

Revisão de técnicas de Aprendizado de Máquina no contexto de Correspondência de Produtos

Eduardo F. Valim¹, Renato P. Ishii¹

¹Faculdade de Computação — Universidade Federal de Mato Grosso do Sul (UFMS)
Campo Grande, MS — Brasil

{eduardo.valim, renato.ishii}@ufms.br

Abstract. This paper conducts a systematic literature review, with the intention of identifying the most effective machine learning techniques for the product matching problem. It begins by establishing a conceptual groundwork for the discussion of the different approaches, presenting an comprehensive overview of the field. Subsequently, it outlines the review methodology and selection protocols. Finally, a critical evaluation of the methods and results found in these studies is conducted, synthesizing their findings.

Resumo. Este trabalho conduz uma revisão sistemática da literatura, com a intenção de identificar as técnicas de aprendizado de máquina mais eficazes para o problema de correspondência de produtos. Inicia-se estabelecendo uma base conceitual para a discussão das diferentes abordagens, apresentando uma visão abrangente da área. Em seguida é estabelecida a metodologia de revisão e os protocolos utilizados na seleção dos trabalhos. Por fim, é realizada uma avaliação crítica dos métodos e resultados encontrados nestes estudos, sintetizando seus resultados.

1. Introdução

O comércio eletrônico global, já solidamente estabelecido, continua em grande expansão. Projeções indicam que até 2025, este setor pode exceder a marca de 7 trilhões de dólares¹. Este imenso mercado não é apenas um fenômeno comercial, mas também uma fonte imensa de dados. Um exemplo claro é o website brasileiro da Amazon, que sozinho oferece mais de 15 milhões de produtos em seu catálogo². Neste cenário, as técnicas de correspondência de produtos emergem como uma importante ferramenta tanto para empresas, na gestão destes dados, como para usuários que desejam comparar ofertas.

Correspondência de produtos é um subcampo de *Entity Matching* (EM), e define o processo de parear entre diferentes ofertas que se referem ao mesmo produto, como exemplifica a figura 1. Tradicionalmente, essa tarefa era realizada por meio de regras fixas, usando comparação de códigos identificadores únicos, como o (*Stock Keeping Unit*) (SKU). Esta abordagem, porém, não consegue distinguir entre itens com ligeiras diferenças, que costumam ter os mesmos códigos, e não pode ser usada para comparar bases de dados que usam chaves diferentes.

A partir do surgimento do aprendizado máquina (AM), novas técnicas mais eficientes começaram a surgir. Através da utilização de campos como o nome, foto e

¹<https://www.insiderintelligence.com/content/worldwide-e-commerce-forecast-update-2022>

²<https://www.amazon.com.br/b?node=18274725011>



Figura 1. Distribuição do número de trabalhos selecionados por ano

descrição do produto, passou a ser possível treinar modelos capazes de reconhecer produtos idênticos, sem a necessidade de catalogação manual, em bases de dados diferentes e com informações incompletas e inconsistentes. Inicialmente usando técnicas mais simples como Máquina de Vetores de Suporte (SVM) e agrumanto (como *K-means*), hoje observa-se uma ampla gama de abordagens mais sofisticadas, como Redes Neurais Convolucionais (CNNs), Processamento de Linguagem Natural (NLP) e mais recentemente, os poderosos modelos de arquiteturas *transformer*.

Neste cenário, por meio de uma revisão sistemática da literatura, usando um protocolo definido, este trabalho visa compreender quais são os modos de realizar correspondência de produtos mais eficazes da atualidade. Na Seção 2, são introduzidos alguns conceitos teóricos para familiarizar o leitor com os jargões da área. Na Seção 3, é definida a metodologia de revisão utilizada, detalhando os passos na seleção dos artigos, as bases utilizadas e os critérios de seleção. Finalmente, na Seção 4, são discutidos os resultados encontrados, abordados os desafios encontrados pelos autores e apresentando as técnicas mais prevalentes, e as mais eficientes.

