

# 02-05

## Algorithme Apriori

420-A58-SF — Algorithmes d'apprentissage non supervisé — Été 2021  
Spécialisation technique en intelligence artificielle — M. Swawola, M.Sc.

---

**NOUS ÉCLAIRONS.  
VOUS BRILLEZ.**

---

FORMATION CONTINUE  
ET SERVICES AUX ENTREPRISES



# Sommaire

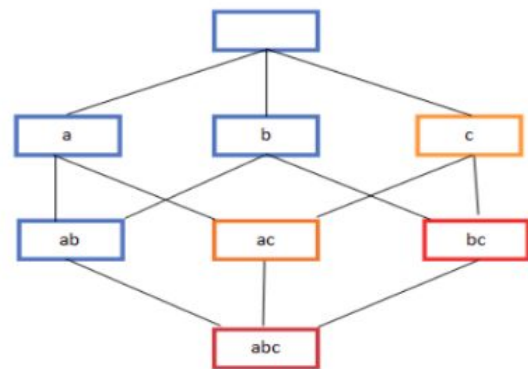
1. Algorithme Apriori
2. Pour aller plus loin ...
3. Lectures et références

# Sommaire

1. Algorithme Apriori
2. Pour aller plus loin ...
3. Lectures et références

# Algorithme Apriori

- L'algorithme Apriori est une approche nécessitant deux passages sur les données pour trouver les **itemsets fréquents**
- Se base sur la **monotonie**
  - Si un ensemble d'items  $I$  apparaît au moins  $s$  fois, alors chaque sous-ensemble  $J$  de  $I$  aussi
- **Contraposée pour les paires**
  - Si l'item  $i$  n'apparaît pas dans  $s$  baskets, alors aucune paire incluant  $i$  ne peut apparaître dans  $s$  baskets



# Algorithme Apriori

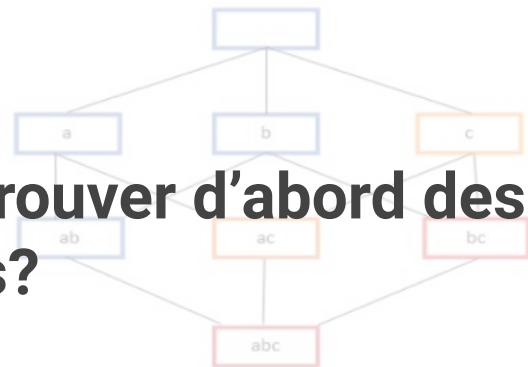
- L'algorithme Apriori est une approche nécessitant deux passages sur les données pour trouver les **itemsets fréquents**

- Se base sur la **monotonie**

**À partir de ces constatations, comment trouver d'abord des paires d'items fréquents?**

- **Contraposée pour les paires**

- Si l'item  $i$  n'apparaît pas dans  $s$  baskets, alors aucune paire incluant  $i$  ne peut apparaître dans  $s$  baskets



# Algorithme Apriori - Étape #1

- Lire tous les baskets et compter le nombre de chaque items
- Les items apparaissant au moins  $s$  fois sont considérés **fréquents**

$B_1 = \{\text{lait, coca-cola, bière}\}$	$B_2 = \{\text{lait, pepsi, jus}\}$
$B_3 = \{\text{lait, bière}\}$	$B_4 = \{\text{coca-cola, jus}\}$
$B_5 = \{\text{lait, pepsi, bière}\}$	$B_6 = \{\text{lait, coca-cola, bière, jus}\}$
$B_7 = \{\text{lait, coca-cola, bière, jus}\}$	$B_8 = \{\text{bière, coca-cola}\}$



<b>lait</b>	5
<b>coca-cola</b>	5
<b>bière</b>	6
<b>pepsi</b>	2
<b>jus</b>	4

# Algorithme Apriori - Étape #1

- Lire tous les baskets et compter le nombre de chaque items
- Les items apparaissant au moins  $s$  fois sont considérés **fréquents**

