

Ex - Sequential Pattern

Sequential Pattern – Exercise 1

a) (3 points) Given the following input sequence

$\langle \begin{matrix} \{A\} \\ t=0 \end{matrix} \quad \begin{matrix} \{B,F\} \\ t=1 \end{matrix} \quad \begin{matrix} \{E\} \\ t=2 \end{matrix} \quad \begin{matrix} \{A,B\} \\ t=3 \end{matrix} \quad \begin{matrix} \{A,C,D\} \\ t=4 \end{matrix} \quad \begin{matrix} \{F\} \\ t=5 \end{matrix} \quad \begin{matrix} \{B,E\} \\ t=6 \end{matrix} \quad \begin{matrix} \{C,D\} \\ t=7 \end{matrix} \rangle$

show all the occurrences (there can be more than one or none, in general) of each of the following subsequences in the input sequence above. Repeat the exercise twice: the first time considering no temporal constraints (left column): the second time considering min-gap = 1 (i.e. gap > 1, right column). Each occurrence should be represented by its corresponding list of time stamps, e.g.: $\langle 0,2,3 \rangle = \langle t=0, t=2, t=3 \rangle$.

	Occurrences	Occurrences with min-gap = 1
ex.: $\langle \{B\}\{E\} \rangle$	$\langle 1,2 \rangle \langle 1,6 \rangle \langle 3,6 \rangle$	$\langle 1,6 \rangle \langle 3,6 \rangle$
$w_1 = \langle \{A\} \{B\} \{E\} \rangle$		
$w_2 = \langle \{B\} \{D\} \rangle$		
$w_3 = \langle \{F\} \{E\} \{C,D\} \rangle$		

Sequential Pattern – Exercise 1

a) (3 points) Given the following input sequence

$\langle \begin{matrix} \{A\} \\ t=0 \end{matrix} \quad \begin{matrix} \{B,F\} \\ t=1 \end{matrix} \quad \begin{matrix} \{E\} \\ t=2 \end{matrix} \quad \begin{matrix} \{A,B\} \\ t=3 \end{matrix} \quad \begin{matrix} \{A,C,D\} \\ t=4 \end{matrix} \quad \begin{matrix} \{F\} \\ t=5 \end{matrix} \quad \begin{matrix} \{B,E\} \\ t=6 \end{matrix} \quad \begin{matrix} \{C,D\} \\ t=7 \end{matrix} \rangle$

show all the occurrences (there can be more than one or none, in general) of each of the following subsequences in the input sequence above. Repeat the exercise twice: the first time considering no temporal constraints (left column): the second time considering min-gap = 1 (i.e. gap > 1, right column). Each occurrence should be represented by its corresponding list of time stamps, e.g.: $\langle 0,2,3 \rangle = \langle t=0, t=2, t=3 \rangle$.

Answer:

	Occurrences	Occurrences with min-gap = 1
ex.: $\langle \{B\}\{E\} \rangle$	$\langle 1,2 \rangle \langle 1,6 \rangle \langle 3,6 \rangle$	$\langle 1,6 \rangle \langle 3,6 \rangle$
$w_1 = \langle \{A\} \{B\} \{E\} \rangle$		
$w_2 = \langle \{B\}\{D\} \rangle$		
$w_3 = \langle \{F\}\{E\}\{C,D\} \rangle$		

Sequential Pattern – Exercise 1 – Solution

a) (3 points) Given the following input sequence

$\begin{matrix} < \{A\} & \{B,F\} & \{E\} & \{A,B\} & \{A,C,D\} & \{F\} & \{B,E\} & \{C,D\} > \\ & t=0 & t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 \end{matrix}$

show all the occurrences (there can be more than one or none, in general) of each of the following subsequences in the input sequence above. Repeat the exercise twice: the first time considering no temporal constraints (left column): the second time considering min-gap = 1 (i.e. gap > 1, right column). Each occurrence should be represented by its corresponding list of time stamps, e.g.: $\langle 0,2,3 \rangle = \langle t=0, t=2, t=3 \rangle$.

