Formal concept analysis for detecting criminal patterns

Roberto G. Aragón M. Eugenia Cornejo Jesús Medina
Eloísa Ramírez-Poussa Clemente Rubio-Manzano

Dpt. of Mathematics. University of Cádiz, Spain
Emails:
{roberto.aragon,mariaeugenia.cornejo,jesus.medina,
eloisa.ramirez,clemente.rubio}@uca.es

Cádiz, Spain
October, 7th

Formal Concept Analysis

- Formal concept analysis (FCA), introduced by Ganter and Wille, is a mathematical tool based on lattice theory to obtain significant information from relational datasets.
- FCA has been widely studied and applied to different frameworks and different fuzzy extensions have been introduced in the literature.
- These databases contain a set of attributes A and a set of objects B related between them by means of a relation $R \subseteq A \times B$, from which we obtain concepts.
- Establishing a hierarchy among these concepts, we obtain the algebraic structure of a concept lattice from which we can develop the data analysis.

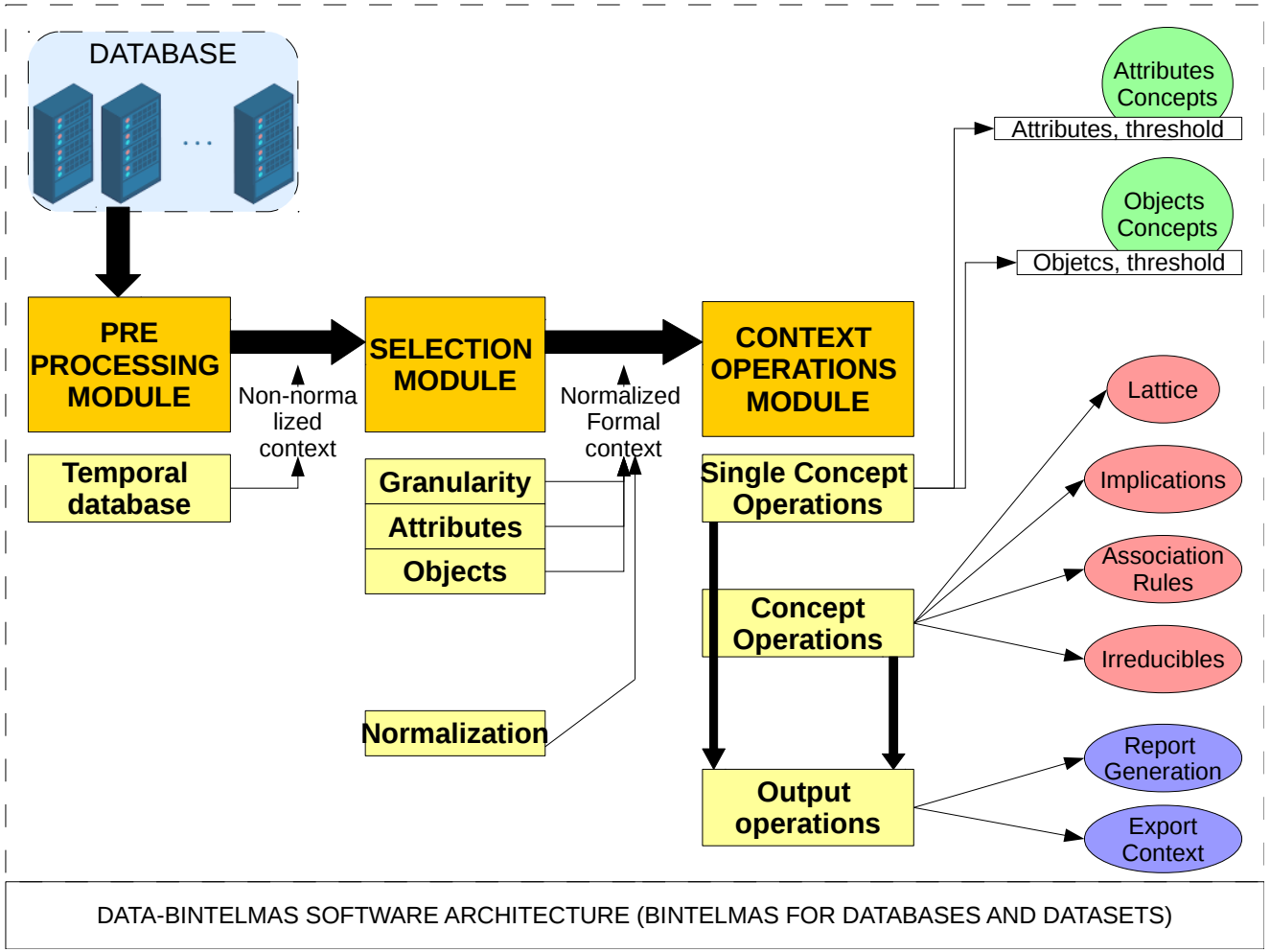
Contribution

- A fuzzy framework based on a residuated lattice will be considered in order to obtain information from a dataset provided by the Ertzaintza.
- Ertzaintza is an institution which is the autonomous police force for the Basque Country, founded in 1982.
- This institution has a great vocation of public service and of the application of new technologies to improve citizen security.
- We will apply FCA to present an initial study on the criminal incidence in the city of Bilbao, Spain.
- During the days that local football club Athletic Club plays a football match at San Mamés Stadium, differentiating between the neighborhoods near and far from the stadium.

BIntelMAS

- Operators and mechanisms used in this work have been implemented in the software BIntelMAS.
- BIntelMAS is a software that integrates different functionalities into a fuzzy framework and it has been developed by the M·CIS research team.
- It employs its own inference engine and it is composed of three main modules: preprocessing, selection and context operations.

