



Machine Learning

Regras de associação: algoritmos

Prof. Hugo de Paula

Algoritmo Apriori

Baseado na ideia de usar conhecimento já obtido dos itemsets anteriores.

- Fase I:

Descobrir todos os conjuntos de itens com suporte maior ou igual ao mínimo suporte especificado pelo usuário.

Algoritmo Apriori

- Um subset de um itemset frequente também é um itemset frequente
 - P. ex., se $\{AB\}$ é um itemset frequente, ambos $\{A\}$ e $\{B\}$ devem ser um itemset frequente

Regras de associação: algoritmo Apriori

- Fase II:

A partir dos conjuntos de itens frequentes, descobrir regras de associação com fator de confiança maior ou igual ao especificado pelo usuário.

Regras de associação: algoritmo *Apriori*

Database D

TID	Itens
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Scan D



C_1

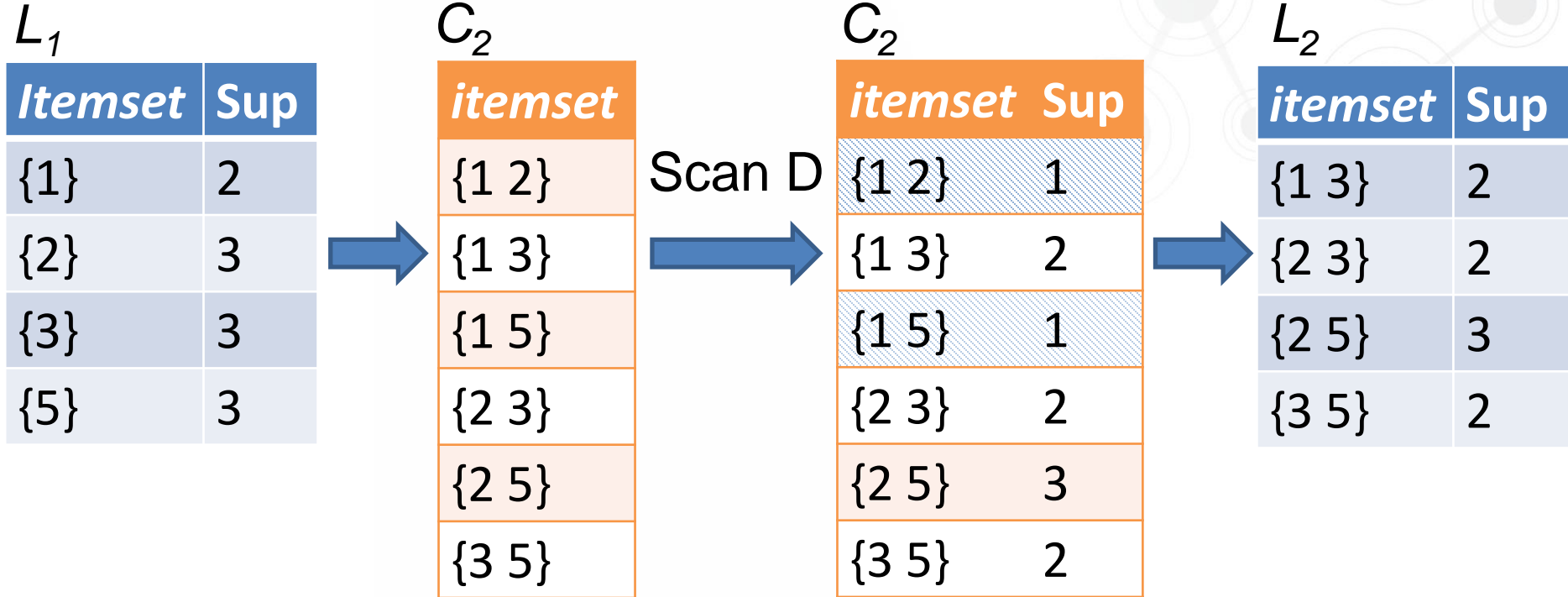
<i>itemset</i>	Sup.
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3



