Building Cognitive Cities with Explainable Artificial Intelligent Systems

Jose M. Alonso[1] and Corrado Mencar[2]

[1] Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
josemaria.alonso.moral@usc.es
[2] Department of Informatics, University of Bari "Aldo Moro", Bari, Italy,
corrado.mencar@uniba.it

Abstract. In the era of the *Internet of Things* and *Big Data*, data scientists are required to extract valuable knowledge from the given data. This challenging task is not straightforward. Data scientists first analyze, cure and pre-process data. Then, they apply Artificial Intelligence (AI) techniques to automatically extract knowledge from data. However, nowadays the focus is set on knowledge representation and how to enhance the human-machine interaction. Non-expert users, i.e., users without a strong background on AI, require a new generation of explainable AI systems. They are expected to naturally interact with humans, thus providing comprehensible explanations of decisions automatically made. In this paper, we sketch how certain computational intelligence techniques, namely interpretable fuzzy systems, are ready to play a key role in the development of explainable AI systems. Interpretable fuzzy systems have already successfully contributed to build explainable AI systems for cognitive cities.

Keywords: Explainable Computational Intelligence, Interpretable Fuzzy Systems, Natural Language Generation, Cognitive Cities

1 Introduction

The quest of "comprehensibility" and "explanation" in the field of Computational Intelligence is rooted in the more general paradigm of Computing With Words (CWW) stated by Zadeh more than two decades ago [1]. CWW comes into play when there is the need of representing and manipulating knowledge expressed in Natural Language (NL). CWW is closely related to the Computational Theory of Perceptions (CTP) [2] and is based on Fuzzy Set Theory (FST) [3].

CWW operates at the semantic level, i.e., inference is carried out by considering the semantic definition of words (and their connectives) in terms of fuzzy sets and operators. In fact, CWW relies on the assumption that fuzziness is intrinsic in the semantics of most linguistic terms. Moreover, as a distinctive feature, it deals naturally with fuzziness at the inference process and produces results represented linguistically.

Fuzzy (rule-based) systems use CWW to represent linguistic knowledge and carry out approximate reasoning under imprecision and uncertainty, which makes their use appealing in many application domains. Interpretable fuzzy systems (IFS) are fuzzy systems whose behavior is *human-centric*, i.e., they can be easily understood, trusted on or accounted for human beings. IFS are especially appreciated in domains with advanced human-computer collaboration such as Medicine [4]. However, it is noteworthy that building IFS is a matter of careful design [5]. In a nutshell, interpretability is not granted by the adoption of FST, which represents a necessary yet not a sufficient requirement for CWW and human-centric computing [6]. Accordingly, fuzzy designers must look for a good interpretability-accuracy trade-off when designing fuzzy systems for specific applications.

IFS combined with NL generation techniques are a recent development [7], well suited to address one of the last challenges stated by the USA Defense Advanced Research Projects Agency (DARPA) [8]:

> "Even though current AI systems offer many benefits in many applications, their effectiveness is limited by a lack of explanation ability when interacting with humans."

Notice that powerful AI techniques like Deep Learning [9] perform very well when dealing with low-level perceptions and pattern recognition tasks but the resulting models act as black boxes, i.e. their main drawback is their lack of explanation ability what makes them hard to be trusted by humans.

Black-box models are not suitable for Cognitive Cities. A Cognitive City [10] evolves from a smart city by introducing collaborative intelligence between the city and its citizens who are willing to produce and receive information (thus requiring mutual communication of information and knowledge between humans and machines). Accordingly, Cognitive Cities should be populated with explainable AI systems able to share with humans their view of the world.

In this paper, we will briefly describe the concept of Interpretability within CWW (Sec. 2), then we outline a novel framework to build explainable AI systems for Cognitive Cities, also enumerating some examples of applications where this framework has been applied (Sec. 3). Finally, Section 4 draws main conclusions and sketch future work.

2 Interpretability as Semantic Co-intension

Human perceptions are defined by intrinsic and extrinsic attributes regarding human senses (sight, smell, taste, touch and hearing). In addition, human pleasantness depends on what people experience (Perceptions), but also on what people expect (Cognitions) which is influenced by common and personal background (e.g., context or mood). In most cases, human beings use NL for describing their perceptions and for construing their experience [11].

