Extracting Emergent Knowledge about the Socioeconomic Urban Contexts

Juan Galán-Páez

Dept. Computer Science and AI -
Universidad de Sevilla
E.T.S. Ingeniería Informática, Avda. Reina
Mercedes s.n. 41012-Seville, Spain
juangalan@us.es

Joaquín Borrego-Díaz

Dept. Computer Science and AI -
Universidad de Sevilla
E.T.S. Ingeniería Informática, Avda. Reina
Mercedes s.n. 41012-Seville, Spain
jborrego@us.es

Jaime de Miguel-Rodríguez

Sushant School of Art and Architecture -
Ansal University Sector 55, Golf Course Road
122003 (Haryana) Gurgaon, India
demiguel.jaime@gmail.com

Abstract

An approach to represent and analyze socioeconomic contexts as well as to reason with them, in order to extract useful conclusions about the social perception emerging from citizens' beliefs and feelings, is introduced. We concentrate here in the formal aspects of the solution, completing this way our work [4].

Introduction and technical background

The increasing availability of digital information about cities and their dynamics represents an enormous challenge for urban planners, data scientists and, in general, for the scientific community involved in the development of smart cities. The collection, processing and use of urban data applied to projects such as designing new applications, urban services or analyzing cities' social dimension, is a key research & development field.

The human understanding of the knowledge information provides, as well as the extraction and processing of digital information, are particularly difficult to tackle due to the big amount of data. For example, the use of portals where citizens interact and, in a more challenging situation, where citizenship socializes their feelings [3]. The problem has a semantic nature, as it is essentially qualitative and context-sensitive.

The **aim of the paper** is to show how to transform information from the WWW about specific topics of the city, generated by institutions, companies and citizens, into (emergent) semantic information useful to study how citizens understand the particular socioeconomic context.

Background: Formal Concept Analysis
Formal Concept Analysis (FCA) [5] provides powerful semantic tools for classification, data mining and KE and Discovery (KD). Among these tools, particularly interesting are concepts extraction and organization, and implication basis. FCA also provides reasoning tools, which can be used to re-organize, enrich and even to predict the evolution of the system [2]. A *formal context* $M = (O, A, I)$ consists of two sets, O (objects) and A (attributes), and a relation $I \subseteq O \times A$. Finite contexts can be represented by a 1-0-table (identifying I with a boolean function on $O \times A$). Given $X \subseteq O, Y \subseteq A$, $X' = \{a \in A \mid oIa$ for all $o \in X\}$ and $Y' = \{o \in O \mid oIa$ for all $a \in Y\}$ can be defined.

Knowledge Bases in FCA consist of *implications between attributes*. An implication is a pair of sets of attributes, written as $Y_1 \to Y_2$. An implication is true with respect to $M = (O, A, I)$ according to the following definition: A subset $T \subseteq A$ *respects* $Y_1 \to Y_2$ if $Y_1 \not\subseteq T$ or $Y_2 \subseteq T$. $Y_1 \to Y_2$ is said to hold in M ($M \models Y_1 \to Y_2$ or $Y_1 \to Y_2$ is an implication of M) if for all $o \in O$, the set $\{o\}'$ respects $Y_1 \to Y_2$.

Let $\mathcal{L}$ be a set of implications and L be an implication.

An implication L **follows from** $\mathcal{L}$ ($\mathcal{L} \models L$) if each subset of A respecting $\mathcal{L}$ also respects L. $\mathcal{L}$ is **complete** if for every implication L, if $M \models L$ then $\mathcal{L} \models L$. $\mathcal{L}$ is **non-redundant** if for each $L \in \mathcal{L}$, $\mathcal{L} \setminus \{L\} \not\models L$. Finally $\mathcal{L}$ is a (implication) **basis** for M if $\mathcal{L}$ is complete and non-redundant. A particular basis is the *Duquenne-Guigues* or so called *Stem* Basis (SB) (see [5]). The reasoning system we use is a production system on the selected implication basis. In FCA, association rules are also implications between sets of attributes. Confidence and support are defined as usual in data mining. The analogous to Stem Basis for association rules is the Luxenburger basis [6]. The reasoning system for SB can be adapted for reasoning with Luxenburger basis. Recall that Y is closed if $Y'' = Y$. Let be $M = (O, A, I)$ a formal context and $Y, Y_1, Y_2 \subset A$.

- Given Y_1, Y_2 closed, $Y_1 \prec Y_2$ if there is not Y closed such that $Y_1 \subset Y \subset Y_2$.

- The **support of** $Y \subseteq A$ is $supp(Y) = |Y'|$. and the support of $L = Y_1 \to Y_2$ is $supp(L) = |(Y_1 \cup Y_2)'|$

- The **confidence** of L is $conf(L) = \dfrac{supp(Y_1 \cup Y_2)}{supp(Y_1)}$

Given γ, δ, the **Luxenburger basis of a context** M with confidence γ and support δ, denoted by $\mathcal{L}(M, \gamma, \delta)$, is

$$\mathcal{L}(M, \gamma, \delta) := \{L : Y_1 \to Y_2 \mid Y_1, Y_2 \text{ closed}, \\ Y_1 \prec Y_2,\ conf(L) \geq \gamma,\ sup(L) \geq \delta\}$$

By convention, $\mathcal{L}(M, 1, 0)$ is the SB basis. Luxenburger basis can be interpreted as an association rules set, thus they allow reasoning under uncertainty. The particular feature of the Luxenburger basis is that it shows some explicit information on how concepts from the concept lattice are related (according to the confidence and support).

