, seja a trajetória “0001 - 19/03/09”. A duração total da trajetória foi de 02:49:57 e o tempo em que o animal desta permaneceu na *Área Norte* foi de 01:05:39. Logo, em relação ao tempo total, o animal permaneceu 38,63% do tempo neste local.

Colar	Data	Duração	Área Norte	Área C. Norte	Área C. Sul	Área Sul
0001	19/03/09	02:49:57	38,63%	49,62%	11,67%	3%
0003	18/03/09	02:19:34	34,68%	22,62%	18,96%	27,05%
0004	18/03/09	04:59:50	21,80%	36,31%	30,40%	13,10%
	20/03/09	03:34:55	55,59%	27,71%	18,69%	2,32%
0006	18/03/09	04:30:08	29,68%	36,31%	40,63%	10,88%
	19/03/09	02:29:58	66,49%	6,40%	6,85%	20,25%

Tabela 6.5: Dados Reais - Período de tempo (em %) em cada trajetória permaneceu em cada área de interesse do ambiente de pastagem.

Como pode ser conferido, a maioria dos animais apresentou uma tendência em permanecer ao menos 20% de seu tempo na *Área Norte* do Mangueiro Digital. O maior índice de permanência nessa área foi registrado no collar 0006 na trajetória do dia 19/03/09. O animal do collar 0004 apresentou o menor índice de permanência na trajetória do dia 18/03/09.

Na *Área Centro-Norte*, com exceção da trajetória “0006 - 19/03/09”, os animais apresentaram uma tendência em permanecer 22% de seu tempo na região. A mesma trajetória pode ser tratada como uma exceção na *Área Centro-Sul*, onde os animais apresentaram uma tendência em permanecer 11% de seu tempo. A *Área Sul* foi a região com os menores índices de permanência. Com exceção das trajetórias “0006 - 19/03/09” e da “0003 - 19/03/09” que apresentaram porcentagens maiores que 20%, as outras permaneceram dentro de um intervalo de 2% a 14%.

Tendo conhecimento das quatro áreas e analisando os resultados acima, as seguintes inferências podem ser tiradas:

- O mangueiro está localizado na *Área Norte*. Como os colares são colocados e retirados nesse lugar, os animais tem uma tendência em permanecer nas imediações da

construção no início e no fim de sua movimentação.

- A *Área Centro-Norte* apresenta uma área de pastagem maior que a *Área Norte* e fica bem próximo do mangueiro. Com alimentação disponível e algumas fontes de sombra, os bovinos tendem a permanecer um bom tempo nessa região para pastar.
- Mesmo possuindo uma área maior para o pastejo e uma grande quantidade de áreas de sombreamento, a grande distância entre a *Área Centro-Sul* e a *Área Norte* (onde os animais iniciam sua interação com o ambiente) acaba afetando o tempo de permanência dos bovinos na *Área Centro-Sul*, ficando bem abaixo dos anteriores.
- Por se tratar da região mais distante, a *Área Sul* acaba sendo a região de menor índice de permanência dos bovinos. O que provavelmente deve garantir a vinda dos animais até aqui é a presença do bebedouro e do cocho.

Vale ressaltar que com a aplicação das outras técnicas do estudo, as inferências realizadas até então poderão ser confirmadas ou desmentidas, além claro, do surgimento de novas inferências.

6.2.1.2 Resultados Obtidos com Dados Simulados

Aplicando agora a mesma metodologia para o conjunto de trajetórias simuladas, os resultados obtidos podem ser conferidos na Tabela 6.6.

Colar	Mov.	Data	Duração	Área Norte	Área C. Norte	Área C. Sul	Área Sul
0009	Rápida	01/08/13	08:15:00	0%	4,04%	62,63%	33,33%
		02/08/13		9,09%	27,27%	46,46%	17,17%
		03/08/13		16,16%	45,45%	39,39%	0%
		04/08/13		1,01%	35,35%	54,55%	10,10%
		05/08/13		0%	43,43%	32,32%	25,25%
0010	Normal	01/08/13	16:30:00	4,04%	35,35%	30,30%	31,31%
		02/08/13		0%	18,18%	39,39%	43,43%
		03/08/13		32,32%	31,31%	11,11%	26,26%
		04/08/13		31,31%	9,09%	30,30%	30,30%
		05/08/13		23,23%	48,48%	26,26%	3,03%
0011	Normal	01/08/13	08:15:00	0%	0%	51,52%	49,49%
		02/08/13		0%	55,56%	44,44%	0%
		03/08/13		0%	15,15%	50,51%	35,35%
		04/08/13		48,48%	52,55%	0%	0%
		05/08/13		0%	8,08%	25,25%	67,68%
0012	Lenta	01/08/13	16:30:00	0%	62,63%	38,38%	0%
		02/08/13		0%	0%	23,23%	77,78%
		03/08/13		0%	93,94%	7,07%	0%
		04/08/13		0%	55,56%	23,23%	22,22%
		05/08/13		0%	0%	42,42%	58,59%

Tabela 6.6: Dados Simulados - Período de tempo (em %) em cada trajetória permaneceu em cada área de interesse do ambiente de pastagem.

Como pode ser observado, o período de permanência na *Área Norte* teve uma queda acentuada em comparação aos resultados obtidos a partir de dados reais. Como os dados

simulados possuem como ponto inicial uma coordenada qualquer do ambiente de pastejo, muitas trajetórias acabaram se iniciando a uma distância considerável do mangueiro. Logo, isso explica o índice tão baixo na *Área Norte*.

Ao observar as trajetórias onde o deslocamento varia entre 2 m até 7.5 m (colares 0011 e 0012), a primeira constatação é que o animal percorre no máximo três regiões diferentes do ambiente. Não existiu na simulação casos onde um animal passou todo seu tempo em uma única região. Nas trajetórias onde o deslocamento varia de 2 m até 15 m (colares 0009 e 0010), o animal já consegue percorrer todas as áreas do ambiente. Em nenhum dos exemplos gerados, o animal permaneceu em apenas uma ou duas regiões.

Enquanto nas trajetórias de dados reais os animais tiveram uma preferência em permanecer nas *Área Norte* e *Área Centro-Norte*, na simulação essa preferência ficou por conta da *Área Centro-Norte* e da *Área Centro-Sul*. Logo, os animais tiveram uma preferência em se estabelecer na parte central do ambiente.

Por fim, levando em conta o alto índice obtido pela *Área Sul* sobre a *Área Norte*, enquanto não estão circulando na região central, os animais tem mais tendência a migrarem para a porção sul da área de pastagem do que a norte.

6.2.2 Localizar Regiões de Concentração na Área de Pastagem

Após verificar o tempo que cada animal permaneceu em uma das quatro áreas do ambiente de pastagem, a segunda etapa deste estudo foca a localização das regiões de concentração, isto é, os lugares na área de pastagem onde os animais possuem uma maior preferência em ficar. Por exemplo, o animal da trajetória “0001 - 19/03/09” pode ter ficado mais tempo na *Área Norte* do Mangueiro Digital, mas enquanto esteve nessa região, onde ele se concentrou?

Para descobrir essas regiões, o primeiro passo é encontrar os conjuntos de pontos de baixa velocidade de cada trajetória, através da primeira etapa do algoritmo **CB-SMoT** (Seção 3.2.2.2). Conjuntos de baixa velocidade são subtrajetórias onde a velocidade entre os pontos é baixa e, dessa forma, fazendo com que os pontos estejam mais próximos, e assim formando um conjunto de pontos mais denso.

Tendo aplicado isso a todas as trajetórias presentes nos colares, o segundo passo é aplicar um algoritmo de agrupamento nos conjuntos de baixa velocidade. No caso deste estudo, o algoritmo será o **DBSCAN** (Seção 3.2.2.2), metodologia que conta com dois parâmetros: um relacionado ao número mínimo de pontos que um conjunto precisa possuir e outro que é a distância mínima entre esses conjuntos.

6.2.2.1 Resultados Obtidos com Dados Reais

Aplicando os algoritmos primeiramente para as trajetórias reais, os valores dos parâmetros foram de 20 metros para a distância e 6 para o número mínimo de pontos em conjunto. Com isso, as regiões de concentração geradas para as trajetórias de cada colar podem ser

observadas na Figura 6.5.