2. Fundamentação Teórica

2.1. Aprendizado de Máquina

Aprendizado de máquina (do inglês *Machine Learning* – ML) é um ramo fundamental da inteligência artificial, que habilita computadores a aprender e melhorar a partir de dados. A ideia central é ter, ao invés de códigos fixos, rigidamente programados, algoritmos que consigam identificar padrões em dados brutos, ajustar seus parâmetros internos e assim “aprender”, para então conseguirem fazer previsões com base em novas informações. [Goodfellow et al. 2016]

Os conjuntos de dados, conhecidos como *datasets*, são uma coleção de amostras caracterizadas por propriedades *features*, sendo uma dessas a informação de interesse, chamada de rótulo, ou *label*. Posto de forma simplificada, o objetivo de um algoritmo de AM é encontrar uma equação, cujas variáveis são as *features*, que prediga o valor do *label*. Percorrendo todo o *dataset*, o processo de treinamento realiza essa tarefa ajustando os pesos (*weights*) da equação por meio de um método iterativo, geralmente minimizando uma função de perda que mede a diferença entre as previsões do modelo e os valores reais dos *labels*. O resultado deste processo é denominado modelo.

Antes do treinamento, os dados passam por uma fase de pré-processamento, onde são transformados em um formato mais adequado para a modelagem. Um exemplo comum é a conversão de datas em variáveis categóricas. No campo do AM, existem quatro categorias principais de aprendizado, cada uma utilizando técnicas específicas. O aprendizado supervisionado treina modelos com dados rotulados para aprender a relação entre *features* e *labels*, como na classificação de e-mails em “spam” ou “não spam”. O aprendizado não supervisionado, por outro lado, trabalha com dados não rotulados, empregando métodos como agrupamento (K-means, DBSCAN) e redução de dimensionalidade (PCA, t-SNE) para descobrir padrões nos dados. O aprendizado semi-supervisionado mescla elementos dos métodos supervisionados e não supervisionados, usando ambos os tipos de dados. Finalmente, o aprendizado por reforço foca em treinar modelos para tomar decisões, aprendendo mediante recompensas baseadas em suas ações.

No processo de AM, cada iteração do modelo envolve etapas cruciais de otimização e regularização. Durante estas fases, técnicas como o gradiente descendente são aplicadas para minimizar os erros do modelo. Um aspecto importante nesta etapa é a prevenção do sobreajuste, que ocorre quando um modelo se ajusta excessivamente aos dados de treinamento, prejudicando sua habilidade de generalizar para novos dados. Para evitar isso, o modelo é treinado, validado e testado usando três conjuntos de dados distintos: treino, validação e teste. O conjunto de treino é usado para ajustar os parâmetros do modelo. O conjunto de validação, separado dos dados de treino, ajuda a avaliar o desempenho do modelo durante o treinamento e ajustes. Por fim, o conjunto de teste é usado para avaliar o desempenho do modelo após o treinamento, fornecendo uma medida da eficácia do modelo em condições reais.

Na avaliação de modelos de AM, utilizam-se métricas como precisão (proporção de identificações positivas corretas), *recall* (proporção de casos positivos reais corretamente identificados) e F1 score (1) (combinação harmônica de precisão e *recall*). Estas métricas avaliam o equilíbrio entre a capacidade do modelo de classificar corretamente exemplos positivos sem incorrer em falsos positivos.

$$\text{F1 Score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (1)$$

Este conjunto de conceitos forma a base do AM, preparando o terreno para abordagens para os próximos tópicos desta seção.

2.2. Redes Neurais

Redes neurais constituem a essência do *Deep Learning* (Aprendizado Profundo), modelando padrões complexos por meio de neurônios artificiais. Interligados em uma arquitetura de rede, esses neurônios processam dados e aprendem ajustando pesos internos. O processo de aprendizado se desdobra em duas fases cruciais:

- **Forward Propagation:** Os dados de entrada são processados sequencialmente pelas camadas da rede.
- **Backpropagation:** O erro calculado é propagado de volta pela rede, permitindo o ajuste dos pesos para minimizar o erro de predição.

