

Análise Associativa

Prof. Nielsen Rechia

nielsen.machado@uniritter.edu.br

Regras de Associação

2

Paradigmas	Supervisionado		Não-supervisionado	
	Classificação	Regressão	Análise associativa	Agrupamento (<i>clustering</i>)
Tarefas	Outros	Outros	Redução de dimensionalidade	Outros

7 tarefas comuns de aprendizado de máquina:
<http://vitalflux.com/7-common-machine-learning-tasks-related-methods/>

Regras de Associação

No aprendizado não supervisionado normalmente não temos a informação da classe das instâncias de treinamento.

Agrupamento é uma tarefa que visa agrupar um conjunto de objetos (instâncias) em diferentes classes (grupos) de objetos de acordo com sua similaridade.

Regras de Associação

Dado: (1) conjunto de transações, (2) cada transação é uma lista de itens (compras de um cliente)

Obter: todas as regras que indicam a ocorrência de um item baseado na ocorrência de outros itens nas transações: ex: $X \rightarrow Y$, onde X e Y são itemsets {Leite, pão}.

TID	Itens
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca-cola
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca-cola

Exemplo de Regras de Associação:

1 - {Fralda} \rightarrow {Cerveja},

2 - {Leite, Pão} \rightarrow {Ovos, Coca-cola}

3 - {Cerveja, Pão} \rightarrow {Leite}

Regras de Associação – Apriori

5

Normalmente executado em 2 sub tarefas

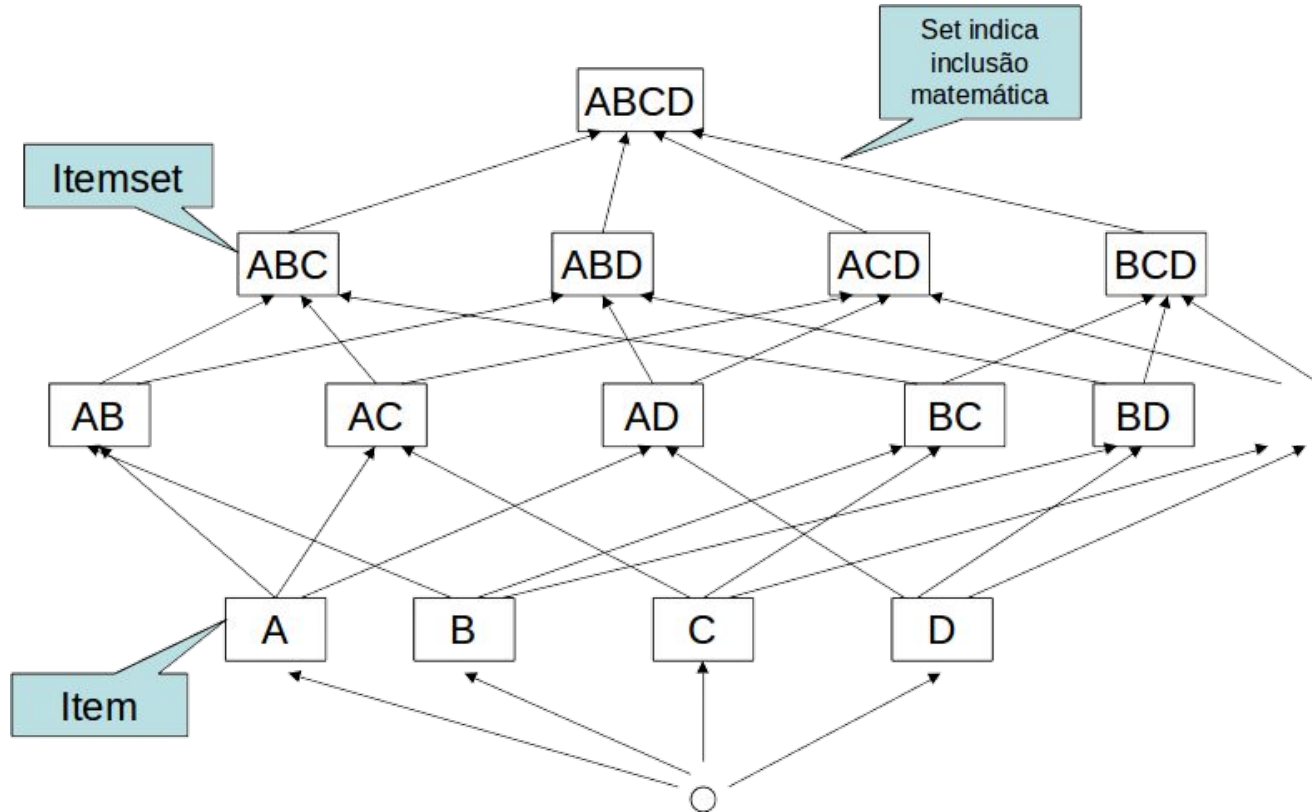
Geração de itens frequentes

