

Frequent Itemsets

Stefan Schaefer, Nils Rethmeier

Motivation



Task: Find items that frequently occur together.

Example: In a store, what *items* do people buy together?

milk & cornflakes (expected)







diapers & beer (unexpected)

^{1 |} SDAA | Frequent Itemsets | Schaefer, Rethmeier | 15. November 2011



Concepts

$$i_k := item$$

$$B := \text{Basket}$$
, a tuple of items i_k

I := itemset $I = \{i_1, i_2, \dots, i_n\}$

n := set size

Example

(milk, corn flakes, bread, eggs) (milk, corn flakes, beer, diapers, chips) (milk, bread, bread, butter, cheese, salad) (milk, corn flakes, beans, bacon)

...

$$I = {\text{milk}}, I = {\text{milk, corn flakes}}$$

 $n = 1, \qquad n = 2$



Concepts

Example

$$s(I) := \text{support/count}$$

of an items set I

$$s(\{\text{milk}\}) = 4$$

 $s(\{\text{milk, corn flakes}\}) = 3$
 $s(\{\text{milk, corn flakes, bread}\}) = 1$

$$s := \text{support threshold}$$

$$s = 3$$

I is a Frequent Itemset
$$I_F$$
 if $s(I) \geq s$.

$$I_F = {\text{milk}}$$

 $I_F = {\text{milk, corn flakes}}$



Concepts

Monotonicity

If *I* is not frequent, then its supersets cannot be frequent either

Example

```
{milk, corn flakes, bread}
is not frequent →
{milk, cornflakes, bread, eggs}
is not frequent
```

```
Frequent Items
Frequent Pairs
Frequent Triples
...
Frequent N-Tupels
```

```
s(\{\text{milk}\}) = 4

s(\{\text{milk, corn flakes}\}) = 3

s(\{\text{milk, corn flakes, bread}\}) = 1
```

A-Priori Algorithm



Given: a list of baskets with items

(milk, corn flakes, bread, eggs) (milk, corn flakes, beer, diapers, chips) (milk, bread, bread, butter, cheese, salad) (milk, corn flakes, beans, bacon)

S.
3 x {bread} Frequent Items

1st pass: find **frequent items** in the baskets.

■ 4 x {milk}, 3 x {corn flakes}, 3 x {bread}

2nd pass: find **frequent pairs** based on frequent items.

■ 3 x {milk, corn flakes}

. . .

n-th pass: find frequent Itemsets of size *n* based on frequent *n-1* itemsets.

Frequent Pairs

Frequent N-Tupels

A-Priori Algorithm



Given: a list of baskets with items

Count Itemsets I of size 1 (items) and filter result against threshold s

	All Items			Frequent Items	
	I	s(I)		l _F	s(I)
count items in the baskets	{milk} {corn flakes} }{eggs} {bread} 	4 3 1 3	filter I if s(I)<=3	{milk} {corn flakes} {bread} 	4 3 3

Count Itemsets I of size 2 (pairs) and filter the results

	Itemsets		Itemset suppo	rts		Frequent Itemsets	
	I		1	s(I)		l _F	s(l)
create pairs using Freq. Items from before	{milk, flakes} }{milk, bread} {flakes, bread} 	determ \ ine s(l) for the pairs	{milk, flakes} }{milk, bread}	3	filter non- freq.	(milk, flakes)	3
			{flakes, bread}	2			

SON Algorithm



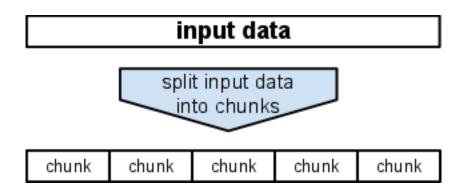
Algorithm of Savasere, Omiecinski, Navathe (SON Algorithm)