Existem diversos tipos de redes neurais, cada uma adequada para tarefas específicas, como:

- **CNNs (Convolutional Neural Networks)**: Estas redes são particularmente eficientes no reconhecimento de padrões espaciais, como texturas e formas em imagens. Elas utilizam operações de convolução que filtram e preservam características espaciais essenciais, tornando-as ideais para aplicações de visão computacional, como detecção de objetos e reconhecimento facial.
- **DNNs (Dense Neural Networks)**: Também conhecidas como redes neurais totalmente conectadas, as DNNs consistem em camadas onde cada neurônio está conectado a todos os neurônios na camada seguinte. Essa estrutura permite que a rede considere todos os recursos de entrada de maneira integrada, sendo útil em tarefas onde a relação entre todos os recursos é importante, como na classificação de dados complexos.
- **RNNs (Recurrent Neural Networks)**: As RNNs são projetadas para lidar com sequências de dados, como séries temporais ou texto. Elas possuem a capacidade de manter um “estado” ou memória dos inputs anteriores, o que é crucial para entender o contexto em sequências. Variações como Long Short-Term Memory (LSTM) e Gated Recurrent Unit (GRU) foram desenvolvidas para superar limitações das RNNs tradicionais, especialmente em relação à captura de dependências de longo prazo em sequências de dados.

Para garantir a generalização destas redes e evitar o sobreajuste, são empregadas técnicas de regularização, como *Dropout* (eliminação aleatória de neurônios durante o treinamento) e *Batch Normalization* (normalização dos lotes).

A compreensão destas ideias é essencial para o estudo do próximo tema, as Arquiteturas *transformers*, um tipo revolucionário de rede neural que amplia significativamente as capacidades da Inteligência Artificial moderna.

2.3. Arquiteturas *Transformer*

As Arquiteturas *transformer* representam um avanço significativo no campo do AM. Introduzidas pelo seminal artigo *Attention is All You Need* [Vaswani et al. 2017], estas arquiteturas destacam-se pela eficiência e eficácia no tratamento de sequências de dados, especialmente texto.

Um aspecto fundamental dos *transformers* é a divisão em duas partes principais: os *encoders* e *decoders*. Os primeiros processam e dedodificam a entrada, capturando as informações contextuais de cada palavra na sequência. Cada *encoder* gera uma representação rica em contexto que é, então, passada aos *decoders*, que utilizam essa representação contextual com as entradas anteriores para gerar a saída sequencial. A depender da arquitetura, são utilizados mais de uma dessas estruturas, como em abordagens multimodais, onde um **enconder** para processamento textual e outro para imagens. Esta estrutura é particularmente eficaz em tarefas como tradução automática, onde o contexto completo da sentença de entrada é necessário para gerar uma tradução precisa.

O componente central de um *transformer* é o mecanismo de atenção, que permite ao modelo ponderar diferentes partes de uma entrada de acordo com sua relevância. Diferente das RNNs e LSTMs, que processam dados sequencialmente, os transformers analisam todas as partes da sequência simultaneamente (processamento paralelo), mantendo o conhecimento do contexto em longas cadeias de *tokens*, resultando em ganhos significativos de desempenho e eficiência.

Em um *transformer*, a atenção é geralmente implementada através do mecanismo de atenção autorregressiva, que permite que cada posição na sequência de entrada influencie todas as outras posições. Este mecanismo é descrito matematicamente por (2).

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (2)$$

Q , K , e V representam as matrizes de consulta, sendo Q (consulta ou *query*) refere-se à parte da entrada para a qual a atenção está sendo calculada. K (chave ou *key*) representa os elementos nos dados de entrada que são comparados com a *query*. V (valor ou *value*) é o conteúdo real associado a cada *key*, contendo as informações efetivamente passadas adiante na rede. A interação entre *query*, *key*, e *value* determina como a atenção é distribuída na sequência de entrada.

