

ALGORITMO APRIORI*

(AGRAWAL AND SRIKANT, 1994)

Cristiane Neri Nobre

Agrawal, R. and Srikant, R. (1994) Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, 12-15 September 1994, 487-499

Regras de Associação com o algoritmo Apriori

- **Duas fases:**

1. Gerar os *item sets* com o **suporte** mínimo especificado;
2. Para cada *itemset*, gerar as regras com a **confiança** mínima especificada.

Regras de Associação com o algoritmo Apriori

Exemplo 1: Prateleiras de mercado

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

Regras de Associação com o algoritmo Apriori

Vamos calcular o suporte dos itemSets?

Lembrando que *suporte* de uma regra é o número de transações que contem todos os itens da transação dividido pelo número total de transações

Regras de Associação com o algoritmo Apriori – Calculando o suporte

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

Produto	Suporte
Leite	2/10 = 0.2
Café	3/10 = 0.3
Cerveja	2/10 = 0.2
Pão	5/10 = 0.5
Manteiga	5/10 = 0.5
Arroz	2/10 = 0.2
Feijão	2/10 = 0.2

Regras de Associação com o algoritmo Apriori

Considerando o suporte ≥ 0.3

Como serão os valores para os ItemSets?

Regras de Associação com o algoritmo Apriori

– Geração dos itensSet

ItemSet = 1?

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

Produto	Suporte
Leite	2/10 = 0.2
Café	3/10 = 0.3
Cerveja	2/10 = 0.2
Pão	5/10 = 0.5
Manteiga	5/10 = 0.5
Arroz	2/10 = 0.2
Feijão	2/10 = 0.2

Regras de Associação com o algoritmo Apriori

– Geração dos itensSet

ItemSet =2?

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

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4

Regras de Associação com o algoritmo Apriori

– Geração dos itensSet

ItemSet = 3?

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

Produto	Suporte
Café, pão e manteiga	3/10 = 0.3

Regras de Associação com o algoritmo Apriori

Segunda fase

Como gerar as regras a partir dos itemSets?

Para cada *itemSet* faça:

- ✓ Gere todas as regras possíveis
- ✓ Mantenha somente as regras com a **confiança mínima** exigida

Regras de Associação com o algoritmo Apriori

Lembrando que o cálculo de confiança é:

$$\text{Confiança } (A \rightarrow B) = \frac{P(A \cup B)}{P(A)} = \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$$

Regras de Associação com o algoritmo Apriori

Como gerar as regras a partir dos itemSets =2?

		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
Produto	Suporte	4	Sim	Sim	Não	Sim	Sim	Não	Não
		5	Não	Não	Sim	Não	Não	Não	Não
Café e pão	3/10 = 0.3	6	Não	Não	Não	Não	Sim	Não	Não
		7	Não	Não	Não	Sim	Não	Não	Não
Café e manteiga	3/10 = 0.3	8	Não	Não	Não	Não	Não	Não	Sim
		9	Não	Não	Não	Não	Não	Sim	Sim
Pão e Manteiga	4/10 = 0.4	10	Não	Não	Não	Não	Não	Sim	Não

Regras de Associação com o algoritmo Apriori

Qual a confiança das regras com ItemSets =2

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4
Produto	Suporte
Café	3/10
Pão	5/10
Manteiga	5/10

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

{Café, pão}

Se Café → Pão

confiança = $3/3 = 1$

Se Pão → Café

confiança = $3/5 = 0.6$

Regras de Associação com o algoritmo Apriori

Qual a confiança das regras com ItemSets =2

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4
Produto	Suporte
Café	3/10
Pão	5/10
Manteiga	5/10

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

{Café, manteiga}

Se Café → manteiga

confiança = $3/3 = 1$

Se Manteiga → Café

confiança = $3/5 = 0.6$

Regras de Associação com o algoritmo Apriori

Qual a confiança das regras com $\text{ItemSets} = 2$

Produto	Suporte
Café e pão	$3/10 = 0.3$
Café e manteiga	$3/10 = 0.3$
Pão e Manteiga	$4/10 = 0.4$
Produto	Suporte
Café	$3/10$
Pão	$5/10$
Manteiga	$5/10$

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

{Pão, manteiga}

Se Pão → manteiga

confiança = $4/5 = 0.8$

Se Manteiga → Pão

confiança = $4/5 = 0.8$

Regras de Associação com o algoritmo Apriori

Se assumirmos confiança ≥ 0.8 , teremos as seguintes regras para os itemSets=2

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

Se Café \rightarrow Pão

confiança = $3/3 = 1$

Se Café \rightarrow manteiga

confiança = $3/3 = 1$

Se Pão \rightarrow manteiga

confiança = $4/5 = 0.8$

Se Manteiga \rightarrow Pão

confiança = $4/5 = 0.8$

~~Se Manteiga \rightarrow Café~~

~~confiança = $3/5 = 0.6$~~

~~Se Pão \rightarrow Café~~

~~confiança = $3/5 = 0.6$~~

Regras de Associação com o algoritmo Apriori

Como gerar as regras a partir dos itemSets = 3?