BIntelMAS architecture



Preliminaries

In this work, we will make use of an algebraic structure called residuated lattice.

Definition

A residuated lattice is a tuple $(L, \leq, *, \rightarrow, 0, 1)$, such as

- $(L, \leq)$ is a complete lattice with 0 and 1 as bottom and top elements.
- $(L, *, 1)$ is a commutative monoide.
- The pair $(*, \rightarrow)$ satisfies the adjoint property, that is, $x * y \leq z$ if and only if $x \leq y \rightarrow z$, for all $x, y, z \in L$.

From this structure the notions of context, derivation operators, concept and concept lattice given in FCA were extended to a fuzzy framework.

Concept-forming operators

On residuated lattices, the following fuzzy extension of FCA arises.

Definition

Given a residuated lattice $(L, \leq, *, \to, 0, 1)$, a *residuated formal context* is a tuple (A, B, R) such that A and B are non-empty sets, and R is a relation $R\colon A \times B \to L$. On a context, the *concept-forming operators* $^{\uparrow}\colon L^B \longrightarrow L^A$ and $^{\downarrow}\colon L^A \longrightarrow L^B$ are defined as:

$$\begin{aligned} g^{\uparrow}(a) &= \inf\{g(b) \to R(a,b) \mid b \in B\} \\ f^{\downarrow}(b) &= \inf\{f(a) \to R(a,b) \mid a \in A\} \end{aligned}$$

for all $g \in L^B$, $f \in L^A$ and $a \in A$, $b \in B$.

Concept lattice

Definition

The set of *residuated formal concepts*, that is, the set

$$\mathcal{C}(A, B, R) = \{\langle g, f\rangle \mid g \in L_2^B, f \in L_1^A \text{ and } g^{\uparrow} = f, f^{\downarrow} = g\}$$

is called *residuated concept lattice*, which forms a complete lattice with the ordering defined by
$\langle g_1, f_1\rangle \preceq \langle g_2, f_2\rangle$ if and only if $g_1 \preceq_2 g_2$ (equivalently $f_2 \preceq_1 f_1$).

Analysis of Bilbao crime dataset

The period considered from Bilbao crime dataset is from 17-08-2018 to 07-04-2019 and we have compared the days of football match (FD, 18 days) with the days of no football match (NFD, 216 days).

