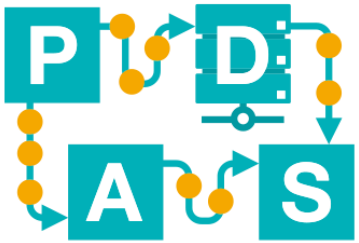
Sequence Mining - Instruction

Lecture 13

IDS-L13-I



Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Sequence Mining

In the Sequence Mining lecture you have seen AprioriAll, an algorithm for mining sequences.

It can be considered a generalized version of Apriori.

Apriori vs. AprioriAll

Can you tell the differences between Apriori and AprioriAll?

- **What are the goals of the two algorithms?**
- **What is the definition of pattern of each algorithm?**
- **Where do the two algorithms look for patterns?**

Apriori vs. AprioriAll

Answers!

- **Apriori looks at items grouped together.**
- **AprioriAll looks at items with a specific partial ordering.**

Apriori vs. AprioriAll

Answers!

- Apriori looks at items grouped together.
- Patterns are sets of items.

- AprioriAll looks at items with a specific partial ordering.
- Patterns are sequences of sets of items.

Apriori vs. AprioriAll

Answers!

- Apriori looks at items grouped together.
- Patterns are sets of items.
- Apriori finds patterns within sets.

- AprioriAll looks at items with a specific partial ordering.
- Patterns are sequences of sets of items.
- AprioriAll finds patterns between sets.

Support

A central concept in pattern mining is *support* (and *support count*): the frequency of appearance (relative or absolute) of a certain pattern within the database.

Support

In this database, find the support count of:

(bc)(de)

b(de)

(bc)de

(ac)(bc)

(bc)(ac)

D = [

<a(bc)d(eb)>,

<(ac)(bc)de>,

<(ac)b(cd)>,

<ab(bc)(cde)>,

<(bc)(bd)(bde)>,

<(abc)(ac)(bc)de>,

<a(bd)c(de)>

]

Support: solutions

In this database, find the support of:

(bc)(de): 2

b(de)

(bc)de

(ac)(bc)

(bc)(ac)

D = [

<a(bc)d(eb)>,

<(ac)(bc)de>,

<(ac)b(cd)>,

<ab(bc)(cde)>,

<(bc)(bd)(bde)>,

<(abc)(ac)(bc)de>,

<a(bd)c(de)>

]

Support: solutions

In this database, find the support of:

(bc)(de): 2

b(de): 3

(bc)de

(ac)(bc)

(bc)(ac)

D = [

<a(bc)d(eb)>,

<(ac)(bc)de>,

<(ac)b(cd)>,

<ab(bc)(cde)>,

<(bc)(bd)(bde)>,

<(abc)(ac)(bc)de>,

<a(bd)c(de)>

]

Support: solutions

In this database, find the support of:

(bc)(de): 2

b(de): 3

(bc)de: 4

(ac)(bc)

(bc)(ac)

D = [

<a(bc)d(eb)>,

<(ac)(bc)de>,

<(ac)b(cd)>,

<ab(bc)(cde)>,

<(bc)(bd)(bde)>,

<(abc)(ac)(bc)de>,

<a(bd)c(de)>

]

Support: solutions

In this database, find the support of:

(bc)(de): 2

b(de): 3

(bc)de: 4

(ac)(bc): 2

(bc)(ac)

D = [

<a(bc)d(eb)>,

<(ac)(bc)de>,

<(ac)b(cd)>,

<ab(bc)(cde)>,

<(bc)(bd)(bde)>,

<(abc)(ac)(bc)de>,

<a(bd)c(de)>

]

Support: solutions

In this database, find the support of:

(bc)(de): 2

b(de): 3

(bc)de: 4

(ac)(bc): 2

(bc)(ac): 1

D = [

<a(bc)d(eb)>,

<(ac)(bc)de>,

<(ac)b(cd)>,

<ab(bc)(cde)>,

<(bc)(bd)(bde)>,

<(abc)(ac)(bc)de>,

<a(bd)c(de)>

]