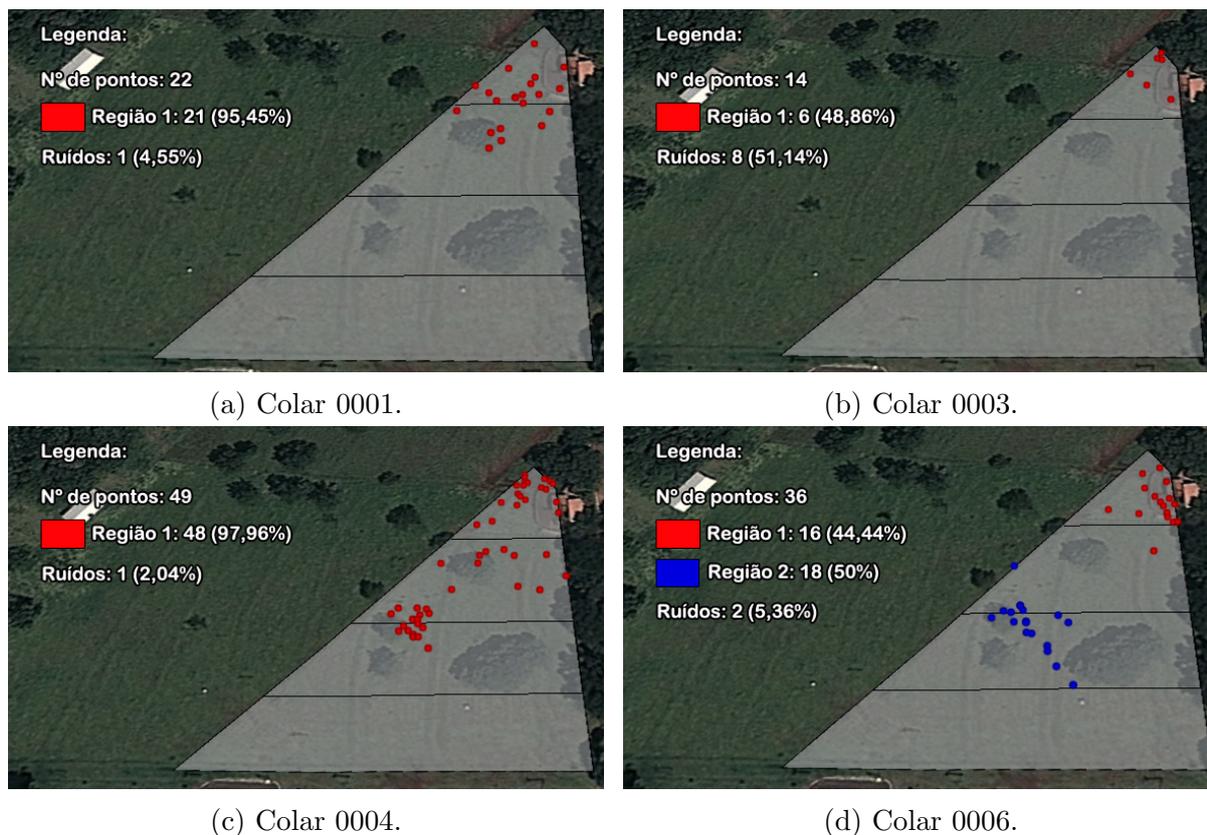


Figura 6.5: Regiões de concentração do estudo empírico (parâmetros: 20 e 6).

Apenas o colar 0006 obteve duas regiões de concentração, enquanto os colares 0001, 0003 e 0004 obtiveram uma única região. Isso pode ser explicado devido a grande distância entre os dois conjuntos de pontos de densidade do colar 0006, onde um agrupamento está localizado na *Área Norte* e o outro na *Área Centro-Sul*.

Os resultados presentes na Figura 6.5 já confirmam algumas informações discutidas durante o algoritmo da seção anterior. Ao observar a *Área Norte*, pode-se observar um grande número de pontos próximos da estrutura do mangueiro. Por ser uma das áreas onde os animais mais permaneceram, grande parte dos pontos de concentração se encontram perto das dependências do mangueiro, lugar onde os colares são colocados. Mesmo apresentando um alto tempo de permanência, há poucos pontos de concentração na *Área Centro-Norte*. E pelos poucos pontos apresentados no mapa, é bem provável que os animais tenham se deslocado mais pela área, sem parar em nenhuma porção específica, do que na *Área Norte*.

Os pontos de concentração da *Área Centro-Sul* mostraram uma preferência dos animais de 0004 e 0006 pela porção oeste da região. Além disso, grande parte dos pontos encontra-se próximos de árvores, mostrando que os animais poderiam estar em busca de sombra. Para a confirmação desta inferência, basta obter as áreas de sombra onde ocorreu uma intersecção com os subconjuntos de pontos de baixa velocidade dos colares 0004 e 0006 e comparar essas informações com os dados dos momentos onde as trajetórias dos mesmos

colares estiveram na sombra.

Inicialmente, a Tabela 6.7 apresenta duas tabelas: a primeira tabela (7a) mostra os conjuntos de pontos de baixa velocidade que interceptaram áreas de sombra (obtidas na Seção 5.3.2.8); já a segunda (7b) mostra as áreas de sombra interceptadas pelas trajetórias do colar (obtidas na Seção 5.3.2.5). Os parâmetros apresentados pela Tabela 6.7a consistem no dia da trajetória onde esses conjuntos foram gerados, o intervalo temporal da trajetória onde estão os conjuntos e a área de sombra interceptada. Na Tabela 6.7b, os parâmetros consistem no dia que a trajetória foi realizada, na área de sombra que a trajetória interceptou e o intervalo de tempo que ela permaneceu na sombra.

Subconj. de baixa velocidade na sombra		
Data	Intervalo Subconj.	Área de Sombra
18/03/09	17:10:27 - 17:20:33	sombra_tarde_arv001
	14:55:28 - 15:25:50	sombra_tardearv002
	15:35:31 - 16:55:28	sombra_tarde_arv3
20/03/09	não há subconjuntos na sombra	

(a) Áreas de Sombra atingidas pelos Subconjuntos de 0004.

Momentos do colar 0004 na sombra		
Data	Área Sombra	Esteve na área entre
18/03/09	sombra_tarde_arv001	13:15:39 - 13:21:04
	sombra_tarde_arv001	17:10:27 - 17:20:33
	sombra_tardearv002	15:15:41 - 15:20:33
	sombra_tarde_arv3	15:35:31 - 15:40:29
	sombra_tarde_arv3	15:45:29 - 15:50:29
	sombra_tarde_arv3	15:56:02 - 16:20:28
20/03/09	sombra_tarde_arv3	14:55:22 - 15:05:24
	sombra_tardearv002	13:35:34 - 13:40:24

(b) Momentos onde o animal do colar 0004 esteve na sombra

Tabela 6.7: Relação de pontos em área de sombra do colar 0004

Observando as áreas de sombra das duas tabelas, pelo nome dessas áreas, pode-se verificar que o animal do colar 0004 só esteve na sombra durante o período vespertino. Consultando a Tabela 6.2, observa-se que todas as trajetórias reais começam ou no início da tarde ou perto do meio dia. Logo, as áreas de sombra encontradas pelos animais ou estarão embaixo das árvores ou estarão ao leste delas.

Comparando a Figura 6.1 com a 6.7, a região de concentração de pontos está ao leste da *Árvore 3*, lugar onde as tabelas indicam a presença animal na região de sombra criada pela árvore durante a tarde. Como os pontos de concentração são formados por subconjuntos de baixa velocidade, o subconjunto em 6.7a que intercepta a sombra vespertina da *Árvore 3* é conjunto de pontos compreendidos no intervalo “15:35:31 - 16:55:28”. Ao observar a Tabela em 6.7b, os momentos em que o animal esteve na área de *sombra sombra_tarde_arv3* coincidem com o intervalo do subconjunto e com a data. Logo, o animal de 0004 estava realmente a procura de sombra.

A Tabela 6.8 utiliza a mesma ideia apresentada na Tabela 6.7, mas dessa vez para os dados do colar 0006.

Subconj. de baixa velocidade na sombra		
Data	Intervalo Subconj.	Área Atingidas
18/03/09	14:45:27 - 14:55:31	sombra_tardearv002
	15:20:29 - 15:50:35	sombra_tarde_arv3
	16:01:06 - 17:00:37	sombra_tarde_arv3
19/03/09	não há subconjuntos na sombra	

(a) Áreas de Sombra atingidas pelos Subconjuntos de 0006.

Momentos do colar 0006 na sombra		
Data	Área Sombra	Esteve na área entre
18/03/09	sombra_tarde_arv2	14:50:29 - 14:55:31
	sombra_tarde_arv3	15:35:33 - 15:45:28
	sombra_tarde_arv3	16:09:06 - 16:16:29
	sombra_tarde_arv3	16:20:33 - 16:45:28
	sombra_tarde_arv3	16:50:28 - 16:55:28
19/03/09	sombra_tarde_arv001	13:05:50 - 13:11:04
	sombra_tarde_arv2	15:00:26 - 15:05:26
	sombra_tardearv004	14:50:27 - 14:55:29

(b) Momentos onde o animal do colar 0006 esteve na sombra

Tabela 6.8: Relação de pontos em área de sombra do colar 0006

Usando o mesmo raciocínio do caso anterior, observa-se na Figura 6.8 que as regiões de concentração do colar 0006 também se encontram ao leste da *Árvore 3*. Consultando a tabela, haverá indicações nas duas tabelas de pontos interceptando a região de sombreamento da árvore durante o mesmo dia no período vespertino. Comparando os subconjuntos de pontos compreendidos nos períodos “15:20:29 - 15:50:35” e “16:01:06 - 17:00:37” (Tabela 8a) com os momentos onde o animal esteve nessa área de sombra (Tabela 8b), os períodos de tempo onde o animal esteve na área de sombra *sombra_tarde_arv3* bate com os intervalos dos dois subconjuntos. Logo, o animal do colar 0006 também estava a procura de uma sombra.

Tanto o animal de 0004 quanto o de 0006 estavam realmente concentrados nas imediações de uma área de sombra. Comparando as Tabelas 6.7b e 6.8b observa-se que no dia 18/03/09, os dois animais estiveram no período “15:35:33 à 16:55:28” concentrados na mesma região de sombreamento. Portanto, ambos estavam juntos durante o período em que estiveram concentrados nas imediações da *Árvore 3*.

É válido ressaltar que os subconjuntos de pontos de baixa velocidade representam uma pequena parte dos pontos de uma trajetória. Logo, é normal que algumas sombras que a trajetória tenha interceptado (Tabelas 6.7b e 6.8b) não estejam na tabela dos subconjuntos (Tabelas 6.7a e 6.8a), uma vez que essas sombras foram interceptadas por pontos que não fazem parte desses conjuntos de concentração.

Por fim, nenhuma concentração foi encontrada na *Área Sul*, o que indica que o único interesse dos animais pelo lugar não é o capim desta região, mas sim o bebedouro. O resultado desta primeira aplicação do algoritmo ainda não atingiu o ideal, uma vez que ainda há grandes agrupamentos de pontos de concentração. Observando-se esses grandes conjuntos, há lugares onde a densidade dos pontos é mais alta do que em outras.