Produto	Suporte
Café, pão e manteiga	3/10 = 0.3

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

{Café, pão, manteiga}

Se Café e pão → manteiga

Se Café e manteiga → pão

Se pão e manteiga → café

Se café → pão e manteiga

Se pão → café e manteiga

Se manteiga → café e pão

Regras de Associação com o algoritmo Apriori

Como calcular a confiança das regras dos itemSets = 3?

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4
Produto	Suporte
Café, pão e manteiga	3/10 = 0.3

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

{Café, pão, manteiga}

Se Café e pão → manteiga

confiança = $3/3 = 1$

Se Café e manteiga → pão

confiança = $3/3 = 1$

Se pão e manteiga → café

confiança = $3/4 = 0.75$

Regras de Associação com o algoritmo Apriori

Como calcular a confiança das regras dos itemSets = 3?

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4
Produto	Suporte
Café, pão e manteiga	3/10 = 0.3

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

{Café, pão, manteiga}

Se café → pão e manteiga

confiança = $3/3 = 1$

Se pão → café e manteiga

confiança = $3/5 = 0.6$

Se manteiga → café e pão

confiança = $3/5 = 0.6$

Regras de Associação com o algoritmo Apriori

Quais as regras iremos selecionar, considerando a confiança ≥ 0.8

Num	Regras	Confiança
1	Se Café \rightarrow Pão	$3/3 = 1$
2	Se Café \rightarrow manteiga	$3/3 = 1$
3	Pão \rightarrow manteiga	$4/5 = 0.8$
4	Manteiga \rightarrow Pão	$4/5 = 0.8$
5	Se Café e pão \rightarrow manteiga	$3/3 = 1$
6	Se Café e manteiga \rightarrow pão	$3/3 = 1$
7	Se café \rightarrow pão e manteiga	$3/3 = 1$

Regras de Associação com o algoritmo Apriori

Como avaliar as melhores regras a partir destas 7 regras?

Iremos calcular o Lift da regra!

Regras de Associação com o algoritmo Apriori

Iremos calcular o Lift da regra: $Lift = \frac{confiança(A \rightarrow B)}{suporte(B)}$

Regras	Confiança	Lift
Se Café → Pão	3/3 = 1	1/0.5=2
Se Café → manteiga	3/3 = 1	1/0.5 = 2
Pão → manteiga	4/5 = 0.8	0.8/0.5 = 1.6
Manteiga → Pão	4/5 = 0.8	0.8/0.5=1.6
Se Café e pão → manteiga	3/3 = 1	1/0.5= 2
Se Café e manteiga → pão	3/3 = 1	1/0.5=2
Se café → pão e manteiga	3/3 = 1	1/0.4=2.5

Produto	Suporte
Leite	2/10 = 0.2
Café	3/10 = 0.3
Cerveja	2/10 = 0.2
Pão	5/10 = 0.5
Manteiga	5/10 = 0.5
Arroz	2/10 = 0.2
Feijão	2/10 = 0.2

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4

Regras de Associação com o algoritmo Apriori

Qual o significado do lift?

Regras	Confiança	Lift	Significado
Se Café → Pão	$3/3 = 1$	$1/0.5=2$	Clientes que compram café tem duas vezes mais chances de comprarem pão

Valor de **lift** igual a 1 para a regra ($A \rightarrow B$) indica que A e B são **independentes**

Um valor de **lift superior a 1** indicam que A e B aparecem mais frequentemente juntos do que esperado. Isso significa que a ocorrência de A tem um efeito positivo sobre a ocorrência de B.

Um valor de **lift menor do que 1** indica que A e B aparecem com menos frequência do que o esperado em conjunto, indicando que a ocorrência de A tem um efeito negativo sobre a ocorrência de B.

Regras de Associação com o algoritmo Apriori

Qual o significado do lift?

Regras	Confiança	Lift	Significado
Se Café → Pão	$3/3 = 1$	$1/0.5=2$	Clientes que compram café tem duas vezes mais chances de comprarem pão

Valor de **lift** igual a 1 para a regra ($A \rightarrow B$) indica que A e B são **independentes**

Um valor de **lift superior a 1** indicam que A e B aparecem mais frequentemente juntos do que esperado. Isso significa que a ocorrência de A tem um efeito positivo sobre a ocorrência de B.

Um valor de **lift menor do que 1** indica que A e B aparecem com menos frequência do que o esperado em conjunto, indicando que a ocorrência de A tem um efeito negativo sobre a ocorrência de B.

Regras de Associação com o algoritmo Apriori

- O problema principal dos algoritmos de regras de associação é que o especialista precisa dar valor às regras geradas
- Para isso, pode ser necessário jogar atributos fora, aqueles que aparecem com muitas frequência, por exemplo.

Regras de Associação com o algoritmo Apriori

Investiguem outros algoritmos de regras de associação:
eclat, fp-grow!

Neste vídeo, aprendemos:

1. O funcionamento do algoritmo Apriori

No próximo vídeo, veremos o funcionamento deste algoritmo no **ambiente WEKA**