$B_1 = \{\text{lait, coca-cola, bière}\}$	$B_2 = \{\text{lait, pepsi, jus}\}$
$B_3 = \{\text{lait, bière}\}$	$B_4 = \{\text{coca-cola, jus}\}$
$B_5 = \{\text{lait, pepsi, bière}\}$	$B_6 = \{\text{lait, coca-cola, bière, jus}\}$
$B_7 = \{\text{lait, coca-cola, bière, jus}\}$	$B_8 = \{\text{bière, coca-cola}\}$

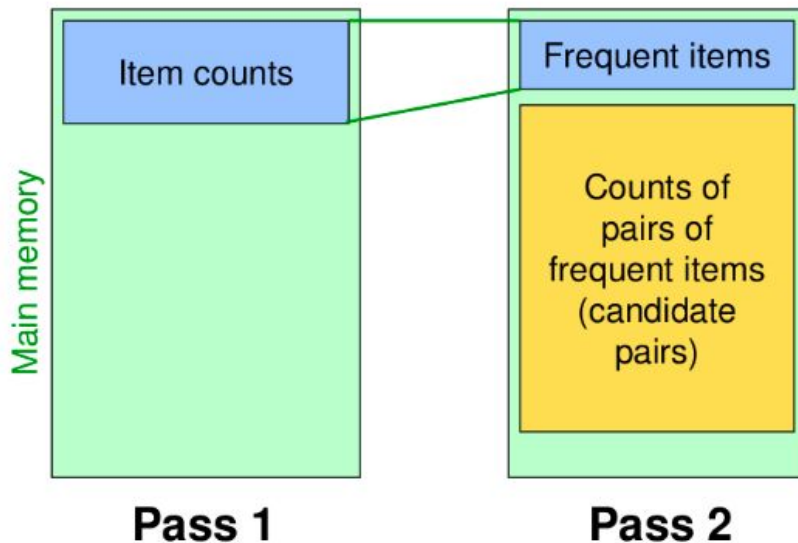


<b>lait</b>	5
<b>coca-cola</b>	5
<b>bière</b>	6
<b>pepsi</b>	2
<b>jus</b>	4

Si  $s = 3$

# Algorithme Apriori - Étape #2

- Lire à **nouveau** tous les baskets et compter le nombre de paires (2-tuple) dont chaque élément est fréquent