Como o objetivo é encontrar essas regiões de grande densidade, a solução é uma segunda aplicação da metodologia utilizando um valor menor para distância e um maior para o conjunto mínimo de pontos. Isso fará com que as regiões de baixa concentração sejam retiradas no próximo resultado, garantindo um refinamento na saída. Aplicando novamente o DBSCAN, dessa vez com uma distância de 15 metros e um grupo mínimo de 8 pontos, o resultado apresentado Figura 6.6 já demonstra uma grande diferença em relação ao anterior.

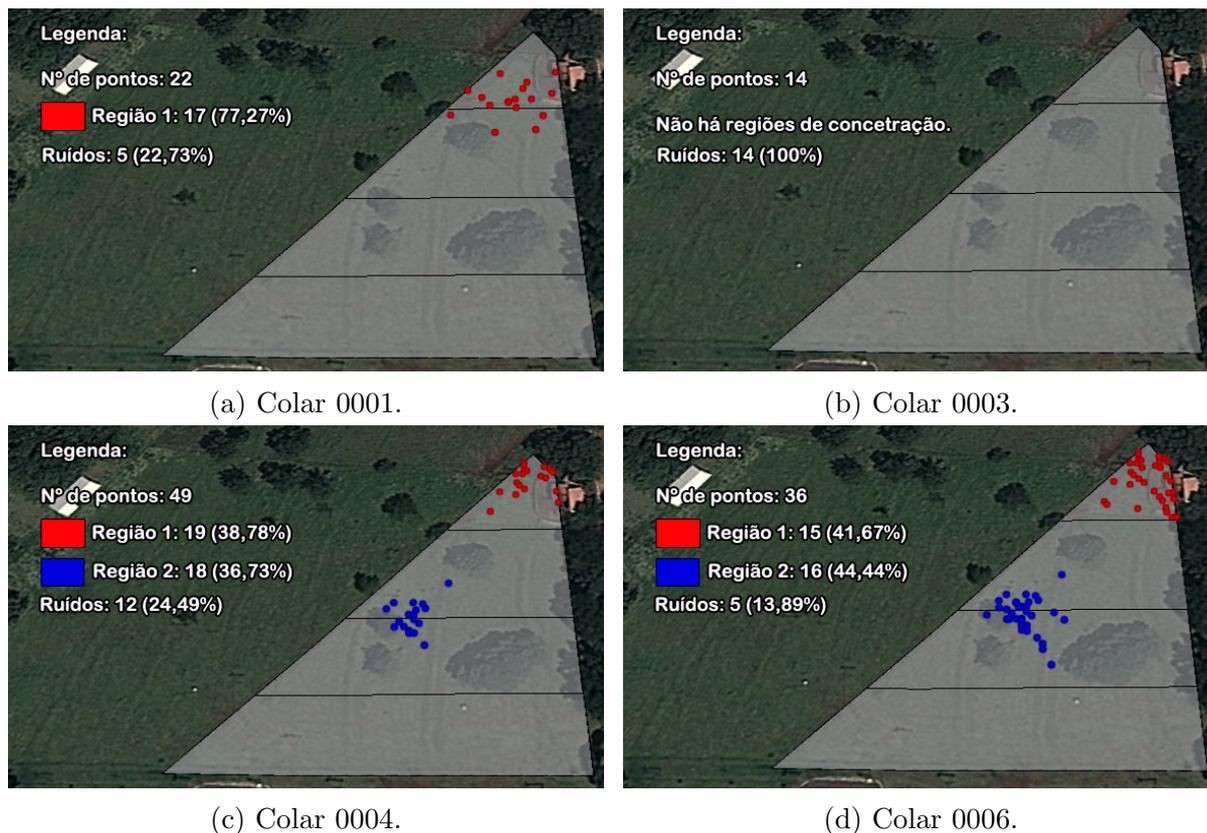


Figura 6.6: Regiões de concentração do estudo empírico (parâmetros: 15 e 8).

Comparando as duas imagens (6.5 e 6.6), é possível observar que parte dos pontos que se encontravam mais afastados dos agrupamentos foram removidos. Dois casos interessantes a se observar são em relação ao colar 0003 e ao 0004. No primeiro, o pequeno agrupamento localizado na *Área Norte* deixou de existir devido as novas exigências do algoritmo. Isso significa que não existem regiões de concentração para as trajetórias deste colar utilizando esses parâmetros.

No caso do colar 0004, o refinamento dos parâmetros acabou por retirar os pontos da *Área Centro-Norte* (considerados como ruídos) e por gerar a divisão do grande agrupamento da Figura 6.6 em dois conjuntos de pontos. Logo, o aumento na precisão do algoritmo resultou em duas regiões de concentração com uma densidade de pontos maior que a anterior.

Diminuindo a distância e aumentando o tamanho mínimo do conjunto de pontos, a tendência é a remoção dos pontos mais afastados e na divisão dos grandes agrupamentos, resultando apenas em região mais densas.

6.2.2.2 Resultados Obtidos com Dados Simulados

Chega a vez de aplicar os algoritmos para as trajetórias simuladas. Usando novamente como parâmetros os valores 20 (para a distância) e 6 (para o tamanho mínimo do conjunto), as regiões de concentração geradas para as trajetórias de cada colar podem ser

observadas na Figura 6.7.

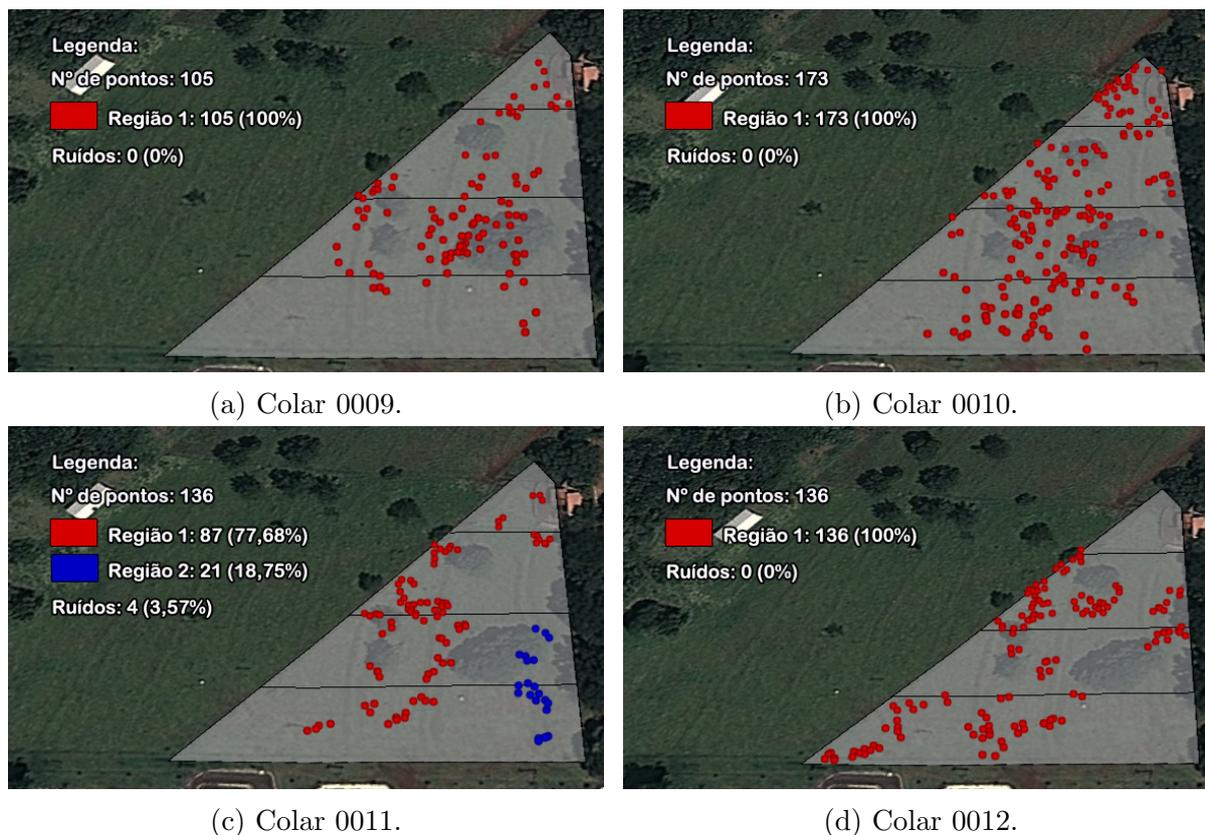


Figura 6.7: Regiões de concentração da simulação (parâmetros: 20 e 6).

A primeira impressão que se tem comparando o resultado dos dados simulados com os reais é que no simulado há regiões mais densas e maiores que nos reais. Isso pode ser explicado pelo fato de cada colar possuir nada menos que 500 pontos. Portanto, uma quantidade maior de pontos irá produzir regiões maiores do que um conjunto pequeno.

Com exceção do cenário apresentado pelo colar 0011, a aplicação do DBSCAN gerou através dos pontos de cada colar um grande e único conjunto de pontos espalhados pelo ambiente. Esses grandes conjuntos foram criados devido aos parâmetros de 20 metros e da simulação gerar pontos em intervalos até 15 metros. Como o valor espacial do algoritmo é muito mais elevado que o valor de distância entre os pontos, dificilmente haverá divisão de conjuntos e eliminação de ruídos.

O colar 0011 foi o único a apresentar duas regiões de concentração nesta primeira aplicação do algoritmo. Mesmo possuindo uma distância entre os pontos de no máximo 15 metros, parte dos pontos de baixa velocidade se concentrou na porção leste próximo da cerca enquanto outra parte dos pontos se concentrou próximo à cerca da porção oeste e em alguns pontos da porção central. Logo, a distância entre esses agrupamentos contribuiu para a geração de duas regiões de concentração.