Problem

- input data too large for RAM
 - swapping slows down processing

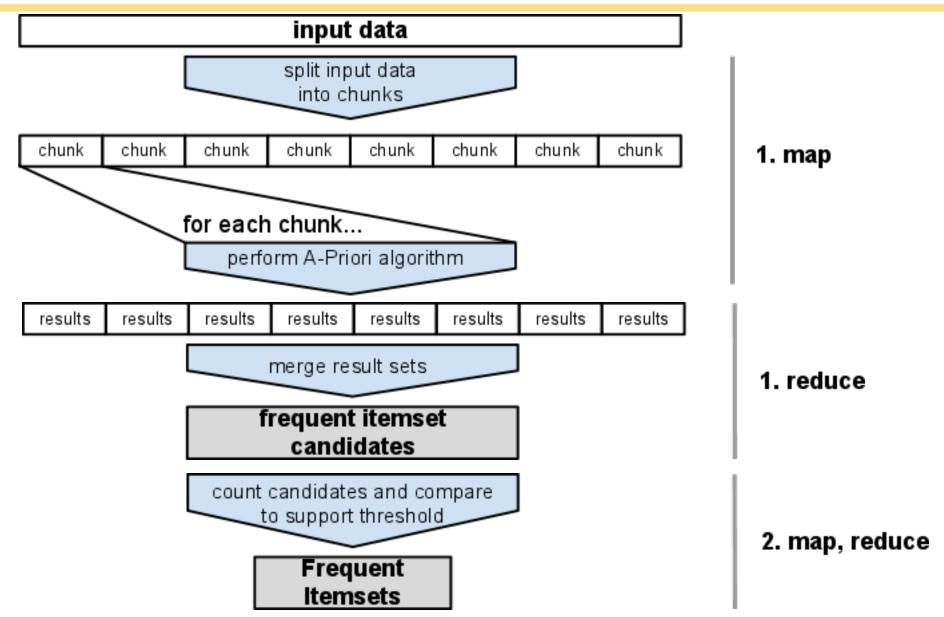
Solution (Data parallelism)

- split input data into chunks
- o process each chunk in parallel
 - Map & Reduce



SON Algorithm





Summary

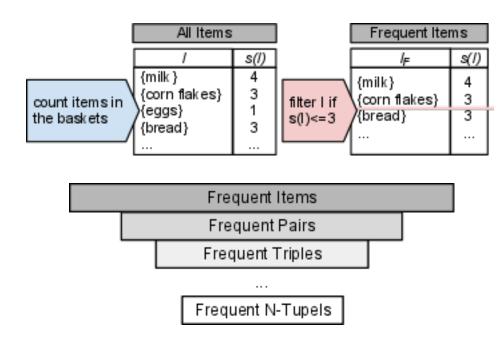


Frequent Itemset applications

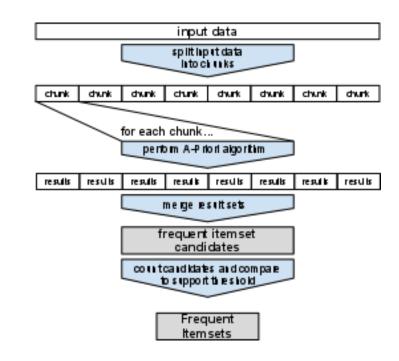
- market analysis
- plagiarism detection
- related concepts



A Priori Algorithm



SON Algorithm





Thank you.





Concepts

$i_k := item$

B := Basket, a tuple of items i_k

Example

{milk, corn flakes, bread, eggs} {milk, corn flakes, beer, diapers, chips} {milk, bread, bread, butter, cheese, salad} {milk, corn flakes, beans, bacon}



Concepts

Example

$$s(I) := \text{support/count}$$

of an items set I

$$s(\{\text{milk}\}) = 4$$

 $s(\{\text{milk, corn flakes}\}) = 3$
 $s(\{\text{milk, corn flakes, bread}\}) = 1$

$$s := \text{support threshold}$$

$$s = 3$$

I is a Frequent Itemset
$$I_F$$
 if $s(I) \geq s$.

$$I_F = \{ \text{milk} \}$$
 $I_F = \{ \text{milk}, \text{corn flakes} \}$



Concepts

$$i_k := item$$

$$B := \text{Basket}$$
, a tuple of items i_k

I := itemset

$$I = \{i_1, i_2, \dots, i_n\}$$

$$n := \text{set size}$$

Example

{milk, corn flakes, bread, eggs}
{milk, corn flakes, beer, diapers, chips}
{milk, bread, bread, butter, cheese, salad}
{milk, corn flakes, beans, bacon}

...

$$I = {\text{milk}}, I = {\text{milk, corn flakes}}$$

 $n = 1, \qquad n = 2$