Essa arquitetura é aplicada em uma variedade de tarefas, como tradução automática, geração de texto e compreensão de linguagem, demonstrando resultados superiores em relação às abordagens anteriores. Além disso, sua aplicabilidade estende-se a outras áreas, como análise de séries temporais e processamento de imagens, indicando sua versatilidade e potencial para inovação contínua.

Entre os vários tipos *transformers* é importante destacar o BERT (*Bidirectional Encoder Representations from Transformers*), vastamente utilizado nos trabalhos aqui apresentados. Desenvolvido pelo Google, é um modelo de processamento de linguagem natural. Inova por usar técnicas de aprendizado profundo para entender o contexto das palavras nas frases, analisando-as em ambas as direções (esquerda e direita). Isso permite que o BERT tenha uma atuação superior em tarefas como compreensão de texto, tradução, e reconhecimento de fala, contribuindo para a eficiência e eficácia do modelo em lidar com a linguagem natural, sendo de extrema utilidade no processamento de descrições de produtos. Entre suas variações, está o RoBERTa, (*Robustly Optimized BERT Approach*), modelo maior, treinado por mais tempo e com mais dados, se sobressai em relação as outras, como será visto adiante.

Os *transformers*, com sua eficiência e flexibilidade, revolucionaram o AM, oferecendo novas perspectivas e possibilidades para o avanço das técnicas de análise de dados, incluindo aplicações em correspondência de produtos.

2.4. Outras Técnicas de AM

Avanços no campo do AM têm levado ao desenvolvimento de técnicas sofisticadas que ampliam as capacidades dos modelos tradicionais. As seguintes abordagens representam algumas das inovações mais significativas na área.

- **Processamento de Linguagem Natural (NLP):** Utiliza modelos de AM para interpretar, gerar e traduzir linguagem humana. Técnicas recentes, como redes neurais de atenção e *transformers*, têm se destacado por sua habilidade em capturar contextos complexos em grandes volumes de texto, superando as limitações das *RNNs* (Redes Neurais Recorrentes) na manutenção de dependências de longo prazo.
- **Aprendizado Multimodal e Contrastivo:** Esta abordagem integra informações de múltiplas modalidades, como texto, imagem e som, em uma única

representação. O aprendizado contrastivo, especificamente, foca em distinguir exemplos semelhantes e distintos durante o treinamento. Isso é feito por meio de funções de perda que penalizam a proximidade de representações de exemplos de diferentes classes e promovem a similaridade entre exemplos da mesma classe no espaço de representação.

- **Modelos Baseados em Grafos:** Modelos de grafos em AM são usados para representar dados com relações complexas e interconectadas. Eles funcionam criando nós que representam entidades e arestas que representam as relações entre essas entidades. Algoritmos como Redes Neurais de Grafos (*GNNs*) são capazes de aprender a estrutura do grafo e as características dos nós, oferecendo percepções sobre dados estruturados de forma não-linear. Há também os Grafos de Conhecimento que complementam essa abordagem, organizando informações de maneira semântica. Eles não só mapeiam entidades e relações, mas também adicionam significado e contexto, usando propriedades nos nós e arestas.
- **Refinamento (do inglês *fine-tuning*):** Também conhecida como aprendizado de transferência, esta técnica envolve a aplicação de conhecimentos adquiridos em um domínio para um domínio relacionado. Geralmente, isso é feito retreinando grandes modelos já treinados, como os baseados em *transformers*, em grandes conjuntos de dados para tarefas mais específicas através do *fine-tuning*. Esse processo adapta as camadas superiores do modelo à nova tarefa, aproveitando os conhecimentos gerais aprendidos durante o pré-treinamento.

Essas técnicas ilustram os desenvolvimentos atuais em AM, refletindo progressos significativos na área. Elas são essenciais para aprimorar e expandir as capacidades dos modelos tradicionais de AM, oferecendo ferramentas mais robustas e adaptáveis para uma variedade de desafios analíticos e de modelagem.