The following table shows a general analysis that exhibits the proportion of crimes in the different areas of Bilbao city. This will give us a global vision of the behavior of crime in this city as well as the influence of football matches in it.

Percentages of crimes

Incidence(number/days)	FD (18 days)	NFD (216 days)	crime number
Bilbao	71.72	68.93	1291|14890
Nearby neighborhoods rate(number/total)	38.39 53.52%	34.07 49.43%	691|7360
Far neighborhoods. rate(number/total)	33.33 46.48%	34.86 50.57%	600|7530

According to the information provided by the Ertzaintza, we will designate as Zone 1 the area that encompasses the neighborhoods near the San Mamés Stadium and as Zone 2 the area that includes all the neighborhoods that are considered far from the stadium.

Crimes every day

Crimes which are committed every day, together with their minimum values

The following tables present the information about the intents of the top elements of the concept lattices associated with the concept lattice corresponding to FD and NFD in Zone 1 in the left side and Zone 2 in the right side.

Four corners

Intent of the top concept associated with FD in Zone 1

Petty larceny	0.05

Support=18
Percentage=1
Objects=18

Intent of the top concept associated with FD in Zone 2

Fraud	0.07
Petty larceny	0.15
Theft with force	0.17

Support=18
Percentage=1
Objects=18

Intent of the top concept associated with NFD in Zone 1

Petty larceny	0.05

Support=216
Percentage=1
Objects=216

Intent of the top concept associated with NFD in Zone 2

Fraud	0.07
Petty larceny	0.12

Support=216
Percentage=1
Objects=216

Consequences

- We observe that "petty larceny" is the unique crime appearing in Zone 1 whereas a wider range of crimes appear in Zone 2.
- This means that the criminality is higher in Zone 2 and every day have a constant minimum of criminal activity.
- Notice that the obtained truth-values associated with the crimes are minimum thresholds verified by all considered days. Therefore, all football match days satisfy these crimes.
- The differences between FD and NFD in Zone 2 may be due to the different number of days to be considered.

Focus on UPF

Undermine the police function, increasing risk in public safety (UPF)

The following tables show the intents of attribute-concepts related to "undermine the police function, increasing risk in public safety (UPF)" with a high incidence (0.75), that is, $\langle \phi^{\downarrow}_{\text{UPF},0.75}, \phi^{\downarrow\uparrow}_{\text{UPF},0.75} \rangle$.

Four corners of UPF

FD Zone 1

Fraud	0.2
Petty larceny	0.7
Less severe injuries	0.22
UPF	0.75
Theft with violence	0.67

Support=2.25; Objects=6
Percentage=0.125

FD Zone 2

Threats	0.2
Damage to other people's property	0.71
Fraud	0.07
Petty larceny	0.15
UPF	0.75
Theft with force	0.17

Support=2.25; Objects=6
Percentage=0.125

NFD Zone 1

Petty larceny	0.05
UPF	0.75

Support=17.25; Objects=47
Percentage=0.08

NFD Zone 2

Fraud	0.07
Petty larceny	0.15
UPF	0.75

Support=18; Objects=54
Percentage=0.08

Analysis

- From the tables we can see that the number of crimes associated with $\phi_{\mathrm{UPF},0.75}$ on football days is greater than the number of crimes when there are not football matches.
- The increase in crime associated with this fuzzy attribute is also reflected in the percentages, which are greater in the tables corresponding to FD.
- There are differences in the type of crimes that take place on FD in Zone 1 and Zone 2.
- For example, in Zone 2, "damage to other people's property" appears with a high degree, while in Zone 1 "petty larceny" and "theft with violence" have the highest degrees.
- We could also infer that a greater permissiveness exists in the behavior of the police force that is in Zone 1 on FD.
- This fact is based on the crimes, that appear with a high degree of UPF in Zone 1, are worse than those that we find with a high degree in Zone 2.

