

Oi,

Somos do curso de Sistemas de Informação da Universidade Franciscana, e esse ebook é um produto *exclusivo* criado pra você.

Nele, você pode ter um gostinho de como é uma das primeiras aulas do seu futuro curso.

Ficou curioso? Então conheça nosso universo.



Inteligência Artificial – Mineração de dados

A Mineração de dados é o processo de descoberta automática de informações úteis em grandes depósitos de dados.

O processo de mineração de dados é composto de várias etapas, não triviais, interativas e iterativas, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis, a partir de grandes conjuntos de dados.

A análise destes dados é feita por meio de técnicas de software para encontrar padrões e regularidades em conjuntos de dados. O computador é responsável por encontrar os padrões por meio da identificação de regras e características implícitas nos dados. Cada conjunto de dados é diferente, por esta razão há várias formas de abordagens dos dados para se extrair estas informações novas.

É possível “achar ouro” em lugares inesperados na medida em que o software de mineração de dados extrai padrões antes não discerníveis ou não tão óbvios que ninguém tenha notado antes.

É possível fazer uma analogia com a mineração efetuada em minas de ouro ou outros minerais. Pois grandes volumes de dados são “peneirados” na tentativa de se encontrar alguma coisa de valor.

Com este intuito, pode-se tentar auxiliar em várias áreas do mundo real com os resultados encontrados no processo de mineração de dados, por exemplo:

- Qual produto de alta lucratividade venderia mais com a promoção de um item de baixa lucratividade, analisando os dados dos últimos dez anos?

- Quais são os clientes potenciais para praticar fraudes?
- Quais clientes gostariam de comprar o novo produto X?
- Que genes são determinantes para o diagnóstico de um determinado tipo de doença?

Um exemplo do uso de mineração de dados na vida real faz com que os seguros de carros sejam mais baratos para mulheres por exemplo, que em cidades grandes, por haver maior risco de colisão e roubo, estes sejam mais caros.

Um exemplo da utilização foi realizado e chamado de “The weather problem”, visto na Figura 1. Observa-se nesta figura anotações realizadas em 14 dias em um campo de futebol 7, onde o proprietário queria saber se as condições climáticas interferiam na ocorrência de jogo ou não.

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Figura 1 - Anotações feitas durante os 14 dias

Na Figura 1 têm-se 5 atributos, Outlook, Temperature, Humidity, Windy e o atributo classe Play. O autor da pesquisa anotou os dados de se estava chovendo ou nublado ou com sol, qual a temperatura do dia dividida em quente frio ou média, se a humidade era normal, alta ou baixa, se havia ou não vento e finalmente se naquelas características, haveria ou não jogo.

O Software começa a calcular e encontrar padrões entre os atributos e desta forma vai fazendo as previsões, como pode ser Observado na Figura 2.

```
If outlook = sunny and humidity = high then play = no
If outlook = rainy and windy = true   then play = no
If outlook = overcast                 then play = yes
If humidity = normal                  then play = yes
If none of the above                  then play = yes
```

Observa-se na Figura 2 que algumas regras combinando atributos foram definidas pelo software, como se estiver ensolarado e a humidade for alta, não haverá jogo, se estiver chovendo e com vento, também não haverá jogo. Também alguns atributos podem fornecer as informações sem a necessidade de combinação, como a terceira regra que diz que sempre que estiver nublado, haverá jogo.

Claro que esta é uma observação de uma base de dados pequenas, mas em grandes bases de dados, a forma de definição de padrões é a mesma,

utilizando-se diferentes tipos de algoritmos consegue-se agrupar dados e realizar previsões em várias áreas, como o surgimento de uma determinada doença, quando determinados genes sofrem mutações em um mesmo organismo.

Descoberta de conhecimento ou Knowledge Discovery in Database (KDD) é um outro termo para o processo de Mineração de Dados.

Data Mining: Practical Machine Learning Tools and Techniques

Autores: Ian H. Witten e Eibe Frank

Editora: Elsevier, 2ª edição (2005)

Agradecemos a leitura e
esperamos você na
Universidade Franciscana.

