



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Instituto de Ciências Exatas e de Informática

Marcos Antonio Lommez Candido Ribeiro¹

Lista #5

Inteligência Artificial

¹Aluno de Graduação em Ciência da Computação – tonilommez@hotmail.com

1) Considere que em um determinado supermercado foram efetuadas as seguintes transações:

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Utilizando-se o algoritmo Apriori, um suporte mínimo aceitável de 0.3 e confiança de 0.8, o número de ItensSets 1, 2, 3 e de regras a partir desta base de dados são:

Produto	Suporte
Leite	$2/10 = 0.2$
Cafe	$3/10 = 0.3$
Cerveja	$2/10 = 0.2$
Pao	$5/10 = 0.5$
Manteiga	$5/10 = 0.5$
Arroz	$2/10 = 0.2$
Feijao	$2/10 = 0.2$

ItemSet = 1	
Produto	Suporte
Cafe	$3/10 = 0.3$
Pao	$5/10 = 0.5$
Manteiga	$5/10 = 0.5$

ItemSet = 2	
Produto	Suporte
Cafe e Pao	3/10 = 0.3
Cafe e Manteiga	3/10 = 0.3
Pao e Manteiga	4/10 = 0.4

ItemSet = 3	
Cafe, Pao e manteiga	3/10 = 0.3

Regras		
Se	Entao	Confianca
Cafe	Pao	3/3
Cafe	Manteiga	3/3
Pao	Manteiga	4/5
Manteiga	Pao	4/5
Cafe e Pao	Manteiga	3/3
Cafe e Manteiga	Pao	3/3
Cafe	Pao e Manteiga	3/3

2) Considerando-se o código que está em Módulos/Apriori.ipynb, rode o código com a base acima e confira os resultados.

Resposta na questao 3

3) Considerando-se o código que está em Módulos/Apriori.ipynb, altere-o para que ele imprima os temsets gerados, com os respectivos suportes

	Antecedente	Consequente	suporte	confianca	lift
4	[Cafe]	[Pao, Manteiga]	0.3	1.0	2.5
1	[Cafe]	[Pao]	0.3	1.0	2.0
0	[Cafe]	[Manteiga]	0.3	1.0	2.0
6	[Cafe, Pao]	[Manteiga]	0.3	1.0	2.0
5	[Cafe, Manteiga]	[Pao]	0.3	1.0	2.0
2	[Manteiga]	[Pao]	0.4	0.8	1.6
3	[Pao]	[Manteiga]	0.4	0.8	1.6

4) Considerando-se o código que está em Módulos/Apriori.ipynb, altere-o para que ele gere regras de associação quando não há presença do produto. Ou seja, gostaria de ver regras da seguinte forma:

Quem não leva álcool leva detergente;

Quem não leva detergente leva arroz, etc

A resposta que eu proponho seria criar novas colunas que funcionariam como "nao" de um objeto, exemplo a coluna "cerveja" seria transformada em duas colunas, uma original a original e a outra inversa com "nao_cerveja". A partir disso basta rodar o algoritmo normalmente e o resultado final seria assim:

	Antecedente	Consequente	Suporte	Confiança	Lift
0	[Manteiga]	[NOT_Leite]	0.375	1.0	1.0
1	[Pao]	[NOT_Leite]	0.375	1.0	1.0
2	[Cafe]	[NOT_Cerveja]	0.375	1.0	1.0
3	[Manteiga]	[NOT_Cerveja]	0.500	1.0	1.0
4	[Pao]	[NOT_Cerveja]	0.500	1.0	1.0
...
62	[NOT_Feijao, Cafe]	[Pao, Manteiga]	0.375	1.0	2.0
63	[Pao, Cafe]	[Manteiga, NOT_Feijao]	0.375	1.0	1.6
64	[Manteiga, NOT_Feijao, Cafe]	[Pao]	0.375	1.0	1.6
65	[Pao, Manteiga, Cafe]	[NOT_Feijao]	0.375	1.0	1.0
66	[Pao, NOT_Feijao, Cafe]	[Manteiga]	0.375	1.0	1.6

A partir disso basta filtrar as regras como quiser. Para ver aquelas onde o antecedente NAO leva a um caso de SIM, basta filtra-las dentre a lista geral gerada.

5) Investigue o funcionamento da biblioteca mlxtend para geração de regras de associação.

A biblioteca mlxtend é uma ferramenta de mineração de dados que oferece funcionalidades abrangentes para a geração de regras de associação, incluindo métodos como Apriori e FP-growth e Eclat.

Para usar o algoritmo Apriori na geração de conjuntos de itens frequentes, você normalmente começaria convertendo seus dados de transação em um DataFrame com codi-

ficação one-hot. Depois por exemplo, você poderia usar a função Apriori para encontrar conjuntos de itens frequentes que atendam a um limite mínimo de suporte especificado usando parametros como o suporte minimo (minsup) e confianca minima (minconf). Após a conclusão da execução, a biblioteca disponibiliza um conjunto de regras de associação, completas com métricas como suporte e confiança, que ajudam na interpretação dos resultados.

6) Faça uma resenha do artigo “A Literature Survey on Association Rule Mining Algorithms.pdf” que está no CANVAS.

O artigo apresenta uma análise abrangente de algoritmos na área de mineração de regras de associação (ARM). O texto mostra como os algoritmos são usados na prática e dentro de quais ramos, como por exemplo na análise de cestas de supermercados, diagnóstico médicos, sensores e etc. A partir disso o artigo explica como os algoritmos são em seus termos em comuns, se baseando em duas medidas bases, sendo estas “suporte” e “confiança”.

Inicialmente os algoritmos são divididos e explicados em algumas grandes classificações, aqueles baseados em layouts horizontais sendo o principal o Apriori, layouts verticais e baseados em dividir para conquistar (projected layout). Além disso, explora algoritmos mais específicos, como FSG e GSPAN, que têm aplicações em encontrar subgrafos e subestruturas frequentes.

O artigo também discute algoritmos específicos, como DIC e DHP. O algoritmo DIC reduz o número de transições no banco de dados ao particioná-lo em intervalos de tamanho fixo. O algoritmo DHP, por outro lado, emprega uma técnica de hashing para restringir o número de conjuntos de itens candidatos, gerar com eficiência grandes conjuntos de itens e reduzir o tamanho do banco de dados de transações.

Em suma, o artigo é um recurso para fornecer um panorama geral dos algoritmos de mineração de regras de associação detalhando métricas, funcionalidades e características de cada grupo de algoritmos.