Como poucos pontos foram eliminados no resultado final desta execução, é necessária uma segunda execução com uma mudança de parâmetros para assim refinar a próxima

saída. Desta forma, aplicando os parâmetros 15 para distância e 8 para o tamanho mínimo do conjunto, o resultado obtido é apresentado na Figura 6.8.

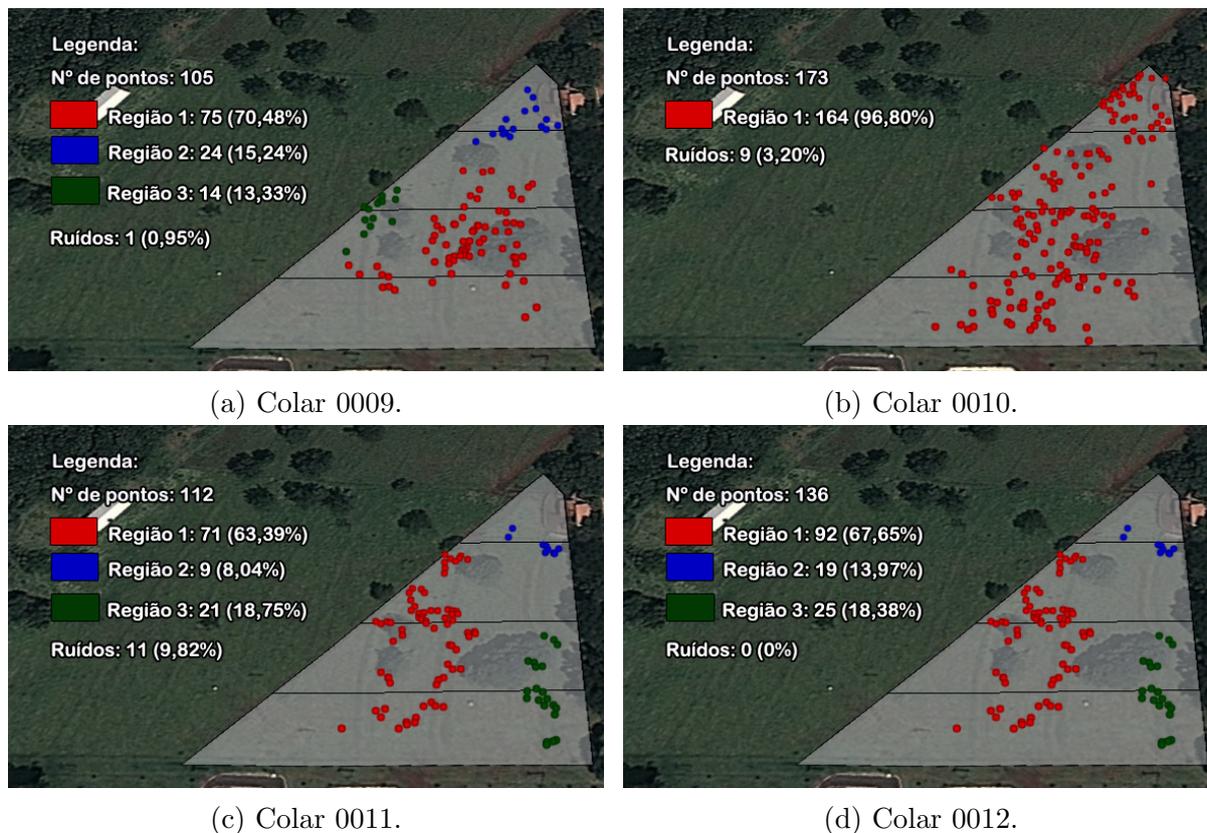


Figura 6.8: Regiões de concentração da simulação (parâmetros: 15 e 8).

Comparando as duas imagens (6.7 e 6.8), é possível observar que parte dos pontos que se encontravam mais afastados dos agrupamentos foram removidos. O aumento na precisão do algoritmo acabou por criar novas regiões, resultantes a partir da quebra dos grandes agrupamentos de pontos (Figura 6.7).

Entretanto, a exceção foi o colar 0010, onde a saída da aplicação do DBSCAN ainda gerou um grande conjunto de pontos de baixa velocidade espalhados pela área de pastejo. Uma vez que ainda há uma grande quantidade de pontos e extensos agrupamentos de pontos, essa condição permite que uma nova execução do algoritmo de agrupamento com um aumento na precisão dos parâmetros seja realizada. Portanto, usando o valor 10 para distância e 10 para o número mínimo de pontos em conjunto, a Figura 6.9 apresenta um resultado ainda mais refinado.

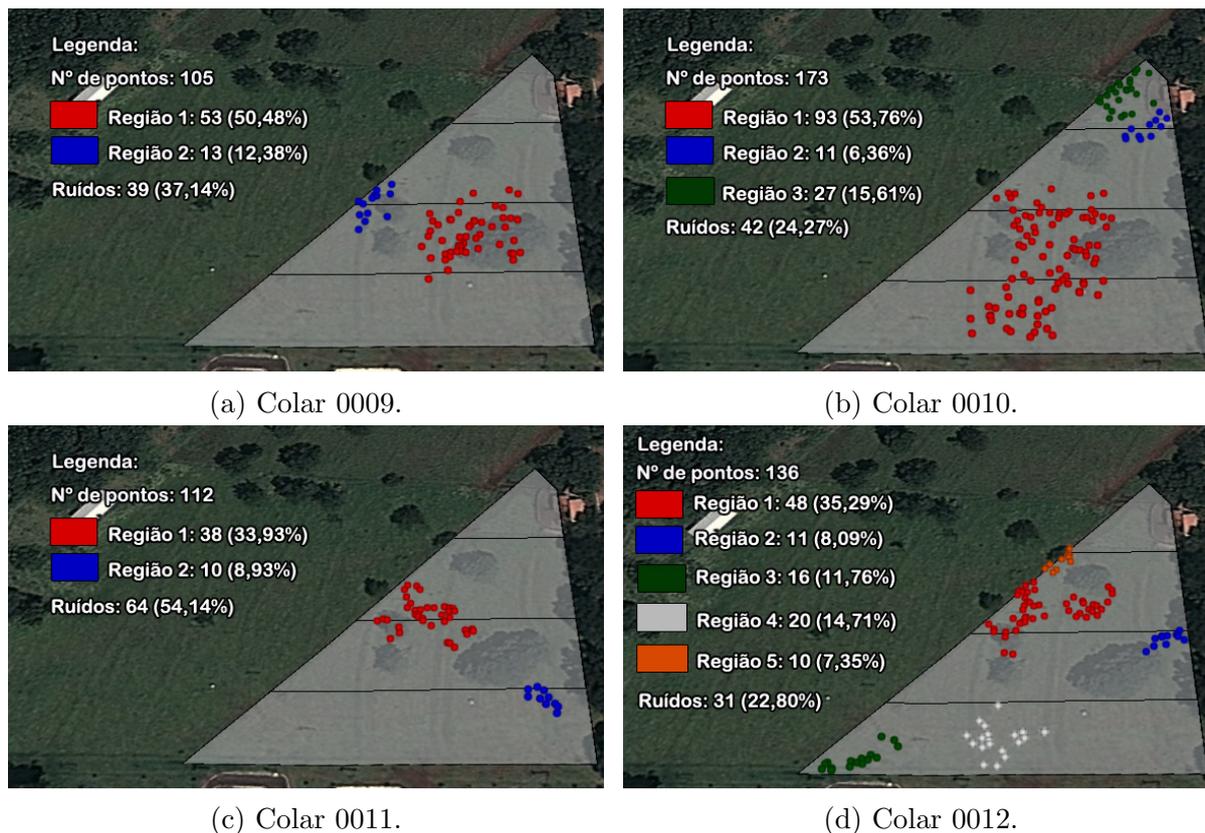


Figura 6.9: Regiões de concentração da simulação (parâmetros: 10 e 10).

Em comparação com os resultados anteriores, a Figura 6.9 apresenta um resultado muito mais expressivo com regiões onde os pontos apresentam uma grande densidade entre si. O colar 0010 que apresentava apenas um grande conjunto de pontos agora apresenta três conjuntos. Alguns agrupamentos das imagens anteriores foram limados nesse último resultado por não apresentarem a densidade requerida pelos parâmetros.

Se as informações de concentração apresentadas fossem retiradas do monitoramento de quatro animais distintos, algumas inferências já poderiam ser feitas. O animal do colar 0009, por exemplo, teve a preferência de se concentrar na *Área Centro-Sul*, próximo às árvores. Observando ainda resultados anteriores, o animal chegou a se concentrar perto da área do mangueiro, concentrando-se na divisão da *Área Norte* com a *Área Centro-Norte*. O animal do colar 0010 teve uma preferência em se concentrar na porção central da *Área Centro-Sul* e *Área Sul*. Como a *Área Sul* não possui árvores na porção ocupada pelo animal, o principal interesse ao ocupar a região não é a sombra, mas sim a pastagem. O animal inclusive foi o único a apresentar uma grande concentração na *Área Norte*.

Já o animal do colar 0011 foi o que apresentou o menor número de pontos em suas regiões de concentração. Assim como o animal do colar 0009, o animal se concentrou em porções do ambiente onde há árvores, o que pode indicar a preferência em pastar em locais onde possa existir áreas de sombra. Por fim, o animal do colar 0012 permaneceu em várias porções diferentes da área de pastagem. Na *Área Centro-Norte*, ele concentrou-se entre a porção central e a cerca ao oeste, local que apresenta algumas árvores. Na *Área Centro-Sul*, concentrou-se nas árvores da porção leste. E por fim, na *Área Sul* permaneceu na

região central e no extremo oeste do ambiente. Ao que tudo indica, apesar de não haver árvores, a *Área Sul* apresenta uma pastagem agradável aos animais.