3. Metodologia da Revisão

A abordagem metodológica adotada neste trabalho foi baseada nas diretrizes de [Kitchingham and Charters 2007], que conceitua Revisão Sistemática da Literatura como:

“Uma forma de estudo secundário que utiliza uma metodologia bem definida para identificar, analisar e interpretar todas as evidências disponíveis relacionadas a uma questão de pesquisa específica de maneira imparcial e (até certo ponto) reproduzível”.

Nesta seção, são detalhados todos os passos tomados no processo de revisão deste estudo, definindo um protocolo que garante sua reprodutibilidade. Primeiro são descritas a pergunta de pesquisa e as bases de dados utilizadas, seguido pelos critérios de seleção e extração dos dados. Através desta abordagem rigorosa e estruturada, busca-se uma abrangente o atual cenário das técnicas de correspondência de produtos, visando reduzir ao máximo possíveis distorções e vieses.

3.1. Condução da Pesquisa

A pesquisa por literatura relevante iniciou-se pela escolha de bases de dados acadêmicas reconhecidas no campo de Ciência da Computação, foram selecionados seis bancos: ACM Digital Library, ArXiv, IEEE Xplore, Mendeley, SpringerLink e Web of Science.

| Base de Dados | Encontrados | Aprovados |
|----------------------|--------------------|------------------|
| ACM Digital Library | 58 | 4 |
| ArXiv | 10 | 4 |
| IEEE Xplore | 8 | 3 |
| Mendeley | 11 | 3 |
| SpringerLink | 235 | 1 |
| Web Of Science | 15 | 2 |
| Total | 337 | 17 |

Tabela 1. Distribuição dos trabalhos encontrados e aprovados por base de dados.

Com o intuito de conseguir uma ampla amostra de trabalhos, não foi escolhida uma expressão de busca muito restrita, sendo selecionados os seguintes termos: “product matching” AND (“machine learning” OR “deep learning” OR “artificial intelligence”, adaptando esta estrutura conforme a ferramenta, no período de 01/01/2018 a 12/12/2023. Após as buscas, foram encontrados 279 estudos que se adequavam ao escopo desta revisão, mostrando técnicas de AM para correspondência de produtos, apresentando seus resultados, conforme a Tabela 1.