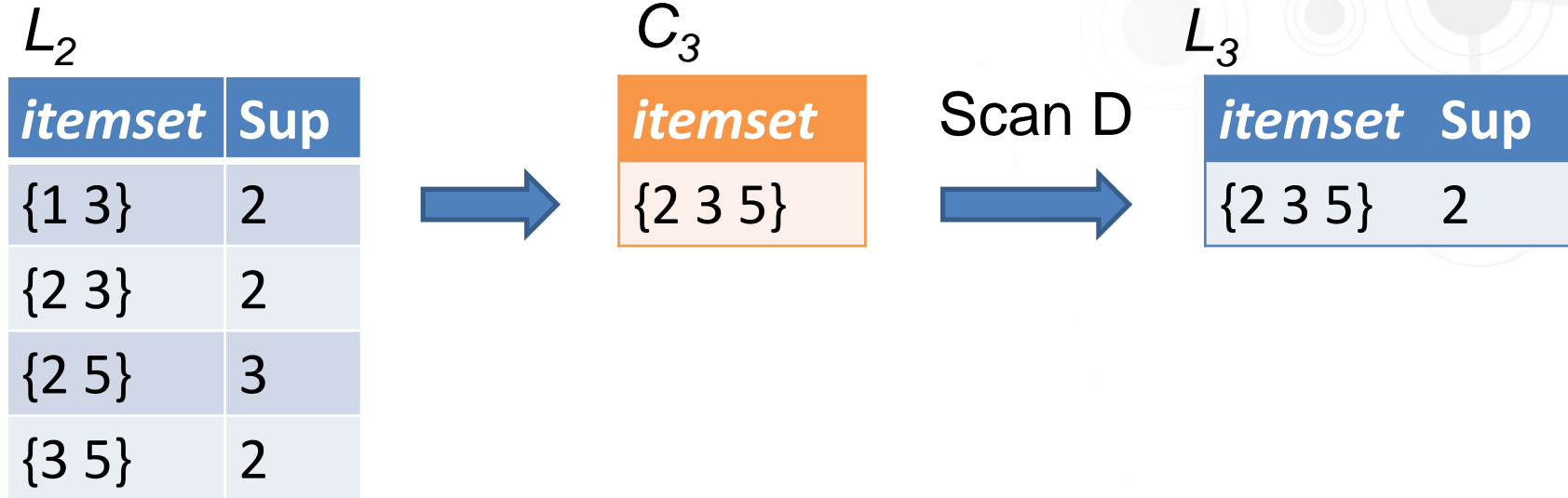
L_1

<i>Itemset</i>	Sup.
{1}	2
{2}	3
{3}	3
{5}	3

Regras de associação: algoritmo *Apriori*



Regras de associação: algoritmo *Apriori*



Algoritmo FP-growth

Método de geração de padrões frequentes de itens sem a geração de candidatos.

Mais eficiente e mais escalável que o algoritmo Apriori.

Algoritmo FP-growth

Percorre o banco de dados apenas duas vezes.