6.2.3 Identificar Áreas Evitadas pelos Animais

Após descobrir o período em que cada animal passou em cada uma das áreas e quais foram os lugares no área de pastagem onde cada animal se concentrou, agora é a vez de identificar quais são as regiões em que os animais evitaram. Em nosso estudo, um animal evita uma certa região quando ele não visita o lugar e quando ele se desvia intencionalmente daquele lugar.

Como já explicado na Seção 5.2.2.4, cada porção de espaço do ambiente de pastagem é representado por uma circunferência. Todos os pontos centrais dos círculos são dispostos no ambiente como se fossem um *grid*. Quando o algoritmo é aplicado, o nível de *avoidance* de cada área é representado pela cor da borda de cada circunferência. Quando a borda é (**verde**), o valor é 0 e indica que as trajetórias atravessaram a região sem qualquer desvio realizado. Quando a cor é **amarela**, o valor está compreendido no intervalo $0 < i \leq 0.5$, onde i é o índice e indica uma baixa presença de desvios. Quando a cor é **vermelha**, o índice está compreendido entre $0.5 < i \leq 1$, onde i é o valor e indica uma alta presença de desvios. Por fim, uma borda (**branca**) indica que a área não foi visitada uma única vez pelas trajetórias e é representado pelo valor -1.

Os parâmetros usados foram os mesmos, tanto para as trajetórias de dados reais quanto para as trajetórias de dados simulados. Os valores 5, 10 e 12 foram escolhidos respectivamente para os seguintes parâmetros: raio da circunferência (área alvo-objeto), raio da região de desvios e comprimento mínimo em metros da subtrajetória. O uso de todos esses parâmetros é explicado na Seção 3.2.2.2.

6.2.3.1 Resultados Obtidos com Dados Reais

A Figura 6.10 representa os resultados obtidos na aplicação da metodologia para o conjunto de dados reais.

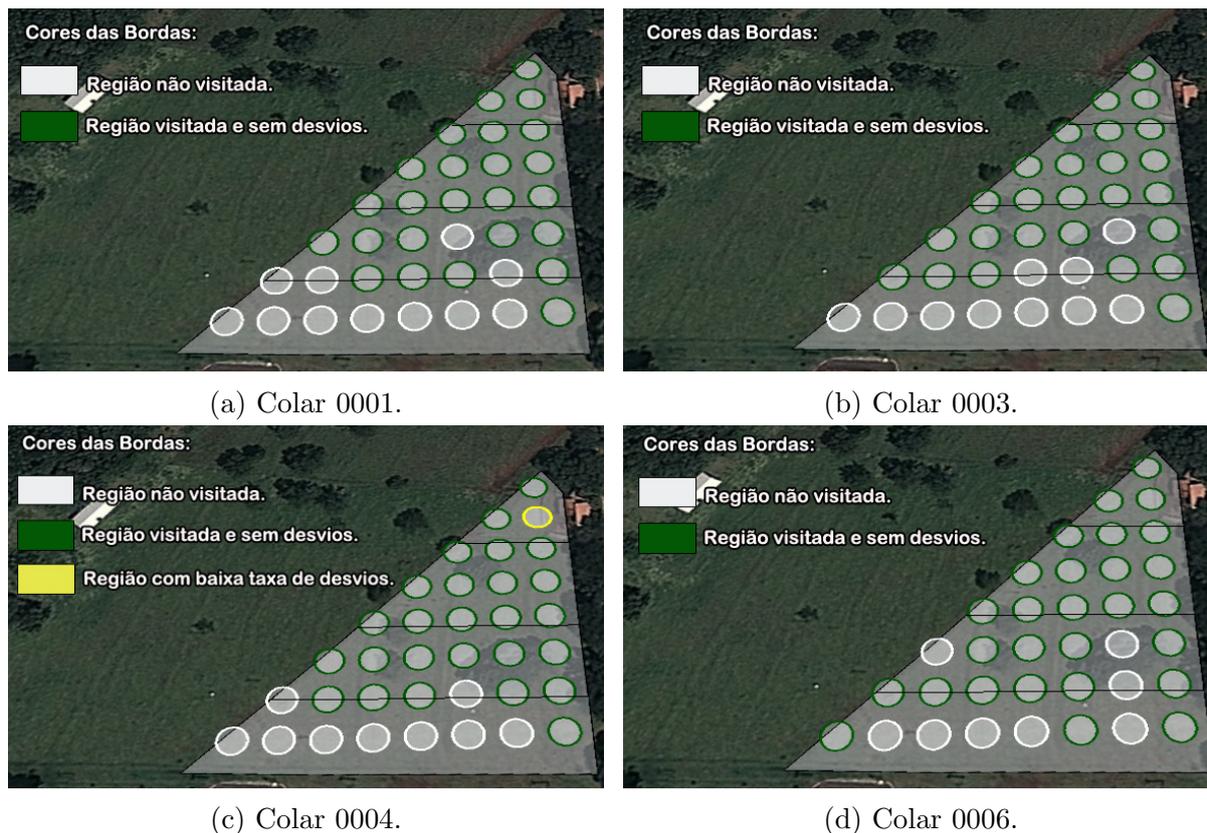


Figura 6.10: Área de Pastagem após a aplicação do padrão *avoidance* sobre o conjunto de trajetórias reais.

Como pode ser observado na imagem acima, o resultado do experimento casa perfeitamente com os resultados anteriormente observados. Com exceção de uma baixa incidência de desvio no mangueiro, relacionado ao animal do colar 0004, a grande parte da área de pastagem não foi evitada por nenhum dos quatro animais.

Em relação a *Área Sul*, praticamente todo o espaço que compreende a área está tomado pela cor branca, o que indica que o animal não esteve em boa parte da região, exceto a região compreendida pelo bebedouro o que prova que a maior parte dos deslocamentos em direção a *Área Sul* são por causa da água.

6.2.3.2 Resultados Obtidos com Dados Simulados

A Figura 6.11 representa os resultados obtidos na aplicação da metodologia para o conjunto de dados simulados.

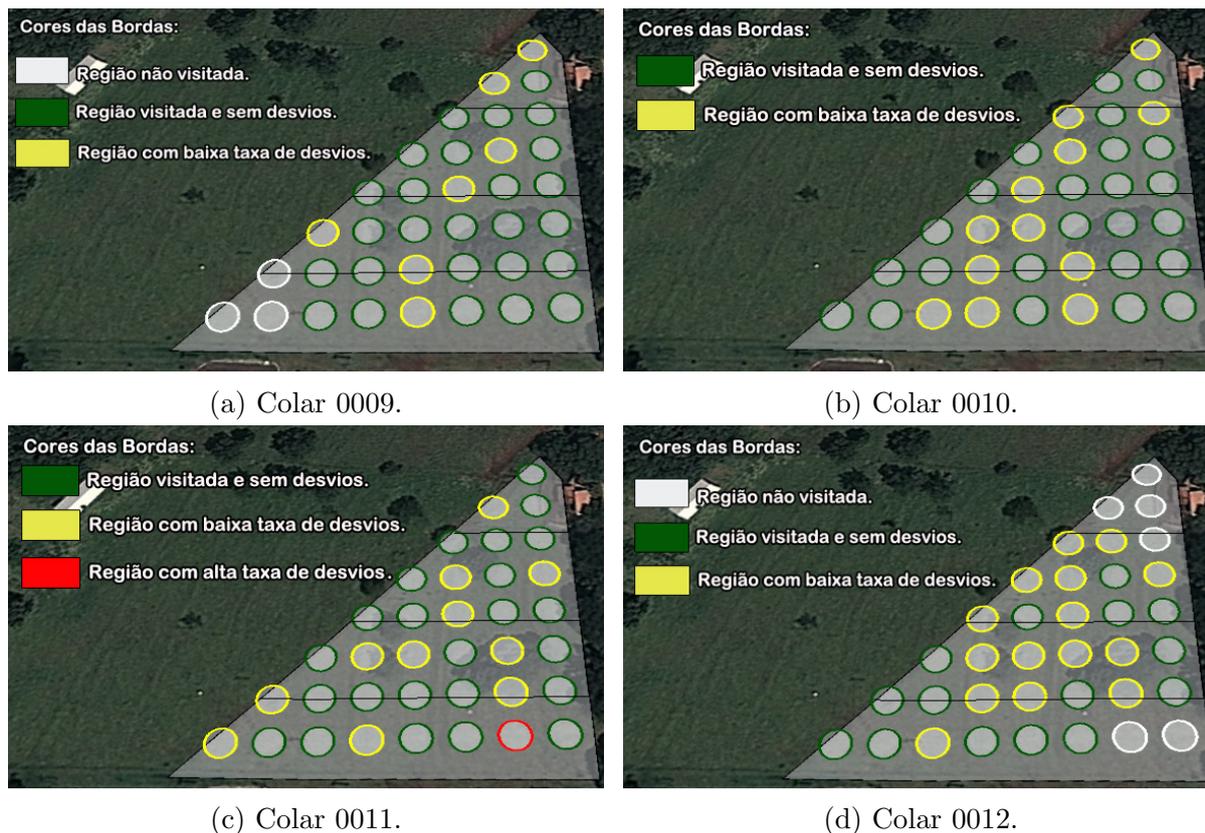


Figura 6.11: Área de Pastagem após a aplicação do padrão *avoidance* sobre o conjunto de trajetórias simuladas.

Em relação ao conjunto de dados reais, o resultado dos dados simulados apresentou uma incidência maior de desvios do que os dados de natureza real. Isso aconteceu devido ao fato das trajetórias simuladas apresentarem muito mais pontos e assim, a oportunidade de usufruir mais do espaço do ambiente de pastagem. E isso é comprovado se comparar o número áreas não visitadas de ambos os modelos.