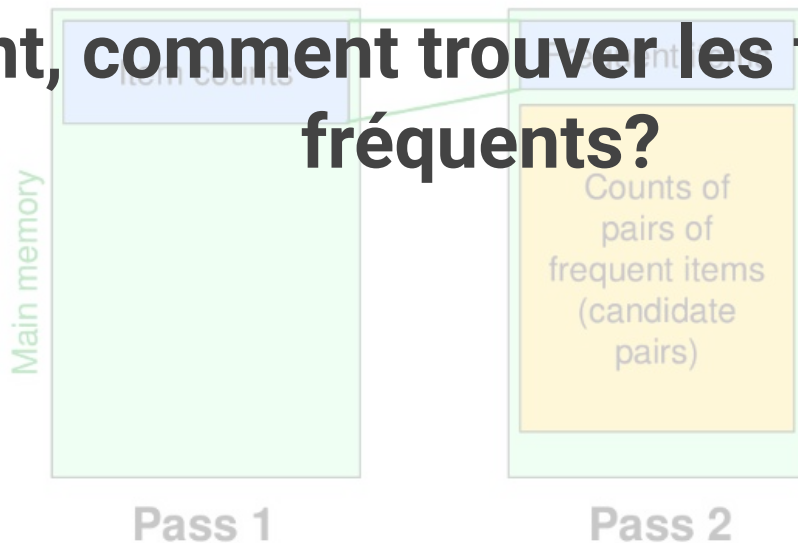




# Algorithme Apriori - Étape #2

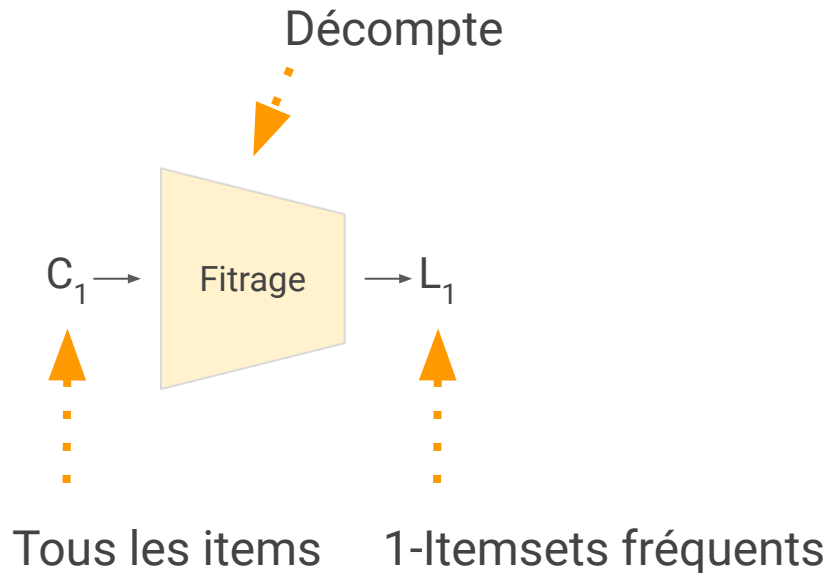
- Lire à **nouveau** tous les baskets et compter le nombre de paires (2-tuple) dont chaque élément est fréquent

**Maintenant, comment trouver les triplets d'items fréquents?**



# Triplets fréquents, etc ...

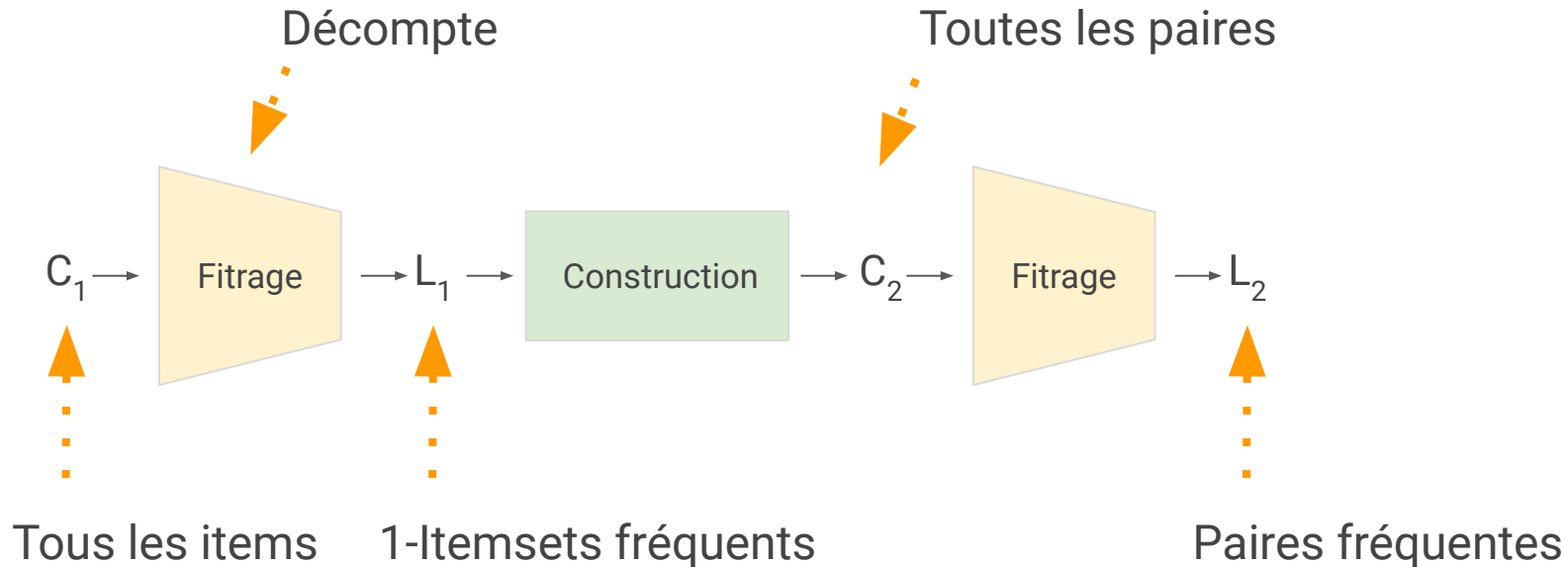
- Pour chaque  $k$ , construction de deux ensembles:  $C_k$  et  $L_k$