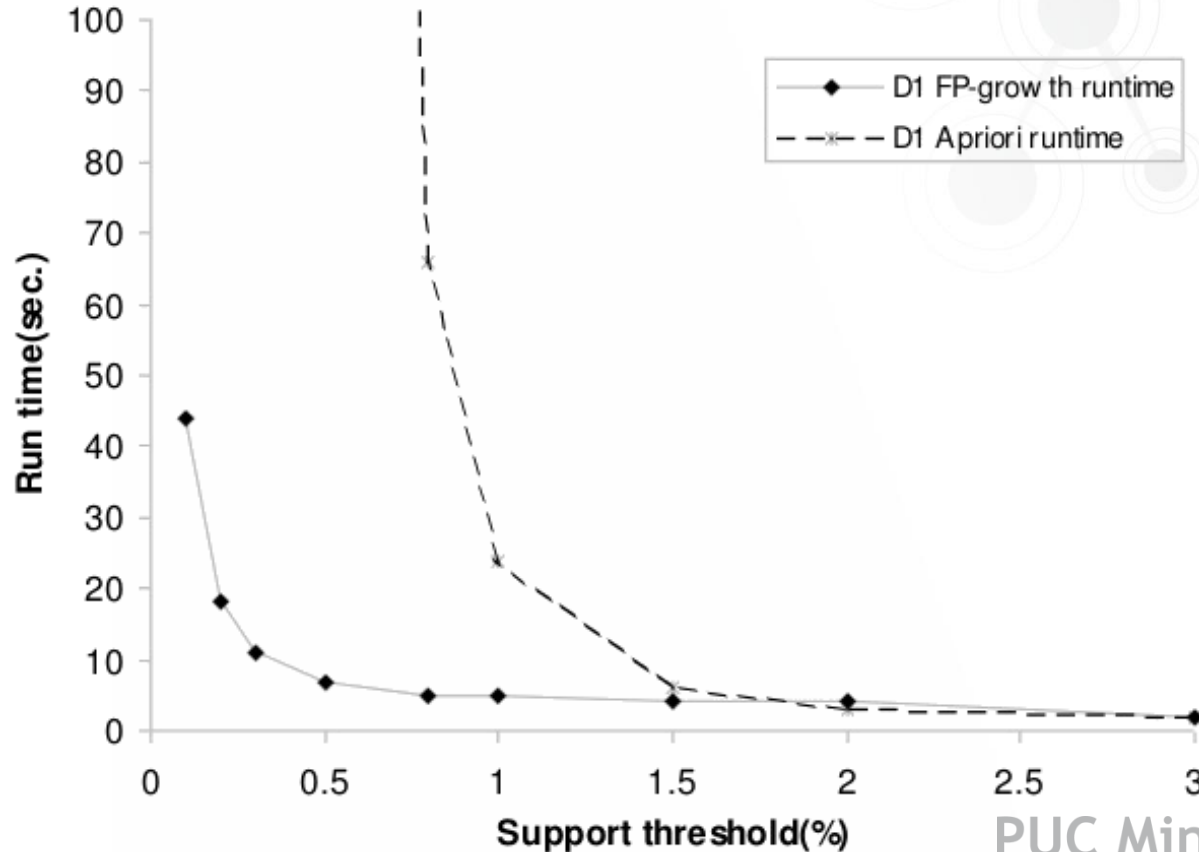
- Fase I:

Construir uma estrutura de dados compacta chamada FP-tree.

- Fase II:

Extrair *itemsets* frequentes diretamente da FP-tree.

Regras de associação: comparação



Aviso legal

O material presente nesta apresentação foi produzido a partir de informações próprias e coletadas de documentos obtidos publicamente a partir da Internet. Este material contém ilustrações adquiridas de bancos de imagens de origem privada ou pública, não possuindo a intenção de violar qualquer direito pertencente à terceiros e sendo voltado para fins acadêmicos ou meramente ilustrativos. Portanto, os textos, fotografias, imagens, logomarcas e sons presentes nesta apresentação se encontram protegidos por direitos autorais ou outros direitos de propriedade intelectual.

Ao usar este material, o usuário deverá respeitar todos os direitos de propriedade intelectual e industrial, os decorrentes da proteção de marcas registradas da mesma, bem como todos os direitos referentes a terceiros que por ventura estejam, ou estiveram, de alguma forma disponíveis nos slides. O simples acesso a este conteúdo não confere ao usuário qualquer direito de uso dos nomes, títulos, palavras, frases, marcas, dentre outras, que nele estejam, ou estiveram, disponíveis.

É vedada sua utilização para finalidades comerciais, publicitárias ou qualquer outra que contrarie a realidade para o qual foi concebido. Sendo que é proibida sua reprodução, distribuição, transmissão, exibição, publicação ou divulgação, total ou parcial, dos textos, figuras, gráficos e demais conteúdos descritos anteriormente, que compõem o presente material, sem prévia e expressa autorização de seu titular, sendo permitida somente a impressão de cópias para uso acadêmico e arquivo pessoal, sem que sejam separadas as partes, permitindo dar o fiel e real entendimento de seu conteúdo e objetivo. Em hipótese alguma o usuário adquirirá quaisquer direitos sobre os mesmos.

O usuário assume toda e qualquer responsabilidade, de caráter civil e/ou criminal, pela utilização indevida das informações, textos, gráficos, marcas, enfim, todo e qualquer direito de propriedade intelectual ou industrial deste material.



PUC Minas
Virtual

© PUC Minas • Todos os direitos reservados, de acordo com o art. 184 do Código Penal e com a lei 9.610 de 19 de fevereiro de 1998.
Proibidas a reprodução, a distribuição, a difusão, a execução pública, a locação e quaisquer outras modalidades de utilização sem a devida autorização da Pontifícia Universidade Católica de Minas Gerais.