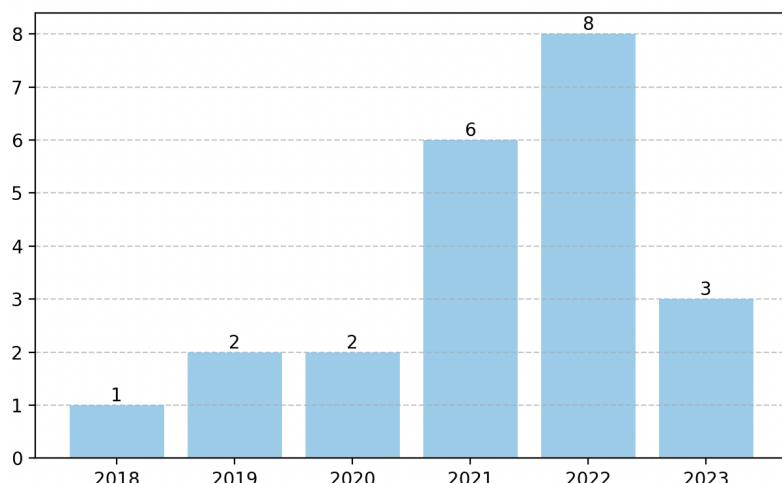


Figura 2. Distribuição do número de trabalhos selecionados por ano

3.2. Critérios de Inclusão e Exclusão

A filtragem subsequente das publicações foi faseada. O primeiro filtro excluiu os trabalhos que não se encaixavam no escopo e objetivos desta revisão. Já o segundo, selecionou os artigos que disponibilizam *links* funcionais para o código-fonte de seus experimentos, a fim de possibilitar a replicação dos métodos. No terceiro e último filtro, foi avaliada a credibilidade das publicações, usando o critério de classificação o Estrato Capes de B2 ou superior, através da ferramenta Qualis. Para os não encontrados neste site, foi ainda checado se a publicação possuía *Highest Percentil* maior ou igual a 25% na plataforma Scopus ou no Fator de Impacto de Jornal (JCR). Após este processo, ao desconsiderar os duplicados, foram selecionados 17 estudos, distribuídos conforme a Figura 2 e listados na Tabela 2, constituindo o conjunto de estudo desta revisão.

Tabela 2. Trabalhos selecionados

| Ano | Autor(es) | Título |
|------|--------------------|--|
| 2018 | Ristoski P et al. | A machine learning approach for product matching and categorization |
| 2019 | Primpeli A et al. | The WDC Training Dataset and Gold Standard for Large-Scale Product Matching |
| 2020 | Li J et al. | Deep cross-platform product matching in e-commerce |
| 2020 | Akritidis L et al. | A self-verifying clustering approach to unsupervised matching of product titles |
| 2021 | Gupte K et al. | Multimodal Product Matching and Category Mapping: Text+Image based Deep Neural Network |
| 2021 | Kim H et al. | Intent-based Product Collections for E-commerce using Pretrained Language Models |
| 2021 | Huang L et al. | Metric Learning Based Vision Transformer for Product Matching |
| 2021 | Foxcroft J et al. | Product Matching Lessons and Recommendations from a Real World Application |
| 2022 | Zhang D et al. | Multi-Faceted Knowledge-Driven Pre-Training for Product Representation Learning |
| 2022 | Mozdzonek M et al. | Multilingual Transformers for Product Matching - Experiments and a New Benchmark in Polish |
| 2022 | Shin W et al. | e-CLIP: Large-Scale Vision-Language Representation Learning in E-commerce |
| 2022 | Peeters R et al. | Supervised Contrastive Learning for Product Matching |
| 2022 | Chiu J et al. | Cross-Encoder Data Annotation for Bi-Encoder Based Product Matching |
| 2022 | Estrada R et al. | An Entity-Matching System Based on Multimodal Data for Two Major E-Commerce Stores in Mexico |
| 2022 | Feng C et al. | Multimodal Representation Learning-Based Product Matching |
| 2023 | Almagro M et al. | LEA: Improving Sentence Similarity Robustness to Typos Using Lexical Attention Bias |
| 2023 | Liu W et al. | Cross-platform product matching based on entity alignment of knowledge graph with ra |

4. Análise dos Trabalhos

Os artigos selecionados foram analisados em sua integralidade, focando nos métodos utilizados e nos resultados obtidos pelos autores. Ao observar a Figura 2, nota-se um notável aumento pelo interesse pelo tema nos últimos anos, mostrando que esse o campo de correspondência de produto ainda permite evolução significativa. Isso é observado na diversidade de abordagens adotados pelos autores, o que será detalhado nesta sessão.

A primeira tendência, é a transição de modelos tradicionais de AM, como Regressão Logística, CNNs [Akritidis et al. 2020] e *Conditional Random Field* (CRF) [Ristoski et al. 2018], para métodos que utilizam técnicas híbridas como o DeepMatcher [Primpeli et al. 2019], modelo sensível à sequência (RNN bidirecional) com um mecanismo de atenção, e mais notavelmente para *frameworks* que utilizam variações do *BERT* e outros *Transformers*, presentes em todos os trabalhos a partir de 2021. Porém, é importante notar que algoritmos clássicos ainda desempenham bem, conforme registrado por [Foxcroft et al. 2021], onde o modelo de Floresta Aleatório se saiu melhor em quase todos os *datasets* testados, atingindo F1 scores de 0,82 a 0,94, perdendo para o *transformer RoBERTa*, apenas no conjunto Amazon/Google, o qual é mais rico em textualidade.