- $C_k = k$ -itemsets **candidats**
- $L_k = k$ -itemsets **réellement fréquents**

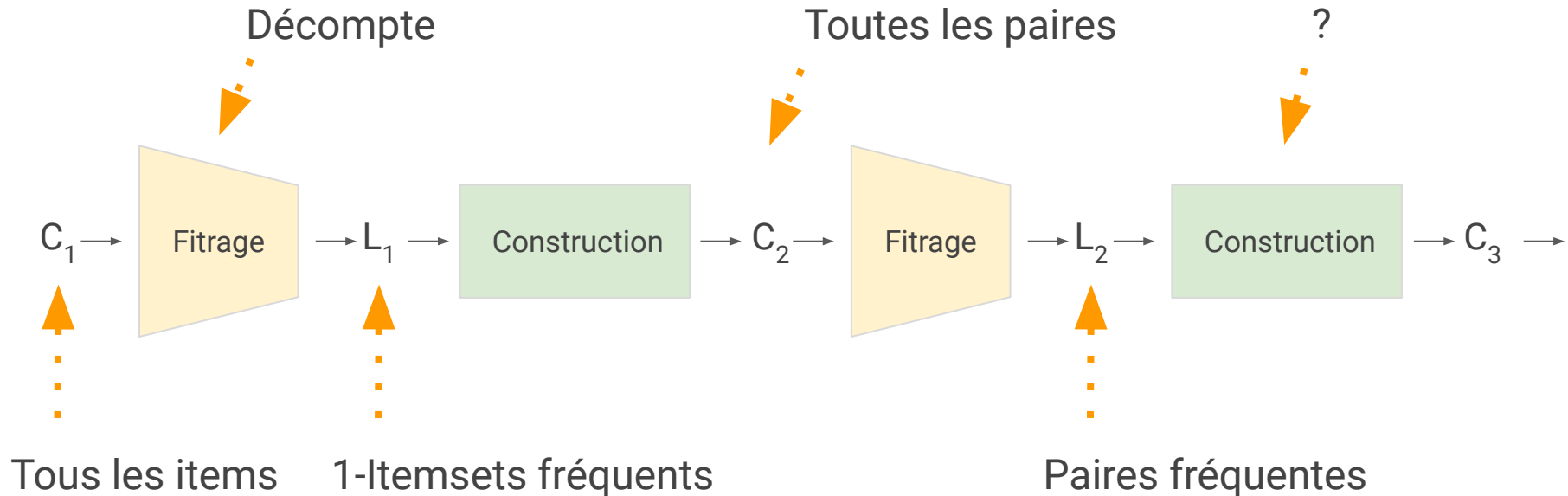
# Triplets fréquents, etc ...

- Pour chaque  $k$ , construction de deux ensembles:  $C_k$  et  $L_k$



# Triplets fréquents, etc ...

- Pour chaque  $k$ , construction de deux ensembles:  $C_k$  et  $L_k$



# Triplets fréquents, etc ...

## ■ Étapes de l'algorithme Apriori

- $C_1 = \{ \{b\}, \{c\}, \{j\}, \{l\}, \{p\} \}$
- $L_1 = \{ \{b\}, \{c\}, \{j\}, \{l\} \}$
- $C_2 = \{ \{b,c\}, \{b,j\}, \{b,l\}, \{c,j\}, \{c,l\}, \{l,j\} \}$
- $L_2 = \{ \{b,l\}, \{b,c\}, \{c,l\}, \{c,j\} \}$
- **$C_3 = ?$**

