WARSAW UNIVERSITY OF TECHNOLOGY
DEVELOPMENT PROGRAMME

Lazy Classification with Contrast Patterns

Marzena Kryszkiewicz

HUMAN CAPITAL
EUROPEAN UNION EUROPEAN SOCIAL FUND
Project is co-financed by European Union within European Social Fund

Contrast Patterns (CPs)

Example training dataset D with two decision classes D_P and D_N.

Outlook	Temperature	Humidity	Windy	Activity
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Overcast	Cool	Normal	True	P
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Rain	Cool	Normal	True	N
Sunny	Mild	High	False	N
Rain	Mild	High	True	N

- A *contrast pattern* is defined as a pattern that occurs in exactly one decision class.
- {Overcast, Hot} is a contrast pattern, which occurs only in the decision class D_P.

2

Lazy Classification with CPs - Input

- D - a decision table with objects assigned to either a positive decision class D_P or a negative decision class D_N.
- T - an object to be classified.

3

Example Input

Example training dataset D with two decision classes D_P and D_N.

Outlook	Temperature	Humidity	Windy	Activity
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Overcast	Cool	Normal	True	P
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Rain	Cool	Normal	True	N
Sunny	Mild	High	False	N
Rain	Mild	High	True	N

- Object to be classified:

T = {Sunny, Mild, High, True}

4

Lazy Classification with CPs - Steps

I. Create a T-related reduced version D' of the decision table D.
II. Find minimal contrast patterns Min(CP(D'$_P$)) and Min(CP(D'$_N$)) in D'.
III. Calculate compact scores of the minimal contrast patterns.
IV. Choose the decision class with the greatest compact score as relevant for T.

5

Step I

- Create a decision table D' that will contain all original decision values and only those conditional values that occur in T.

6

Example: Step I

Original D & T = {Sunny, Mild, High, True} → Reduced D'

Outlook	Temp.	Hum.	Windy	Activity
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Overcast	Cool	Normal	True	P
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Rain	Cool	Normal	True	N
Sunny	Mild	High	False	N
Rain	Mild	High	True	N

⇒

Outlook	Temp.	Hum.	Windy	Activity
		High		P
	Mild	High		P
				P
			True	P
Sunny				P
	Mild			P
Sunny	Mild		True	P
	Mild	High	True	P
				P
Sunny		High		N
Sunny		High	True	N
			True	N
Sunny	Mild	High		N
	Mild	High	True	N

Step II

- Find minimal contrast patterns Min(CP(D'$_P$)) and Min(CP(D'$_N$)) in D'.
- Note:
  - CPs can be calculated in an Apriori-like way.
  - For a candidate pattern, two supports can be determined simultaneously: in D'$_P$ & in D'$_N$.
  - The candidate is a CP if exactly one of the two supports equals 0.

Example: Step II…

Reduced D'

Outlook	Temp.	Hum.	Windy	Activity
		High		P
	Mild	High		P
				P
			True	P
Sunny				P
	Mild			P
Sunny	Mild		True	P
	Mild	High	True	P
				P
Sunny		High		N
Sunny		High	True	N
			True	N
Sunny	Mild	High		N
	Mild	High	True	N

- Candidates of length 1:

{Sunny}$_{2/3}$ {Mild}$_{4/2}$ {High}$_{3/4}$ {True}$_{3/3}$

Result: No Minimal CPs of length 1.

- Candidates of length 2:

{Sunny, Mild}$_{1/1}$ {Sunny, High}$_{0/3}$
{Sunny, True}$_{1/1}$ {Mild, High}$_{2/2}$
{Mild, True}$_{2/1}$ {High, True}$_{1/2}$

Example: Step II…

Reduced D'

Outlook	Temp.	Hum.	Windy	Activity
		High		P
	Mild	High		P
				P
			True	P
Sunny				P
	Mild			P
Sunny	Mild		True	P
	Mild	High	True	P
				P
Sunny		High		N
Sunny		High	True	N
			True	N
Sunny	Mild	High		N
	Mild	High	True	N

- Candidates of length 2:

{Sunny, Mild}$_{1/1}$ {Sunny, High}$_{0/3}$
{Sunny, True}$_{1/1}$ {Mild, High}$_{2/2}$
{Mild, True}$_{2/1}$ {High, True}$_{1/2}$

Result: {Sunny, High} ∈ Min(CP(D'$_N$)).