O resultado deste algoritmo manteve também uma coerência com os resultados das técnicas apresentadas anteriormente. Por exemplo, nas regiões com bordas brancas dos cenários representados pelos colares 0009 e 0012, ao se comparar essas regiões com as áreas de concentração obtidas na Seção 6.2.2.1, observa-se que não há qualquer concentração de pontos sobre os lugares onde a cor branca surgiu. Ainda em relação ao colar 0012, na *Área Norte* há indicação de que não houve visitas em sua região. O fato comprovado pela Tabela 6.6, onde tempo que o animal permaneceu no local é 0 para todos os dias.

Com apenas uma única incidência da borda de cor vermelha, a cor amarela foi a mais predominante em relação às áreas onde houve ocorrência de desvios. O fato de uma área indicar a presença de desvios não que um animal tenha deixado de percorrer ou permanecer no local. Ao comparar as Figuras de Regiões de Concentração presentes na Seção 6.2.2.2, observa-se que várias regiões de concentração encontram-se em áreas da Figura 6.11 que apresentam bordas amarelas, isto é, uma baixa taxa de desvios.

Observando os resultados obtidos na simulação, uma explicação para as áreas onde

ocorrem os desvios poderia se basear na alimentação animal. Segundo Carvalho e Moraes [85], em um ambiente de pastagem, o bovino tem preferência em buscar regiões onde há abundância de forragem, isto é, pastos mais altos. Se anteriormente essas regiões apresentavam uma grande concentração por parte dos animais, isso significa que a forragem dessa região foi consumida e/ou pisoteada durante o pastejo. Logo, em breve este espaço apresentaram um pasto mais baixo, o que não será atraente ao animal. Isto faz com que o animal passe a procurar por regiões de pasto mais alto e passe a deixar de lado as áreas onde a vegetação está mais baixa.

6.3 Considerações Finais

Este capítulo apresentou estudos empíricos e simulados de uso do SATB para a análise comportamental de trajetória de bovinos. Além disso, serviu para demonstrar que a implementação do sistema está condizente ao que foi proposto, permitindo a execução dos estudos de caso.

Mesmo com uma quantidade limitada de dados, o estudo empírico realizado com os dados coletados em 2009 foram surpreendentes. Na aplicação dos três algoritmos, os dados mostraram uma grande coerência entre si além de revelar informações bastante interessantes, como por exemplo, os animais visitarem a *Área Sul* apenas para ir ao bebedouro. Mesmo sendo poucos dados, foi interessante como essas trajetórias se mostraram relacionadas.

Os dados simulados provaram a qualidade das metodologias e do sistema em lidar com um volume maior de dados. Apesar dos conjuntos não se mostrarem relacionados como no caso dos dados reais (até porque foram dados gerados de forma aleatória pelo computador), os resultados dos experimentos sobre cada conjunto mostrou se bem coerente.

Conclusão

Este projeto assumiu como objetivo o uso de metodologias e tecnologias para a manipulação e a análise comportamental de dados espaço-temporais de bovinos, com a finalidade de se obter conhecimento apurado sobre os animais monitorados em uma propriedade. Para isso, era necessário o desenvolvimento de um sistema computacional capaz de empregar as metodologias e tecnologias essenciais para auxiliar o usuário na manipulação e na análise sobre os dados georreferenciados.

Para a resolução da proposta, o conceito de trajetórias semânticas mostrou-se uma interessante opção para o entendimento das ações de bovinos em seu ambiente. Através de suas técnicas, foi possível conhecer tanto o dia-a-dia do animal em uma área de pastagem quanto suas preferências dentro daquele espaço. E mesmo apresentando algumas limitações (Seção 5.5), o estudo das trajetórias bovinas mostrou-se como uma forte aliada no estudo do comportamento animal.

Entre as principais contribuições técnicas deste trabalho, pode-se destacar: o uso do algoritmo **IB-SMoT** para identificar onde, quando e por quanto tempo um animal esteve em uma determinada área; uma ideia inicial para identificar os momentos e os lugares onde o animal esteve na sombra de uma árvore; a adaptação do algoritmo **CB-SMoT** para identificação das regiões de concentração de bovinos no ambiente de pastagem; a adaptação do padrão *Avoidance* para um ambiente aberto; e o uso da **Fórmula de Haversine**, possibilitando aos algoritmos usarem a distância geográfica entre os pontos de posicionamento.

Outra grande contribuição deste estudo foi a implementação do Sistema de Análise de Trajetórias Bovinas (SATB), um sistema capaz de interpretar e armazenar dados georreferenciados contidos em arquivos ARRF e em arquivos produzidos pelos equipamentos do projeto OTAG, cadastrar áreas de pastagens bem como seus elementos geoespaciais, prover a aplicação das metodologias de trajetórias semânticas e possibilitar a visualização das trajetórias e dos resultados obtidos pelos algoritmos em um mapa.

Visando validar o projeto até então realizado, um estudo de caso foi realizado usando um conjunto de dados reais e um conjunto de dados simulados, através de um simulador implementado e anexado ao SATB. Os resultados mostraram a utilidade das técnicas de trajetória semânticas no estudo do comportamento bovino. Uma grande surpresa pode ser observada nos resultados apresentados pelos dados reais que, apesar de serem poucos, possibilitaram inferências bastante interessantes sobre o comportamento dos animais, comprovando eficácia das técnicas.

O projeto integrou os três pilares da pecuária de precisão, que segundo Laca [86], consistem nos conhecimentos de comportamento animal, a tecnologia eletrônica e os sistemas de decisão aplicados ao pastoreio. Essa integração resultou em sistema computacional de análise comportamental a fim de auxiliar o usuário a conhecer as preferências do gado em uma área de pastagem. Por fim, um estudo de caso demonstrou o quanto as metodologias estudadas podem contribuir com a pecuária de corte nacional.

Mais do que empregar conceitos de trajetórias semânticas para estudar o comportamento de bovinos em pastejo, este trabalho representa um ponto de partida para algo muito maior. Com o SATB já é possível estudar o comportamento dos animais através de seus deslocamentos espaço-temporais em uma área de pastagem. Adicionando novas variáveis a esse escopo, como por exemplo, o peso do animal, a movimentação de sua cabeça, a temperatura e a umidade relativa do ar, abre-se uma gama maior de possibilidades para manipulação e análise de dados, bem como a capacidade de realizar inferências com maior precisão.

Com passar do tempo, o *software* irá incorporar novas variáveis e técnicas de análise de dados, bem como aperfeiçoará as metodologias presentes em sua estrutura. Com o crescimento e o desenvolvimento do sistema, em um futuro próximo será possível responder uma das grandes questões relacionadas à análise comportamental de bovinos, que consiste em prever as ações realizadas pelo animal e entender os motivos de tais ações. Em outras palavras, a grande ambição da área é saber o que o animal está “pensando”.

Trabalhos Futuros

Para a evolução do conceito teórico do projeto, um ponto a ser desenvolvido é a ideia da área de sombreamento. Utilizando informações como a altura e largura de uma árvore, pode-se prever o tamanho, a localização e o formato da área de sombra, de acordo com o período do dia e a época/estação do ano.

Outro ponto interessante a se considerar é a realização de novos experimentos contando com novos tipos sensores para os animais, como por exemplo, sensores de movimento (*motion sensor*) nos moldes daqueles usados por Ungar *et al.* [81]. Sensores de ambiente, como sensores de temperatura (usados por Johnson e Margalho em [87] para o estudo das variações climáticas na Amazônia) e sensores de umidade (usados por Pereira *et al.* [88] para avaliar a susceptibilidade da vegetação do cerrado a queimadas), poderiam também fazer parte de um eventual estudo futuro.

O uso de novos sensores, tanto para os animais quanto para o ambiente, implica tanto na evolução do sistema quanto na evolução dos algoritmos de análise comportamental implementados. Dispondo de outras fontes de dados, os cálculos precisarão ser repensados de forma que possa utilizá-los em novas inferências de comportamento animal.

Para a evolução do Sistema de Análise de Trajetórias Bovinas, as seguintes implementações são essenciais:

- A implementação do Padrão de Perseguição, descrito em [6]. Este padrão per-

mite identificar se uma trajetória está seguindo outra trajetória presente no mesmo espaço. Dentro do contexto da pecuária, este padrão poderia ajudar a identificar o líder de um rebanho e, principalmente, se um touro esteve atrás de uma vaca no cio.

- A integração de um serviço de mapas mais flexível. Além de possibilitar uma melhoria na visualização dos resultados, isso colocaria um fim à dependência do *Google Earth* instalado na máquina.
- Criação de papéis para usuários, onde cada papel possui permissões de execução diferenciadas. Por exemplo, um usuário com um papel de consulta poderia executar as metodologias de trajetórias, entretanto, esse mesmo papel não permitiria o gerenciamento dos dados ambientais.
- Criação de funções para gerenciar a remoção e a edição de trajetórias no Banco de Dados. Caso a criação de papéis para usuários seja implementada, o gerenciamento dos dados espaço-temporais precisa ser permitido apenas para usuários com papel de administrador.

Por fim, é necessário uma validação dos parâmetros usados pelos algoritmos de trajetórias semântica. Apesar dos parâmetros utilizados no Estudo de Caso (Capítulo 6) terem oferecido resultados bastante relevantes, ainda é preciso que vários testes sejam efetuados a fim de encontrar os melhores valores para cada algoritmo.