Answer:

	Occurrences	Occurrences with min-gap = 1
ex.: $\langle \{B\}\{E\} \rangle$	$\langle 1,2 \rangle \langle 1,6 \rangle \langle 3,6 \rangle$	$\langle 1,6 \rangle \langle 3,6 \rangle$
$w_1 = \langle \{A\} \{B\} \{E\} \rangle$	$\langle 0,1,2 \rangle \langle 0,1,6 \rangle \langle 0,3,6 \rangle$	$\langle 0,3,6 \rangle$
$w_2 = \langle \{B\}\{D\} \rangle$	$\langle 1,4 \rangle \langle 1,7 \rangle \langle 3,4 \rangle \langle 3,7 \rangle \langle 6,7 \rangle$	$\langle 1,4 \rangle \langle 1,7 \rangle \langle 3,7 \rangle$
$w_3 = \langle \{F\}\{E\}\{C,D\} \rangle$	$\langle 1,2,4 \rangle \langle 1,2,7 \rangle \langle 1,6,7 \rangle \langle 5,6,7 \rangle$	none

Sequential Pattern – Exercise 2

a) (3 points) Given the following input sequence

$\begin{matrix} < \{B,F\} & \{A\} & \{A,B\} & \{C,D,F\} & \{E\} & \{B,E\} & \{C,D\} > \\ t=0 & t=1 & t=2 & t=3 & t=4 & t=5 & t=6 \end{matrix}$

show all the occurrences (there can be more than one or none, in general) of each of the following subsequences in the input sequence above. Repeat the exercise twice: the first time considering no temporal constraints (left column): the second time considering $\text{max-gap} = 4$ (i.e. $\text{gap} \leq 4$, right column). Each occurrence should be represented by its corresponding list of time stamps, e.g.: $\langle 0,2,3 \rangle = \langle t=0, t=2, t=3 \rangle$.

	<i>Occurrences</i>	<i>Occurrences with max-gap = 4</i>
$w_1 = \langle \{B\} \{E\} \rangle$		
$w_2 = \langle \{B\} \{D\} \rangle$		
$w_3 = \langle \{F\} \{B\} \{C,D\} \rangle$		

Sequential Pattern – Exercise 2 – Solution

a) (3 points) Given the following input sequence

$\begin{matrix} < \{B,F\} & \{A\} & \{A,B\} & \{C,D,F\} & \{E\} & \{B,E\} & \{C,D\} > \\ t=0 & t=1 & t=2 & t=3 & t=4 & t=5 & t=6 \end{matrix}$

show all the occurrences (there can be more than one or none, in general) of each of the following subsequences in the input sequence above. Repeat the exercise twice: the first time considering no temporal constraints (left column): the second time considering $\text{max-gap} = 4$ (i.e. $\text{gap} \leq 4$, right column). Each occurrence should be represented by its corresponding list of time stamps, e.g.: $\langle 0,2,3 \rangle = \langle t=0, t=2, t=3 \rangle$.

Answer:

	<i>Occurrences</i>	<i>Occurrences with max-gap = 4</i>
$w_1 = \langle \{B\} \{E\} \rangle$	$\langle 0,4 \rangle \langle 0,5 \rangle \langle 2,4 \rangle \langle 2,5 \rangle$	$\langle 0,4 \rangle \langle 2,4 \rangle \langle 2,5 \rangle$
$w_2 = \langle \{B\} \{D\} \rangle$	$\langle 0,3 \rangle \langle 0,6 \rangle$ $\langle 2,3 \rangle \langle 2,6 \rangle$ $\langle 5,6 \rangle$	$\langle 0,3 \rangle$ $\langle 2,3 \rangle \langle 2,6 \rangle$ $\langle 5,6 \rangle$
$w_3 = \langle \{F\} \{B\} \{C,D\} \rangle$	$\langle 0,2,3 \rangle \langle 0,2,6 \rangle \langle 0,5,6 \rangle$ $\langle 3,5,6 \rangle$	$\langle 0,2,3 \rangle \langle 0,2,6 \rangle$ $\langle 3,5,6 \rangle$

GSP – Exercise 1

b) **(3 points)** Simulate the execution of the GSP algorithm on the following dataset of sequences, assuming a minimum support threshold of 60%.