- Candidates of length 3:

{Sunny, Mild, True}$_{1/0}$
{Mild, High, True}$_{1/1}$

Example: Step II

Reduced D'

Outlook	Temp.	Hum.	Windy	Activity
		High		P
	Mild	High		P
				P
			True	P
Sunny				P
	Mild			P
Sunny	Mild		True	P
	Mild	High	True	P
				P
Sunny		High		N
Sunny		High	True	N
			True	N
Sunny	Mild	High		N
	Mild	High	True	N

- Candidates of length 3:

{Sunny, Mild, True}$_{1/0}$
{Mild, High, True}$_{1/1}$

Result: {Sunny, Mild, True} ∈ Min(CP(D'$_P$)).

- No candidates of length 4

Step III

- Calculate compact scores of the minimal contrast patterns in D'$_P$ and D'$_N$, where:

$$compactScore(D_i) = \frac{sup(\bigvee \mathrm{Min}(\mathrm{CP}(D_i)))}{|D_i|}.$$

Example: Step III

Reduced D’

Outlook	Temp.	Hum.	Windy	Activity
		High		P
	Mild	High		P
				P
			True	P
Sunny				P
	Mild			P
Sunny	Mild		True	P
	Mild	High	True	P
				P
Sunny		High		N
Sunny		High	True	N
			True	N
Sunny	Mild	High		N
	Mild	High	True	N

- $|D'_P| = 9$ objects
- $Min(CP(D'_P)) = \{\{Sunny, Mild, True\}\}$.
- *compactScore*(D'_P) =

$$\frac{sup(\vee Min(CP(D'_P)))}{|D'_P|} =$$

1/9.

- $|D_N| = 5$ objects
- $Min(CP(D'_N)) = \{\{Sunny, High\}\}$.
- *compactScore*(D'_N) =

$$\frac{sup(\vee Min(CP(D'_N)))}{|D'_N|} =$$

3/5.

13

Example: Step IV

Decision table D

Outlook	Temp.	Hum.	Windy	Activity
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Overcast	Cool	Normal	True	P
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Rain	Cool	Normal	True	N
Sunny	Mild	High	False	N
Rain	Mild	High	True	N

- $Min(CP(D'_P)) = \{\{Sunny, Mild, True\}\}$.
- *compactScore*(D'_P) = 1/9
- $Min(CP(D'_N)) = \{\{Sunny, High\}\}$.
- *compactScore*(D'_N) = 3/5
- Object T = {Sunny, Mild, High, True} will be classified to the decision class with greater compact score; that is, to D_N.

14

Dealing with Continuous Attributes

- The values of each continuous attribute in:
  - decision table D and
  - in the object being subject to classification

  should be normalized to the interval [0, 1].
- T-reduced decision table D’, should not contain values of a continuous attribute *c* that do not belong to [v_c - α%, v_c + α%], where $v_c \in$ T.
- The remaining values of continuous attribute *c* should be replaced with v_c in D’.

15

Example: Normalization

Normalization of D

Age	…	Activity
18	…	P
45	…	P
38	…	P
60	…	P
19	…	P
50	…	P
39	…	P
22	…	P
44	…	P
48	…	N
25	…	N
32	…	N
56	…	N
40	…	N

⟹

Age	…	Activity
0.18	…	P
0.45	…	P
0.38	…	P
0.60	…	P
0.19	…	P
0.50	…	P
0.39	…	P
0.22	…	P
0.44	…	P
0.48	…	N
0.25	…	N
0.32	…	N
0.56	…	N
0.40	…	N

Normalization of T

T = {28, …} ⟹ T = {0.28, …}

16

Example: Determining D’

Normalized D → D’

Age	…	Activity
0.18	…	P
0.45	…	P
0.38	…	P
0.60	…	P
0.19	…	P
0.50	…	P
0.39	…	P
0.22	…	P
0.44	…	P
0.48	…	N
0.25	…	N
0.32	…	N
0.56	…	N
0.40	…	N

⟹

Age	…	Activity
	…	P
	…	P
	…	P
	…	P
	…	P
	…	P
	…	P
	…	P
	…	P
	…	N
0.28	…	N
0.28	…	N
	…	N
	…	N

- Let
  - Normalized T = {0.28, …}
  - α% = 5%
- Hence, values matching 0.28 belong to:

  [0.23, 0.33]

17

References

- Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao: Instance-Based Classification by Emerging Patterns. PKDD 2000: 191-200
- Hongjian Fan, Kotagiri Ramamohanarao: Fast Discovery and the Generalization of Strong Jumping Emerging Patterns for Building Compact and Accurate Classifiers. IEEE Trans. Knowl. Data Eng. 18(6): 721-737 (2006)
- Pawel Terlecki, Krzysztof Walczak: Efficient Discovery of Top-K Minimal Jumping Emerging Patterns. RSCTC 2008: 438-447

18