The use of FST and CWW for designing comprehensible intelligent systems fits with the intuitive idea that many concepts represented in NL have a

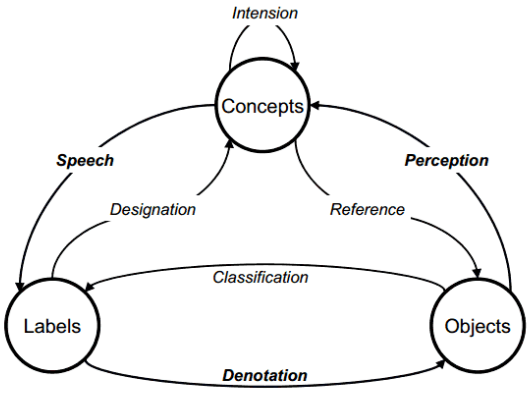

Fig. 1. Flow diagram of a perception-based system.

perception-based nature. An agent can perceive objects of the reality and associate them to concepts which can be verbalized by means of linguistic labels (see Fig. 1). On the one hand, perception is a cognitive act of transforming sensory data into mental representations. On the other hand, linguistic labels designate concepts which refer to the objects of the world. Fuzzy logic (in the wide sense) moves from the assumption that objects belong to a reality that is characterized by continuous features, which are perceived by sensors able to observe and process such continuity. In consequence, the concepts reflect the continuity of the reality and, therefore, accommodate the properties of graduality (because objects' features vary without sharp boundaries) and granularity (because each concept refers to a multitude of objects). Graduality and granularity are the key components of fuzzy sets which are the basic building blocks of several theories and methodologies, including CWW.

Thus, FST seems very suitable to formalize perception-based intelligent system which are interpretable to human beings. However, the use of FST to design intelligent systems is not enough to ensure interpretability. More than 30 years ago, Michalski introduced his "Comprehensibility Postulate" (CP) [12], which can be conveniently used as a starting point for discussing comprehensibility (or interpretability) in knowledge-based systems:

> "The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities. Components of these descriptions should be comprehensible as single *chunks* of information, directly *interpretable in natural language*, and should relate quantitative and qualitative concepts in an integrated fashion."

Notice that the CP was formulated without explicitly considering FST; nevertheless, it highlights several key-points that are worth observing. Firstly, the

human-centrality of the results of a computer induction process, which should be described symbolically as a necessary condition to communicate information. However, such symbols are not "empty", but they should represent chunks of information, namely, *information granules*, that are groups of data tied together by semantic relationships (e.g., proximity or similarity) [13, 14].

Moreover, such symbols should be directly interpretable in NL. However, this does not boil down to a simple selection of symbols within a corpus, i.e., a vocabulary made of NL terms for a specific domain. In fact, the CP requires the *interpretation* of symbols to be in NL. This is a requirement on the *semantics* of symbols and relations between them, i.e., on the information granules they denote. More precisely, on one hand information granules (resulting from computing processes) should conform with concepts a human can conceive; on the other hand, NL terms convey an implicit semantics (which depends also on the context), that is shared among all human beings speaking the same language. Therefore, a symbol coming from NL can be used to denote an information granule only if the implicit semantics of the symbol highly matches with the explicit semantics of the information granule.

This relation between both implicit and explicit semantics is called "semantic co-intension". It is inspired by the concept of model co-intension of Zadeh [15]:

> "In the context of modeling, co-intension is a measure of proximity of the input/output relations of the object of modeling and the model. A model is co-intensive if its proximity is high."

We use the Zadeh's notion of co-intension to relate the semantics defined by information granules with the underlying semantics held by the used symbols: the results of an inductive process are interpretable if the explicit semantics embedded in the model is co-intensive with the implicit semantics inferred by the user while interpreting the model. In this sense, we think FST and CWW can naturally contribute to explainable AI.

3 Explainable AI Systems for Cognitive Cities

The concept of "Cognitive City" was first introduced by Novak [16] and it emphasizes the role of learning, memory creation and experience retrieval as central processes for coping with current challenges of efficiency, sustainability and resilience [17]. In a Cognitive City, an intelligent and distributed collaboration takes place between the city and its citizens who act as "sensors" as well as "recipients". Thus, it arises a novel form of intelligence, which may be called "collaborative intelligence", where people and machines collaborate to solve complex problems [18]. Of course, effective human-machine communication requires mutual understanding. To achieve this goal, a Cognitive City must be populated with explainable AI systems able to share with humans their *semantic co-intensive* view of the world.

We have already empirically shown the benefits of combining IFS with NL generation techniques to make the interaction between humans and fuzzy systems more natural [7]. In addition, we described the role of interpretable fuzzy

systems in designing Cognitive Cities in [19]. In short, we created a new framework (see Fig. 2) where we combined (1) the so-called Linguistic Description of Complex Phenomena (LDCP) [20], which implements and enhances Zadeh's CTP, and (2) the NL generation pipeline proposed by Reiter and Dale [21].