Attribute implications – Decision rules

We will show three tables that include interesting **fuzzy attribute implications on FD in Zone 1**. Considering that the "petty larceny" is a crime that takes place every day with truth-value 0.05.

Intent of the top concept associated with FD in Zone 1

Petty larceny	0.05

Support=18

Percentage=1

Objects=18

Attribute implications: Theft with intimidation

We obtain the following attribute implications:

- The first attribute implication informs us that whenever there is some "theft with intimidation", there will also be "theft with violence".

Antecedent	{Theft with intimidation/0.01 }
Consequent	{Petty larceny/0.07, Theft with intimidation/0.25, Theft with violence/0.17 }
Support	0.444
Confidence	1.0

Attribute implication: less severe injuries

- The second implication shows that the appearance of "less severe injuries" ensures the onset of "theft with violence".

Antecedent	{Less severe injuries/0.01 }
Consequent	{Petty larceny/0.09, Less severe injuries/0.11, Theft with violence/0.17 }
Support	0.611
Confidence	1.0

Attribute implication: UPF

- The third implication indicates that whenever there is an offense of "undermine the police function, increasing risk in public safety", then offenses such as "fraud", "less severe injuries" and "theft with violence" are committed.

Antecedent	{Undermine the police function, increasing risk in public safety/0.01 }
Consequent	{Fraud/0.2, Petty larceny/0.3, Less severe injuries/0.22, Undermine the police function, increasing risk in public safety/0.25, Theft with violence/0.33 }
Support	0.333
Confidence	1.0

Attribute implications FD and Zone 2

Take into account that in **Zone 2 (FD)** every day satisfies at least Fraud/0.07, Petty larceny/0.15 and Theft with force/0.17.

Intent of the top concept associated with FD in Zone 2

Fraud	0.07
Petty larceny	0.15
Theft with force	0.17

Support=18
Percentage=1
Objects=18

Attribute implications: Theft with intimidation Z2

- Concerning "theft with intimidation", we can think that it implies more crimes in Zone 2, however, the support is half of Zone 1.

Antecedent	{Theft with intimidation/0.01 }
Consequent	{Fraud/0.27, Petty larceny/0.23, Theft with force/0.25, Theft with intimidation/0.25, Less severe injuries/0.13, Damage to other people's property/0.14}
Support	0.222
Confidence	1.0

Attribute implication: less severe injuries Z2

- We have that some "less severe injuries" does not imply any other crime in Zone 2.

Antecedent	{Less severe injuries/0.01}
Consequent	{Less severe injuries/0.13 }
Support	0.778
Confidence	1.0

Attribute implication: UPF Zone 2

- Similar meaning to the attribute-concepts studied previously, for which a less truth-degree of UPF is required. This reinforces the assertions given to this crime.

Antecedent	{Undermine the police function, increasing risk in public safety/0.01}
Consequent	{Threat/0.2, Damage to other people's property/0.29, Undermine the police function, increasing risk in public safety/0.25}
Support	0.333
Confidence	1.0

Conclusions and future work

- The study given in this work has shown that FCA extracts relevant information from the dataset provided by the regional Spanish security force Ertzaintza.
- Different concepts and implications on two zones in Bilbao and in different kinds of days: match and non-match days have been analyzed.
- The preliminary interpretations have advertised that FD has a greater criminality, the police force is more permissive in Zone 1 in these days.
- In addition, "theft with intimidation" implies other crimes such as "theft with violence" (FD), "theft with force" and "damage to other people's property" (NFD).
- In the future, more concepts and implications will be inspected.

Formal concept analysis for detecting criminal patterns

Roberto G. Aragón M. Eugenia Cornejo Jesús Medina
Eloísa Ramírez-Poussa Clemente Rubio-Manzano

Dpt. of Mathematics. University of Cádiz, Spain
Emails:
{roberto.aragon,mariaeugenia.cornejo,jesus.medina,
eloisa.ramirez,clemente.rubio}@uca.es

Cádiz, Spain
October, 7th