

Aplicações de IA em Data Mining

Gabriel Leal Chittolina Amaral

E-commerce

- Para aumentar as vendas de uma loja virtual e, possivelmente gerar novos empregos e até causar uma boa experiência de usuário, é essencial que se tenha bons algoritmos de recomendação e identificação de padrões para sugestão de novos itens baseado em padrões de compra

Problema

- Como sugerir produtos que normalmente são comprados juntos, de forma com que os usuários acabem por comprar mais itens na loja, além de gerar uma melhor experiência de usuário?

Solução: Market Basket Analysis

- O Market Basket Analysis [1] é uma técnica de mineração de dados usada pelos varejistas para aumentar as vendas, entendendo melhor os padrões de compra dos clientes. Envolve a análise de grandes conjuntos de dados, como histórico de compras, para identificar agrupamentos de produtos, bem como produtos que provavelmente serão comprados juntos.
- Normalmente se utiliza o Apriori Algorithm.
- De maneira abstrata e simplificada, o algoritmo produz os seguintes resultados:
 - Se {sanduiche, biscoitos} -> então -> { bebida }

Comparador de preços

- Hoje em dia temos alguns comparadores de preços robustos, onde podemos pesquisar por produtos que queremos comprar, e tendo como resultado uma lista de locais onde aquele produto é vendido, bem como seus preços, prazo de entrega etc. Exemplos: **Buscapé** ou **Zoom**. Nesta aplicação, será mostrado de maneira simplificada como esse tipo de sistema funciona.

Problema

- Como criar uma base de dados com todos os produtos agregados e, ao receber requisições de pesquisa, conseguir agrupar eles e identificar os produtos iguais de maneira a mostrar pro usuário os preços daquele determinado produto, de maneira eficiente?

Solução: Mix de algoritmos e Deep learning

- Normalmente são utilizados um conjunto de técnicas [2] para identificar produtos iguais:
 - Similaridade entre os títulos dos anúncios - algoritmo spacy ou GPT-3
 - Comparação de preços - algoritmo de clusterização (K-means, por exemplo) e detecção de outliers
 - Similaridade de imagens
 - Extração de atributos

Medicina

- A IA é extensamente utilizada na Medicina para diversas finalidades, como identificação de tumores, descoberta de novos medicamentos, etc. No entanto, existe um problema em alguns casos. Não se tem dados suficientes para se chegar a resultados satisfatórios.

Problema

- Para que a IA consiga chegar a resultados satisfatórios é necessário um grande conjunto de dados, o que se torna difícil na área de Medicina devido à privacidade dos dados e baixa quantidade.

Solução: GAN - Generative Adversarial Network

- Duas redes neurais trabalhando de maneira conjunta [3]
- Uma gera dados (ou imagens, por exemplo) sintéticas, enquanto a outra realiza a validação
- Caso não seja válida, a rede neural responsável pela geração, gera novamente até que se tenha um resultado válido
- Com isso é possível gerar diversos dados sintéticos (sem problemas com privacidade, conseqüentemente) que podem ser utilizados por diferentes instituições para diferentes finalidades, em algoritmos de IA para obtenção de resultados satisfatórios

Detecção de fraudes

- Para garantir a segurança dos sistemas bancários e consistência, é necessário evitar ao máximo as fraudes, minimizando gastos desnecessários e evitando problemas de segurança com os usuários.

Problema

- Como detectar transações fraudulentas e barrá-las de forma a minimizar os gastos e aumentar a consistência e confiabilidade dos sistemas bancários?

Solução: Detecção de outliers

- Existe, na verdade, mais de uma possível solução para este problema. No entanto, selecionei a de detecção de outliers.
- Utiliza-se um método de aprendizagem não supervisionado, mais especificamente "detecção de outlier comportamental" [4], onde são detectados transações que fogem do padrão das transações não fraudulentas, levantando flags de alerta.
- Podem ser utilizados métodos supervisionados, porém estes só são úteis para detecção de fraudes que já ocorreram anteriormente, não sendo possível detecção de novos tipos de fraude

Referências

[1] - Market Basket Analysis 101: Anticipating Customer Behavior

<https://smartbridge.com/market-basket-analysis-101/>

[2] - 5 Best Use Cases For Product Matching In Ecommerce & How You Can Implement Each One

<https://www.width.ai/post/product-matching-in-ecommerce>

[3] - How Artificial Intelligence can be used for Data Mining

<https://ai-med.io/neurology/gans-in-medicine/>

[4] - Data mining Techniques for Fraud Detection

<http://ijcsit.com/docs/Volume%204/Vol4Issue1/ijcsit2013040101.pdf>