

Data Mining II

June 20th, 2016

Exercise 1 - Classification – alternative methods (11 points)

Exercise 2 - Sequential patterns (11 points)

Given the following input sequence

<	{A}	{B}	{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	

A) 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 $\text{min-gap} = 1$ (i.e. $\text{gap} > 1$, right column). Each occurrence should be represented by its corresponding list of time stamps, e.g.: $\langle 0,2,3 \rangle = \langle t=0, t=2, t=3 \rangle$.

B) list all the subsequences that contain at least 2 elements and satisfy $\text{min-gap}=6$ (i.e. all gaps must be >5).

{A} {B}
{A} {E}
{A} {B,E}
{A} {C}
{A} {D}
{A} {C,D}
{B} {C}
{B} {D}
{B} {C,D}

	Occurrences	Occurrences with min-gap=1
<i>ex.:</i> <{D}{E}>	<1,2> <1,6> <4,6>	<1,6> <4,6>
$w_1 = \langle \{A\} \{B\} \{D\} \rangle$	<0,1,4> <0,1,7> <0,3,4> <0,3,7> <0,6,7> <3,6,7> <4,6,7>	<0,3,7>
$w_2 = \langle \{A\} \{E\} \rangle$	<0,2> <0,6> <3,6>	<0,2> <0,6> <3,6>
$w_3 = \langle \{E\} \{C,D\} \rangle$	<2,4> <2,7> <6,7>	<2,4> <2,7>

Exercise 3 - Time series / Distances (10 points)

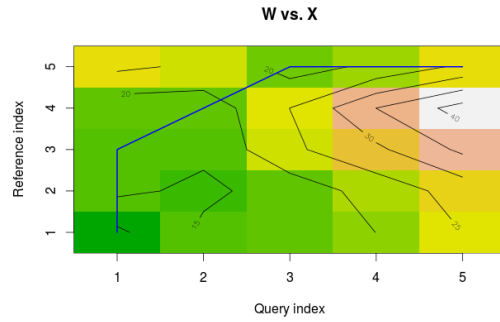
Given the following dataset of time series (on the left):

W	< 1, 3, 11, 13,15 >
X	< 10, 8, 1, 2, 10 >
Y	< 9, 17, 8, 1, 13,3,1 >

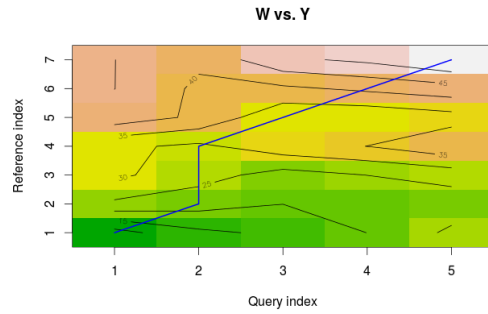
	W	X	Y
W			
X			
Y			

- 1) Compute the matrix of distances among all pairs of time series (on the right) adopting a Dynamic Time Warping distance, and computing the distances between single points as $d(x,y) = |x - y|$. For each pair of time series compared also show the matrix used to compute the final result.

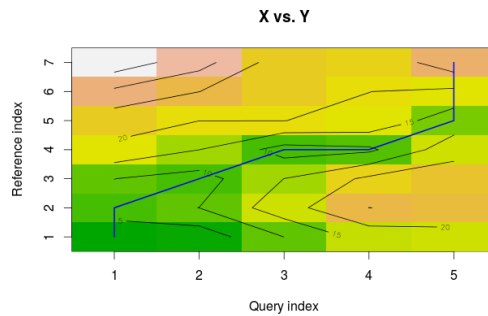
	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	9	16	16	17	26
[2,]	16	14	16	17	24
[3,]	17	17	24	25	18
[4,]	20	22	29	35	21
[5,]	25	27	36	42	26



	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	8	24	31	31	43	45	45
[2,]	14	22	27	29	39	39	41
[3,]	16	20	23	33	31	39	49
[4,]	20	20	25	35	31	41	51
[5,]	26	22	27	39	33	43	55



	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	1	8	10	19	22	29	38
[2,]	2	10	8	15	20	25	32
[3,]	10	18	15	8	20	22	22
[4,]	17	25	21	9	19	20	21
[5,]	18	24	23	18	12	19	28



- 2) Which distances will change if we constrain the DTW with a “Sakoe-Chiba Band ” of size $r=2$, i.e. the maximum misalignment allowed in the matching is of 2 positions?

NONE