$\{A\} \rightarrow \{B\} \rightarrow \{C\} \rightarrow \{D\}$
 $\{A\} \rightarrow \{B\} \rightarrow \{C\} \rightarrow \{C\}$
 $\{D\} \rightarrow \{C\} \rightarrow \{B\} \rightarrow \{C\}$
 $\{A\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{C\} \rightarrow \{E\}$

GSP – Exercise 2

b) (3 points) Running the GSP algorithm on a dataset of sequences, at the end of the second iteration it found the frequent 3-sequences on the left, and at the next iteration it generated (among the others) the candidate 4-sequences on the right. Which of the candidates will be **pruned**, and why?

Frequent 3-sequences

$\{A B\} \rightarrow \{C\}$	$\{A\} \rightarrow \{D\} \rightarrow \{C\}$
$\{A B\} \rightarrow \{D\}$	$\{B\} \rightarrow \{C\} \rightarrow \{C\}$
$\{A\} \rightarrow \{C D\}$	$\{B\} \rightarrow \{C\} \rightarrow \{D\}$
$\{B\} \rightarrow \{C D\}$	$\{B\} \rightarrow \{D\} \rightarrow \{C\}$
$\{A\} \rightarrow \{C\} \rightarrow \{C\}$	$\{D\} \rightarrow \{C\} \rightarrow \{C\}$
$\{A\} \rightarrow \{C\} \rightarrow \{D\}$	$\{D\} \rightarrow \{C\} \rightarrow \{D\}$

Candidates

1. $\{A B\} \rightarrow \{C D\}$
2. $\{A\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{D\}$
3. $\{B\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{D\}$
4. $\{A B\} \rightarrow \{D\} \rightarrow \{C\}$
5. $\{A B\} \rightarrow \{C\} \rightarrow \{D\}$

GSP – Exercise 2 – Solution

b) (3 points) Running the GSP algorithm on a dataset of sequences, at the end of the second iteration it found the frequent 3-sequences on the left, and at the next iteration it generated (among the others) the candidate 4-sequences on the right. Which of the candidates will be **pruned**, and why?

Frequent 3-sequences

$\{A B\} \rightarrow \{C\}$	$\{A\} \rightarrow \{D\} \rightarrow \{C\}$
$\{A B\} \rightarrow \{D\}$	$\{B\} \rightarrow \{C\} \rightarrow \{C\}$
$\{A\} \rightarrow \{C D\}$	$\{B\} \rightarrow \{C\} \rightarrow \{D\}$
$\{B\} \rightarrow \{C D\}$	$\{B\} \rightarrow \{D\} \rightarrow \{C\}$
$\{A\} \rightarrow \{C\} \rightarrow \{C\}$	$\{D\} \rightarrow \{C\} \rightarrow \{C\}$
$\{A\} \rightarrow \{C\} \rightarrow \{D\}$	$\{D\} \rightarrow \{C\} \rightarrow \{D\}$

Candidates

1. $\{A B\} \rightarrow \{C D\}$
2. $\{A\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{D\}$
3. $\{B\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{D\}$
4. $\{A B\} \rightarrow \{D\} \rightarrow \{C\}$
5. $\{A B\} \rightarrow \{C\} \rightarrow \{D\}$

Answer:

Candidates

1. $\{A B\} \rightarrow \{C D\}$
2. $\{A\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{D\}$ ← **PRUNED**
3. $\{B\} \rightarrow \{D\} \rightarrow \{C\} \rightarrow \{D\}$ ← **PRUNED**
4. $\{A B\} \rightarrow \{D\} \rightarrow \{C\}$
5. $\{A B\} \rightarrow \{C\} \rightarrow \{D\}$

Missing from frequent 3-sequences

- $A \rightarrow D \rightarrow D$
- $B \rightarrow D \rightarrow D$