# Triplets fréquents, etc ...

## ■ Étapes de l'algorithme Apriori

- $C_1 = \{ \{b\}, \{c\}, \{j\}, \{l\}, \{p\} \}$
- $L_1 = \{ \{b\}, \{c\}, \{j\}, \{l\} \}$
- $C_2 = \{ \{b,c\}, \{b,j\}, \{b,l\}, \{c,j\}, \{c,l\}, \{l,j\} \}$
- $L_2 = \{ \{b,l\}, \{b,c\}, \{c,l\}, \{c,j\} \}$
- $C_3 = \{ \{b,c,l\}, \{b,c,j\}, \{b,l,j\}, \{c,l,j\} \}$

Pour obtenir  $C_3$ , "combiner"  $L_2$   
et  $L_1$

# Triplets fréquents, etc ...

## ■ Étapes de l'algorithme Apriori

- $C_1 = \{ \{b\}, \{c\}, \{j\}, \{l\}, \{p\} \}$
- $L_1 = \{ \{b\}, \{c\}, \{j\}, \{l\} \}$
- $C_2 = \{ \{b,c\}, \{b,j\}, \{b,l\}, \{c,j\}, \{c,l\}, \{l,j\} \}$
- $L_2 = \{ \{b,l\}, \{b,c\}, \{c,l\}, \{c,j\} \}$
- $C_3 = \{ \{b,c,l\}, \{b,c,j\}, \{b,l,j\}, \{c,l,j\} \}$
- $L_3 = \{ \{b,c,l\} \}$

Pour obtenir  $C_3$ , "combiner"  $L_2$   
et  $L_1$

D'une manière générale, pour obtenir  
 $C_k$ , "combiner"  $L_{k-1}$  et  $L_1$

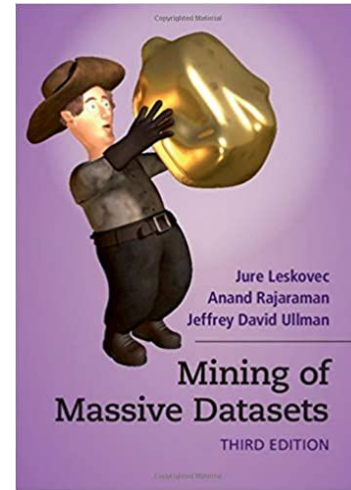
# Sommaire

1. Algorithme Apriori
2. Pour aller plus loin ...
3. Lectures et références



# Pour aller plus loin ...

- Algorithme de Park, Chen, et Yu (PCY)
- Algorithme multistages
- Algorithme multihash
- Algorithme SON
- etc ...



6 Frequent Itemsets  
p. 213-251



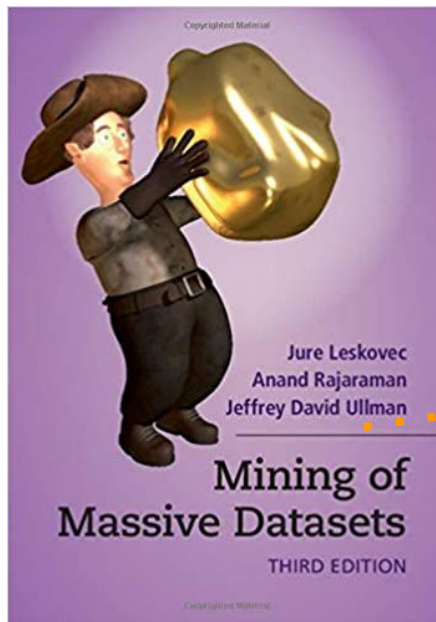
<https://github.com/mswawola-cegep/420-a58-sf.git>

**02-05-A1 et 02-05-A2**

# Sommaire

1. Algorithme Apriori
2. Pour aller plus loin ...
3. Lectures et références

# Lectures



6 Frequent Itemsets  
p. 213-251

Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, **Mining of Massive Datasets, 3rd edition**

# Références

[1] [Mining of Massive Datasets, 3rd edition](#)