Em sua maioria, os trabalhos tratam do problema utilizando apenas informações textuais, porém abordagens multimodais são utilizadas por três artigos. [Gupte et al. 2021], usando Redes Siamesas (RS) mostrou que a integração de texto e imagem mostrou ser mais eficaz na identificação de nuances e detalhes específicos dos produtos, como sabor, tamanho da embalagem, etc., que muitas vezes não são claros apenas pelo texto, o que era considerado um sinal fraco para correspondência, segundo [Ristoski et al. 2018]. [Huang et al. 2021] propõe o ML-VIT, que combina *Vision Transformer* para imagens com *ResNet* para agregação de características textuais, superando as RS com acurácia de 75,52% contra 36,67%. [Zhang et al. 2022] conseguiu um desempenho similar com o *KINDLE*, ainda que usando um *dataset* menor. Por fim, com o *ImageBERT* atingindo um F1 de 0,85, [Estrada-Valenciano et al. 2022] mostrou que o treinamento multimodal é superior aos unimodais e, ao contrário do que se esperava, texto e imagem treinados sozinhos obtiveram resultados similares.

Outra abordagem promissora, é o aprendizado contrastivo, já utilizada por [Gupte et al. 2021]. [Feng et al. 2022], utilizando uma técnica chamada *CoSENT* que distingue explicitamente as diferenças entre itens diversos, atingiu um F1 de 90.57. Mais a frente, [Peeters and Bizer 2022], atingindo F1 de 98,33, o maior dos observados, com o modelo *R-SupCon*, obtido por meio de *fine-tuning* do *RoBERTa*, também utilizado no artigo anterior, com um par de treinamento que passa por um pré-processamento de aprendizado contrastivo supervisionado.

[Chiu and Shinzato 2022] apresenta um *cross-encoder*, também com BERT, que codifica entradas conjuntamente, permitindo uma interação mais complexa e contextual entre elas, melhorando a qualidade dos *embeddings*. [Liu et al. 2023] introduz o RAEA, fundamentado em GNNs, emprega uma abordagem para alinhamento de entidades com grafos de conhecimento, mostrando-se útil na integração de dados de plataformas diferentes. Por fim, [Almagro et al. 2023] foca em aumentar a qualidade de dados com muito de má qualidade, usando um módulo inovador chamado *Lexical-aware Attention* (LEA) em métodos com cross-encoder. Esta camada, integrado ao mecanismo de autoatenção, prioriza tokens com alta similaridade lexical e melhora significativamente o desempenho

em ambientes ruidosos.

Vale notar que apenas sete, dos 17 artigos, disponibilizavam *links* dos códigos de seus experimentos para replicação. Também é importante observar que existem vários *datasets* distintos, impossibilitando uma comparação mais precisa dos métodos. No entanto, alguns autores já começam a usar conjuntos de dados padronizados, como o proposto por [Primpeli et al. 2019], habilitando avaliações mais consistentes de futuros métodos.

5. Conclusão

A revisão sistemática apresentada neste artigo evidencia uma evolução significativa nas técnicas de aprendizado de máquina aplicadas à correspondência de produtos. Observou-se uma transição de modelos tradicionais de AM, para abordagens mais sofisticadas e eficientes que incorporaram arquiteturas *transformer* e estratégias multimodais.

Os resultados indicam uma preferência crescente por *frameworks* que utilizam *fine-tuning* de modelos *transformers*, especialmente o *RoBERTa*, que demonstram capacidade superior em entender e processar a complexidade e nuances do texto, essenciais na correspondência de produtos. Destacando-se por sua habilidade em capturar contexto em ambientes de dados grandes e variados.

Juntamente com os mecanismos de autoatenção, a integração de dados multimodais, combinando texto e imagem, emergiu como uma estratégia promissora, proporcionando uma compreensão mais abrangente dos produtos. Esta abordagem se mostra particularmente eficaz em identificar características específicas dos produtos, muitas vezes não capturadas unicamente por textos.