A Methodology based on FCA. The methodology applied in this work is based on the processing of data to obtain a formal context representing the global knowledge (the called *Monster Model*). This context is usually built by collecting a huge amount of information by means of WWW services. From this global context, a concept lattice can be extracted. This lattice is a complex semantic network with complex topology, providing some insights on the robustness of the data set [1]. The second stage involves the selection of a number of features and thresholds (to discretize data) to produce KBs for reasoning tasks. In this way both, concept lattices and implication basis, are obtained, providing knowledge from the information.

In this paper data resources come from housing market, which represent a highly dynamic data set that plays a key role in the understanding of urban processes. This data set provides a meaningful view of the socioeconomic landscape of our cities. For instance, when focusing on a certain neighborhood, the data can be used to measure sales-price variation indexes upon which substantial knowledge can be gained regarding issues like social integration and diversity of urban areas.

In this case Data comes from *Self-City* platform[1]. This platform allows obtaining real time urban information in a meaningful way, striving to make dynamic data both efficient and essential for city management and decision-making. In order to achieve these goals, Self City involves a complex articulation of visualization techniques, crossing of planning data, clustering algorithms and most importantly, monitoring strategies in tune with agencies' needs. The right setup of these different apparatuses will determine till what extent will, the information served, be assimilated and with this, the chances of reaction upon data.

Reasoning with emergent concepts

Currently, Self-City manages digital information only. Such information can be visualized, but a sound comprehension of its dynamics and conceptual structure requires semantic processing. The understanding of the full dataset needs a semantic interpretation of the concepts involved in this complex system. Contexts are built from the Self-City information flow as for instance temporal and spatial information on for sale housing: objects are for sale homes and attributes provide a qualitative description of these at a given moment; description of the item (i.e. price, dimensions, environment, etc.) and description of the item evolution (i.e. price changes, environment evolution, etc.). Thresholds for attributes have to be selected.

The formal context considered in this work consist of 6000 (approx.) for sale homes in the city of Seville. Also subareas of Seville can be considered (i.e. streets, zones, etc.) for a more detailed analysis. A more precise analysis of conceptual differences can be provided by means of logical reasoning with association rules related to the lattice. Also, it is possible to consider how qualitative patterns are reproduced within the city.

The study of this kind of semantic structures allows, for example, **discovering new concepts**, as a result of exploratory analysis of the Concept Lattice for a particular city area. This allows both understanding and comparing different city areas and neighborhoods. The estimation of

[1] http://self-city.com/

similarity among concept lattices from different cities allows comparing socioeconomic contexts. Fig. 1 shows the similarity between the main streets of a district and a specific Avenue.

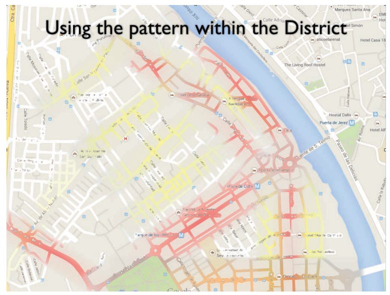

Figure 1: Semantic (social) Patterns in a district from Seville

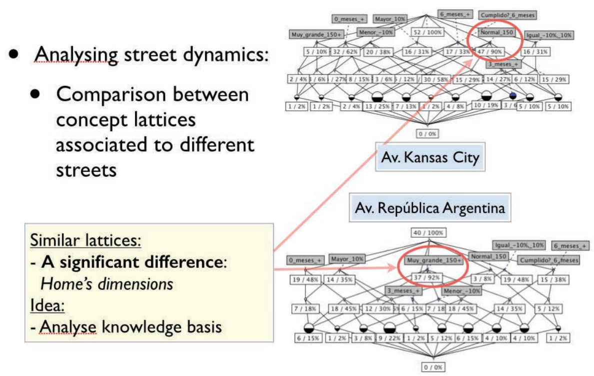

Figure 2: Comparing concept lattices from different streets

It is also possible to **compare citizens' perceptions** on the socioeconomic context for different neighborhoods (see Fig. 2) by using Luxenburger rules. In this case, if M_K is the context for the first street and M_A for the second, it is possible to see if the concept "large floor" is differently perceived by citizens in different neighbors: If "large size floor" attribute is permuted with "medium size floor" in the context M_A, obtaining M'_A, it holds that:

- $\mathcal{L}(M_K, .85, 0) \models \mathcal{L}(M'_A, 0.97, 0)$
- $\mathcal{L}(M'_A, 1, 0) \models \mathcal{L}(M_A, 0.94, 0)$

Acknowledgements

TIC-6064 Excellence project (*Junta Andalucía*) and TIN2013- 41086-P (Spanish Ministry of Economy and Competitiveness), co-financed with FEDER funds

References

[1] G.A. Aranda-Corral, J. Borrego-Díaz, J. Galán-Páez, On the Phenomenological Reconstruction of Complex Systems—The Scale-Free Conceptualization Hypothesis, Systems Research and Behavioral Science 30(6): 716-734 (2013).

[2] G. A. Aranda-Corral, J. Borrego-Díaz, J. Galán-Páez, Complex Concept Lattices for Simulating Human Prediction in Sport. J. Systems Science and Complexity 26(1):117-136 (2012)

[3] D Beraldo, J Galán–Páez. The #OCCUPY network on Twitter and the challenges to social movements theory and research. International Journal of Electronic Governance 6 (4), 319-341 (2014)

[4] J. Borrego-Díaz, J. Galán-Páez, J. De Miguel-Rodríguez, An approach to social perception of socioeconomic contexts by means of concept reasoning. VII International Meeting in Dynamics of Socio-Economic Systems (2014).

[5] B. Ganter and R. Wille, *Formal Concept Analysis - Mathematical Foundations*, Springer-Verlag, 1999.

[6] Luxenburger, M.: Implications partielles dans un contexte. Math. Inf. Sci. hum. 113 (1991) 35–55