Referências Bibliográficas

- [1] IBGE. Produção da pecuária municipal, 2011. Disponível em www.ibge.gov.br/ProducaoPecuaria/ProducaoDaPecuariaMunicipal/2011/comentarios.pdf. Acesso em: 01 de Outubro de 2013.
- [2] Embrapa Gado de Corte. *Manual de Boas Práticas Agropecuárias - Bovinos de Corte*, 2006. Disponível em http://www.jbs.com.br/_doc/bpa_jbs_embrapa.pdf. Acesso em: 01 de Outubro de 2013.
- [3] F. Y. T. Munhoz, M. T. P. de Oliveira, P. M. Miyasato, M. A. I. da Silva, e C. Carromeu. Taurus: Uma plataforma de software para manejo agropecuário com foco em rastreabilidade e usabilidade. *Universidade Federal de Mato Grosso do Sul*, 2010.
- [4] P. C. de F. Carvalho, J. K. da Tridande, J. C. Mezzalira, C. H. E. C. Poli, C. Nabinger, T. C. M. Genro, e H. L. Gonda. Do bocado ao pastoreio de precisão: compreendendo a interface planta-animal para explorar a multi-funcionalidade das pastagens. *Revista Brasileira de Zootecnia*, v.38, p.109-122, 2009.
- [5] Q. I. dos S. Neto. Termus - uma plataforma para aferição remota de temperatura de bovinos da raça brangus. Dissertação de Mestrado, Universidade Federal do Mato Grosso do Sul, 2010.
- [6] V. Bogorny e F. J. Braz. *Introdução a Trajetórias de Objetos Móveis*. Editora Univalle, Primeira Edição, 2012.
- [7] F. L. Vilela, R. A. M. Brito, e K. E. Filho. O mercado da Carne, 2012. Disponível em www.cnpqc.embrapa.br/publicacoes/doc/doc107/028.html. Acesso em: 01 de Outubro de 2013.
- [8] V. Rocco e K. E. Filho. Qualidade da Carne, 2012. Disponível em www.cnpqc.embrapa.br/publicacoes/doc/doc107/029.html. Acesso em: 01 de Outubro de 2013.
- [9] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, e E. Cayirci. Wireless Sensor Networks: a Survey. *Computer Networks*, vol. 38, p. 393-422, 2002.
- [10] A. A. F. Loureiro. Redes de sensores sem fio. Em *Grandes Desafios da Pesquisa em Computação para o período 2006-2016*. Universidade Federal de Minas Gerais, 2006.
- [11] F. A. Silva, T. R. de M. Braga, e L. B. Ruiz. Tecnologia de nós sensores sem fio, Agosto de 2004. Disponível <http://homepages.dcc.ufmg.br/~linnyer/ufmgnsensores.pdf>. Acesso em: 01 de Outubro de 2013.

- [12] L. B. Ruiz. *MANNA: Uma Arquitetura para Gerenciamento de Redes de Sensores Sem Fio*. Dissertação de Doutorado, Departamento de Ciência da Computação da Universidade Federal de Minas Gerais, 2003.
- [13] L.B. Ruiz, L. H. A. Correia, L. F. M. Vieira, D. F. Macedo, E. F. Nakamura, C. M. S. Figueiredo, E. H. B. Maia, M. A. M. Vieira, J. M. S. Nogueira, A. A. F. Loureiro, D. C. da Silva, e A. O. Fernandes. *Arquiteturas para Redes de Sensores Sem Fio*. Em *XXII Simpósio Brasileiro de Redes de Computadores*, 2004.
- [14] A.A.F. Loureiro, J. M. S. Nogueira, L. B. Ruiz, R. A. de F. Mini, E. F. Nakamura, e C. M. S. Figueiredo. *Redes de Sensores Sem Fio*. Em *XXI Simpósio Brasileiro de Redes de Computadores*, 2003.
- [15] M. R. Pereira, C. L. de Amorim, e M. C. S. de Castro. Tutorial sobre Redes de Sensores. *Cadernos do IME - Série Informática*, vol. 14, p. 39-53, Junho de 2003.
- [16] P. B. Velloso. *Transmissão de Voz em Redes Ad Hoc*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Agosto de 2003.
- [17] M. A. Perillo e W. B. Heinzelman. *Fundamental Algorithms and Protocols for Wireless and Mobile Networks*, chapter *Wireless Sensor Network Protocols*, p. 813 – 842. CRC Hall, 2005.
- [18] A. B. Bondi. *Characteristics of Scalability and their Impact on Performance*. *Proceedings of the 2nd international workshop on Software and performance, WOSP '00*, p. 195 - 203, Setembro de 2000.
- [19] I. Dietrich e F. Dressler. *On the Lifetime of Wireless Sensor Networks*. *ACM Transactions on Sensor Networks*, vol. 5, no. 1, artigo no. 5, p. 1 - 39, Fevereiro de 2009.
- [20] V. Chandola e V. Kumar. *Summarization - Compressing Data into an Informative Representation*. *ACM Knowledge and Information Systems*, vol. 12, no. 3, p. 355 - 378, Agosto de 2007.
- [21] D. de O. Cunha. *Redes Sem Fio de Múltiplos Saltos: Protocolos específicos para aplicações e roteamento com suporte a diversidade cooperativa*. Dissertação de Doutorado, Universidade Federal do Rio de Janeiro, Fevereiro de 2008.
- [22] E.C.R. de Oliveira e C.V.N. de Albuquerque. *Avaliação de protocolos de roteamento para redes ad-hoc e RSSF aplicados à TV digital interativa e cidades digitais*. Em *XXXIII Conferencia Latinoamericana de Informatica (CLEI 2007)*, San Jose, Costa Rica, Outubro de 2007.
- [23] B. Krishnamachari, D. Estrin, e S. Wicker. *Modelling data-centric routing in wireless sensor networks*. Em *Proceedings of IEEE INFOCOM 2002*, Junho de 2002.
- [24] J. M. Stochero, A. J. G. Pinto, e J. F. de Rezende. *Roteamento com Agregação de Dados em Redes de Sensores*. Relatório Técnico, Universidade Federal do Rio de Janeiro, 2003.

- [25] J. N. Al-Karaki e A. E. Kamal. Routing techniques in wireless sensor networks: A survey. *Wireless Communications IEEE*, vol. 11, no. 6, p. 6-28, 20 de Dezembro de 2004.
- [26] L. Pelusi, A. Passarella, e M. Conti. Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad hoc Networks. *Communications Magazine, IEEE*, vol. 44, p. 134-141, Novembro de 2006.
- [27] L. Lilien, Z. H. Kamal, V. Bhuse, e A. Gupta. Opportunistic networks: The concept and research challenges in privacy and security. Relatório Técnico, Department of Computer Science Western Michigan University, Kalamazoo, 2005.
- [28] J. Sushant, K. Fall, e R. Patra. Routing in a Delay Tolerant Network. *ACM SIG-COMM'04 Computer Communication Review*, vol. 34, no. 4, p. 145-158, Outubro 2004.
- [29] A. Vahdat e D. Becker. Epidemic routing for partially connected ad hoc networks. Relatório Técnico, Department of Computer Science, Duke University, Durham, NC, 2000.
- [30] B. Burns, O. Brock, e B. N. Levine. MV Routing and Capacity Building in Disruption Tolerant Networks. Em *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, p. 398-408 vol. 1, Março de 2005.
- [31] M. Musolesi, S. Hailes, e C. Mascolo. Adaptive routing for intermittently connected mobile ad hoc networks. *6th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, Junho de 2005.
- [32] J. Leguay, T. Friedman, e V. Conan. Evaluating Mobility Pattern Space Routing for DTNs. Em *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, p. 1-10, Abril de 2006.
- [33] T. Small e Z. J. Haas. The Shared Wireless Infostation Model - A New Ad Hoc Networking Paradigm. Em *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc '03*, p. 233-244. ACM, Junho de 2003.
- [34] S. Jain, R. Shah, W. Brunette, G. Borriello, e S. Roy. Exploiting Mobility for Energy Efficient Data Collection in Wireless Sensor Networks. *ACM/Kluwer Mobile Networks and Applications (MONET)*, vol. 11, no. 3, p. 327-339, Junho de 2006.
- [35] W. Zhao, M. Ammar, e E. Zegura. A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks. *5th ACM international symposium on Mobile ad hoc networking and computing (Mobihoc)*, ACM Press, p. 187-198, Maio de 2004.
- [36] D. P. Agrawal. Designing Wireless Sensor Networks: from Theory to Applications. *Central European Journal of Computer Science*, vol. 1, no. 1, p. 2-18, Março de 2011.

- [37] R. Elmasri e S. B. Navathe. *Sistemas de Bancos de Dados*. Pearson - Addison Wesley, Quarta Edição, 2005.
- [38] H. H. Sferra e Â. M. C. J. Corrêa. Conceitos e aplicações de data mining. *Revista de Ciência e Tecnologia*, V. 11, Nº 22 – pp. 19-34, 2003.
- [39] S. M. Weis e N. Indurkha. Predict data mining. *Morgan Kaufmann Publishers, Inc*, 1999.
- [40] U. Fayyad, G. Piatetsky-Shapiro, e P. Smyth. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, vol. 17, no. 3, p. 37-54, 1996.
- [41] R. J. Brachman e T. Anand. The Process of Knowledge Discovery in Databases. Em Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, e Ramasamy Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, p. 37–57. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996.
- [42] C. H. Devêza. Minerando padrões sequenciais para bases de dados de lojas virtuais, Junho de 2011. Monografia (Bacharel em Ciência da Computação), UFOP (Universidade Federal de Ouro Preto), Ouro Preto/MG.
- [43] J. Han e M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, Elsevier, Segunda Edição, Março 2006.
- [44] H. Y. Shun e X. Wei. A study of spatial data mining architecture and technology. *2nd IEEE International Conference on Computer Science and Information Technology, Peqim, China: IEEE Computer Society, ago, p 163-166*, 2009.
- [45] B. S. Fajardo. Uma extensão nativa do SQL para mineração de trajetórias semânticas, Novembro 2008. Monografia (Bacharel em Ciência da Computação), UFRGS (Universidade Federal do Rio Grande do Sul), Porto Alegre/RS.
- [46] D. Kalyani e S. K. Chaturvedi. A survey on spatio-temporal data mining. *International Journal of Computer Science and Network (IJCSN)*, vol. 1, no. 4, Agosto de 2012.
- [47] M. Ester, A. Frommelt, H. P. Kriegel, e J. Sander. Database primitives for spatial data mining. *Data Mining and Knowledge Discovery, an International Journal, Kluwer Academic Publishers*, 1999.
- [48] S. Spaccapietra, C. Parent, M. L. Damiani, e C. Vangenot. A Conceptual View on Trajectories. *Data and Knowledge Engineering*, 65(1):126–146, 2008.
- [49] P. Laube e S. Imfeld. Analyzing Relative Motion within Groups of Trackable Moving Point Objects. Em *Proceedings of the Second International Conference on Geographic Information Science*, GIScience '02, p. 132–144, London, UK, 2002. Springer-Verlag.
- [50] P. Laube, S. Imfeld, e R. Weibel. Discovering Relative Motion Patterns in Groups of Moving Point Objects. *International Journal of Geographical Information Science*, vol. 19, no. 6, p. 639-668, Julho de 2005.