Outra tendência observada é o uso de aprendizado contrastivo. Esta metodologia aprimora a eficácia dos modelos, otimizando sua capacidade de distinguir entre produtos semelhantes e distintos, resultando em uma precisão notavelmente alta na correspondência de produtos.

Em suma, esta revisão sistemática sintetiza o atual cenário da correspondência de produtos, destacando o papel crucial das técnicas avançadas de aprendizado de máquina na resolução de desafios neste domínio. À medida que o comércio eletrônico continua a crescer e evoluir, estas abordagens oferecem caminhos promissores para o desenvolvimento de soluções mais precisas e eficientes.

Referências

- [Akritidis et al. 2020] Akritidis, L., Fevgas, A., Bozanis, P., and Makris, C. (2020). A self-verifying clustering approach to unsupervised matching of product titles. *Artificial Intelligence Review*, 53:4777–4820.
- [Almagro et al. 2023] Almagro, M., Almazán, E., Ortego, D., and Jiménez, D. (2023). Lea: Improving sentence similarity robustness to typos using lexical attention bias. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, volume 1, pages 36–46. ACM.

- [Chiu and Shinzato 2022] Chiu, J. and Shinzato, K. (2022). Cross-encoder data annotation for bi-encoder based product matching. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 161–168. Association for Computational Linguistics.
- [Estrada-Valenciano et al. 2022] Estrada-Valenciano, R., Muñiz-Sánchez, V., and la Torre-Gutiérrez, H. D. (2022). An entity-matching system based on multimodal data for two major e-commerce stores in mexico. *Mathematics*, 10:2564.
- [Feng et al. 2022] Feng, C., Chen, W., Chen, C., Xu, T., and Chen, E. (2022). *Multimodal Representation Learning-Based Product Matching*, volume 1711 CCIS, pages 180–190. Springer Science and Business Media Deutschland GmbH.
- [Foxcroft et al. 2021] Foxcroft, J., Chen, T., Padmanabhan, K., Keng, B., and Antonie, L. (2021). Product matching lessons and recommendations from a real world application. *Proceedings of the Canadian Conference on Artificial Intelligence*.
- [Goodfellow et al. 2016] Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA, USA. <http://www.deeplearningbook.org>.
- [Gupte et al. 2021] Gupte, K., Pang, L., Vuuyuri, H., and Pasumarty, S. (2021). Multimodal product matching and category mapping: Text+image based deep neural network. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4500–4505. IEEE.
- [Huang et al. 2021] Huang, L., Shao, W., Wang, F., Xie, W., and Wong, K.-C. (2021). *Metric Learning Based Vision Transformer for Product Matching*, volume 13108 LNCS, pages 3–13. Springer Science and Business Media Deutschland GmbH.
- [Kitchenham and Charters 2007] Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.
- [Liu et al. 2023] Liu, W., Pan, J., Zhang, X., Gong, X., Ye, Y., Zhao, X., Wang, X., Wu, K., Xiang, H., Yan, H., and Zhang, Q. (2023). Cross-platform product matching based on entity alignment of knowledge graph with raea model. *World Wide Web*, 26:2215–2235.
- [Peeters and Bizer 2022] Peeters, R. and Bizer, C. (2022). Supervised contrastive learning for product matching. In *Companion Proceedings of the Web Conference 2022*, pages 248–251. ACM.
- [Primpeli et al. 2019] Primpeli, A., Peeters, R., and Bizer, C. (2019). The wdc training dataset and gold standard for large-scale product matching. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 381–386. ACM.
- [Ristoski et al. 2018] Ristoski, P., Petrovski, P., Mika, P., and Paulheim, H. (2018). A machine learning approach for product matching and categorization. *Semantic Web*, 9:707–728.
- [Vaswani et al. 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett,

R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[Zhang et al. 2022] Zhang, D., Liu, Y., Yuan, Z., Fu, Y., Chen, H., and Xiong, H. (2022). Multi-faceted knowledge-driven pre-training for product representation learning. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–12.