Baseado em um threshold mínimo para suporte

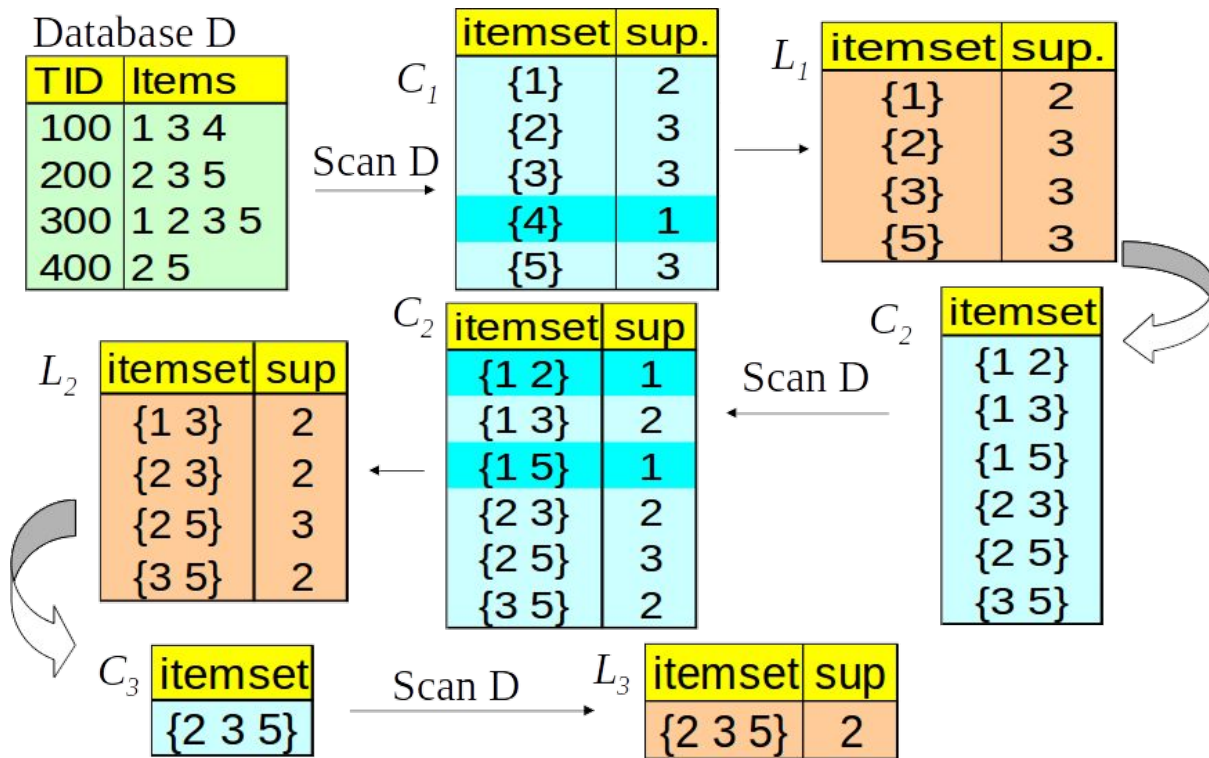
Geração de regras de associação

Baseado em um threshold de confiança e outras medidas como lift

Regras de Associação - Apriori



Regras de Associação - Apriori



Regras de Associação - Apriori

Suporte é a Fração de transações que contém X e Y;

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$$

Confiança mede quão frequente itens em Y aparecem em transações que contém X;

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{s(X)}$$

Lift mede a probabilidade de ocorrência de X e Y independentemente um do outro;

$$lift(X \rightarrow Y) = \frac{c(X \rightarrow Y)}{s(Y)}$$

= 1, indica independencia
> 1, indica correlação positiva
< 1, indica correlação negativa

Regras de Associação - Apriori

9

Exemplo no Jupiter

Exercício para entregar

10

Dados do conjunto de dados em
<https://archive.ics.uci.edu/ml/datasets/online+retail>

Encontrar, para um único país (ex Italia), regras usando o algoritmo apriori, como visto no exemplo no jupyter.

- Utilizar medidas de suporte, confiança e lift

- Testar alguns valores

- Definir possíveis melhores valores para tais medidas utilizadas

- Entregar via BB o código python ou jupyter e o passo a passo realizado (ex, motivos das escolhas dos valores etc.) em um zip.

Conclusão

11

Leitura recomendada:

Capítulo 6, 7 de Introduction to Data Mining