- [51] L. O. Alvares, V. Bogorny, B. Kuijpers, e A. Vaisman. A model for enriching trajectories with semantic geographical information. *ACM-GIS, New York, NY, USA. ACM Press, p. 162-169, 2007.*
- [52] M. C. T. Silva. TETL - Uma Ferramenta de ETL para Trajetórias de Objetos Móveis, Dezembro de 2009. Monografia (Bacharel em Ciência da Computação), UFPE (Universidade Federal de Pernambuco), Recife/PE.
- [53] A. T. Palma, L. O. Alvares, V. Bogorny, e B. Kuijpers. A clustering-based approach for discovering interesting places in trajectories. *Annual Symposium on Applied Computing, ACM SAC'08, Fortaleza, 2008.*
- [54] P. Berkhin. Survey of Clustering Data Mining Techniques. Relatório Técnico, Accrue Software, Inc., 2002.
- [55] A. M. Loy. Identificando padrões comportamentais do tipo avoidance em trajetórias de objetos móveis. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2011.
- [56] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, e D.I. Rubenstein. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with ZebraNet. *ACM SIGPLAN Notices, vol. 37, p. 96-107, 2002.*
- [57] R. N. Handcock, D. L. Swain, G. J. Bishop-Hurley, K. P. Patison, T. Wark, P. Valencia, P. Corke, e C. J. O'Neill. Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing. *Sensors 2009*, 13 de Maio de 2009.
- [58] M. C. Visoli, J. P. Chanet, F. Pinet, e S. Bimonte. Uso de Trajetórias na Rastreabilidade Bovina. Em *VII Congresso Brasileiro de Agroinformática, 2009, Viçosa - MG. Anais do VII Congresso Brasileiro de Agroinformática, 2009.* 5 p.
- [59] E. S. Nadimi, R. N. Jorgensen, V. Blanes-Vidal, e S. Christensen. Monitoring and classifying animal behavior using zigbee-based mobile ad hoc wireless sensor networks and artificial neural networks. *Computer and Electronics in Agriculture, vol. 82, p. 44-54, Março de 2012.*
- [60] D. J. de Moura, I. de A. Naas, M. P. G. Queiroz, e L. dos A. Brunassi. Estimativa de estro em vacas leiteiras utilizando métodos quantitativos preditivos. *Ciência Rural, Santa Maria, vol. 38, no. 8, 2008.*
- [61] A. M. Guimarães, R. A. Silva, P. J. Papaorgji, e J. C. F. da Rocha. Dager: Uma ferramenta para mineração de dados georreferenciados. *Revista de Engenharia e Tecnologia, vol. 4, no. 2, Agosto de 2012.*
- [62] M. A. S. Turine, H. da C. Sandim, C. Carromeu, M. R. Silva, C. O. Soares, E. N. Cáceres, P. P. Pires, M. A. I. da Silva, e Q. I. dos S. Neto. e-SAPI bovis: Plataforma web para rastreabilidade bovina. Em *VII Congresso Brasileiro de Agroinformática - SBIAgro 2009. Viçosa - MG, 2009.*

- [63] L. S. Mota. *Adaptação e interação genótipo-ambiente em vacas leiteiras*. Dissertação de Doutorado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, 1997.
- [64] M. R. Silva. *Middleware cerberus usando RFID para rastreabilidade bovina*. Dissertação de Mestrado, Universidade Federal do Mato Grosso do Sul, 21 de Agosto de 2009.
- [65] Agência Brasil. *Brasil tem 80,9 milhões de usuários de internet, mas expansão nas classes D e E e nas zonas rurais ainda é desafio*, 2013. Disponível em <http://agenciabrasil.ebc.com.br/noticia/2013-06-20/brasil-tem-809-mil-hoes-de-usuarios-de-internet-mas-expansao-nas-classes-d-e-e-e-nas-zonas-rurais-aind>. Acesso em: 01 de Outubro de 2013.
- [66] K. H. Neto, L. G. Nadalete, F. A. Gennari, e A. A. C. de Freitas. *Desenvolvimento de Sistema Web utilizando Arquitetura em Três Camadas e Applets*. Em *I WORKCOMP-SUL, Florianópolis/SC. Anais do I Workcomp-Sul: I Workshop de Computação da Região Sul*, 2004.
- [67] R. Heckel, R. Correia, C. Matos, e L. Andrade. *Architectural transformations: From legacy to three-tier and services*. *Springer Berlin Heidelberg*, p. 139-170, 2008.
- [68] F. Patrice, R. Raphael, M. Laure, e R. Philippe. *OTAG - Electronic Collars to Track Cattle, Manual of Use*, Março de 2009.
- [69] GeoRastro. *Sistema de geodécisão para rastreabilidade e produção sustentável da bovinocultura de corte*, 2012. Disponível em <http://www.cnpm.embrapa.br/projetos/georastro/>. Acesso em: 01 de Outubro de 2013.
- [70] ARFF. *Attribute-Relation File Format*, 2002. Disponível em <http://www.cs.waikato.ac.nz/ml/weka/arff.html>. Acesso em: 01 de Outubro de 2013.
- [71] Google. *Google Earth*, 2013. Disponível em <http://www.google.com.br/intl/pt-BR/earth/index.html>. Acesso em: 01 de Outubro de 2013.
- [72] Mathforum.org. *Deriving the Haversine Formula*, 1999. Disponível em <http://mathforum.org/library/drmath/view/51879.html>. Acesso em: 01 de Outubro de 2013.
- [73] Australian Government Geoscience Australia. *World geodetic system 84*, 2013. Disponível em <http://www.ga.gov.au/earth-monitoring/geodesy/geodetic-datums/other/wgs84.html>. Acesso em: 01 de Outubro de 2013.
- [74] JavaFX 2. *Documentation*, 2013. Disponível em <http://docs.oracle.com/javafx/>. Acesso em: 01 de Outubro de 2013.
- [75] J. Hunter e R. Lear. *Java Document Object Model*, 2000. Disponível em <http://www.jdom.org/index.html>. Acesso em: 01 de Outubro de 2013.
- [76] S. Colebourne e B. S O'Neill. *Joda Time - java date and time API*, 2002. Disponível em <http://joda-time.sourceforge.net/>. Acesso em: 01 de Outubro de 2013.

- [77] PostgreSQL Global Development Group. PostgreSQL - the world's most advanced open source database, 2013. Disponível em <http://www.postgresql.org/>. Acesso em: 01 de Outubro de 2013.
- [78] PostGIS Project Steering Committee. PostGIS - spatial and geographic objects for postgresql, 2013. Disponível em <http://postgis.net/>. Acesso em: 01 de Outubro de 2013.
- [79] Machine Learning Group at the University of Waikato. Weka 3 - data mining software in java, 2013. Disponível em <http://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 01 de Outubro de 2013.
- [80] D. Ganskopp. Manipulating cattle distribution with salt and water in large arid-land pastures: a GPS/GIS assessment. *Applied Animal Behaviour Science*. 73, 251-262, 2001.
- [81] E. D. Ungar, Z. Henkin, M. Gutman, e D. Ganskopp. Inference of animal activity from GPS collar data on free-ranging cattle. *Rangeland Ecology and Management*. 58(3), 256-266, 2005.
- [82] A. Einemo. Interpretation of data from GPS-collars with movement sensors on cattle. *Dept. of Animal Environment and Health, SLU. Studentarbete (Sveriges lantbruksuniversitet, Institutionen för husdjurens miljö och hälsa) vol. 209*, 2008.
- [83] Operational Management, Geodecisional Prototype to Track, e Trace Agricultural Production. Final activity report. *Project number FOOD - CT - 2006 - 043134*, 2010.
- [84] A. I. C. Arce, A. R. B. Tech, S. P. A. Souto, e E. J. X. Costa. Simulador de deslocamento de rebanho bovino para avaliação de monitoramento baseado em redes de sensores sem fio. *Revista Brasileira de Agroinformática*, vol. 8, no. 1, p. 1-16, 2006.
- [85] P. C. F. Carvalho e A. de Moraes. Comportamento ingestivo de Ruminantes: bases para o manejo sustentável do pasto. Em *Ulysses Cecato; Clóves Cabreira Jobim. (Org.). Manejo Sustentável em Pastagem. Maringá-PR: UEM, vol. 1, p. 1-20*, 2005.
- [86] E. A. Laca. Pastoreo de Precisión. *Bioma Campos: innovando para mantener su sustentabilidad y competitividad. Montevideo: Tradinco, vol. 1, p. 29-40*, 2008.
- [87] T. M. Johnson e M. Margalho. Redes de sensores sem fio para monitoramento agro-climatológico na amazônia. Em *Semana Paraense de Informática, Belém*, 2006.
- [88] A. Pereira, H. França, e J. E. dos Santos. Método para Avaliação da Susceptibilidade de Vegetação do Cerrado ao Fogo em Relação a Indicadores Antrópicos. *Anais XI Simpósio Brasileiro de Sensoriamento Remoto, Belo Horizonte, Brasil, INPE, p. 501-508*, 05-10 de Abril de 2003.