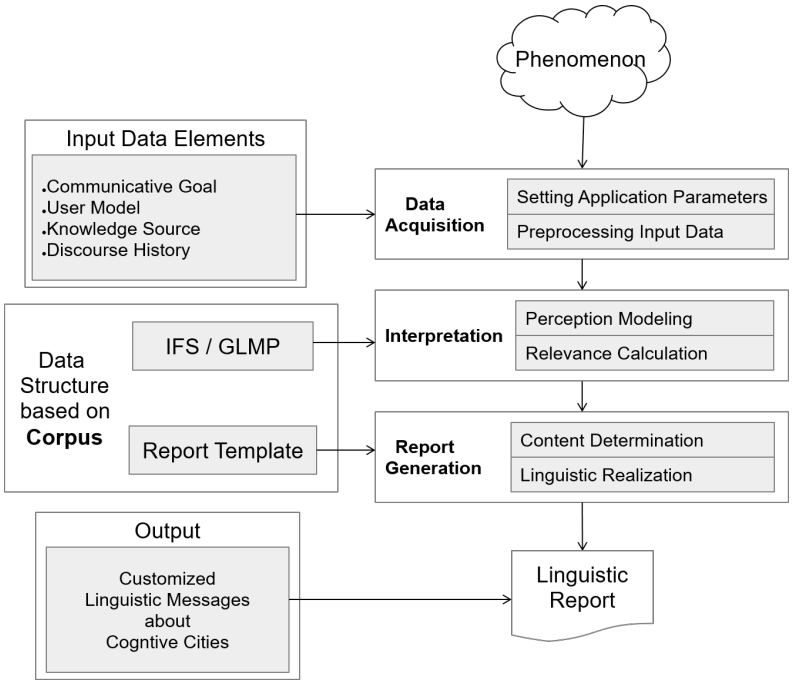

Fig. 2. A Framework for building explainable AI systems in Cognitive Cities.

The framework consists of three main stages:

1. Data acquisition is made through both humans and machines. Raw data are pre-processed in accordance with the communicative goal, the user model, the knowledge source, and the discourse history.
2. Data analysis and interpretation is carried out by IFS which are carefully designed off-line. Notice that domain knowledge is embedded into linguistic rules which relate semantically co-intensive concepts taken from a corpus specific for the target application. They are in the core of the so-called Granular Linguistic Model of Phenomena (GLMP) which consists of a hierarchical network of perception mappings and computational perceptions defined by IFS. Moreover, the relevance calculation is the basis to generate customized messages. Only the most relevant pieces of information will be added to the final report.
3. The extracted knowledge (expressed in the form of linguistic pieces of information coming out of the so-called content determination stage) is presented

to humans in NL (after document planning and careful computational linguistic realization which includes lexicalization, referring expression generation and aggregation).

We have already applied this framework (1) to provide citizens in Gijón (Spain) [19] with information related to the public bus transport; and (2) to provide citizens of European cities with details about their energy consumption [22]. Notice that the last work was carried out in the context of the NatConsumers Horizon2020 EU project[3]. In addition, the interested reader is referred to the R software package called rLDCP [23] which provides a first open-source implementation of this framework.

Let's introduce briefly how the framework sketched in Fig. 2 was implemented for one of the pilots in the NatConsumers project. Input data elements were as follows:

- *Communicative goal.* The main factors influencing residential energy consumption in European countries were identified in a previous study. Here, the goal was the automatic generation of linguistic advice for saving energy at home, remarking the main factors to consider and the main actions to carry out. Three kind of energy consumptions are considered: General, specific and standby consumptions.
- *User model.* We identified the target consumers who will receive the customized messages. Moreover, we classify consumers regarding both attitudinal and physical taxonomies which were previously defined by other partners in the project.
- *Knowledge source.* Information related to the energy consumption for each household involved in the project.
- *Discourse history.* The sequence of messages already generated for each specific consumer.

Fig. 3 shows a schematic excerpt of the IFS/GLMP described in [22]. On the one hand, the perception mappings (PM) at the lowest level of the hierarchy (e.g., $1PM_{CD}$) are implemented by strong fuzzy partitions with linguistic terms established in accordance with the corpus given for this application domain. Thus, we preserve interpretability at partition level. Then, PMs in upper levels ($2PM$) are implemented by fuzzy IF-THEN rule sets, also paying attention to interpretability constraints. For example, $2PM_{RD}$ includes rules such as "IF Average Cluster Consumption is High AND Hourly Household Consumption is High THEN The Household Consumption is Similar to Other Households in the Same Cluster".

In addition, Fig. 3 includes some illustrative examples of sentences which were generated as instantiations of the related templates, after running the fuzzy inference process, for a specific case. Moreover, Fig. 4 depicts the final report generated for the same illustrative case. Notice that the report combines both NL texts and graphs with the aim of maximizing interpretability.

[3] H2020 Coordination and Support Action (CSA) [http://www.natconsumers.eu]

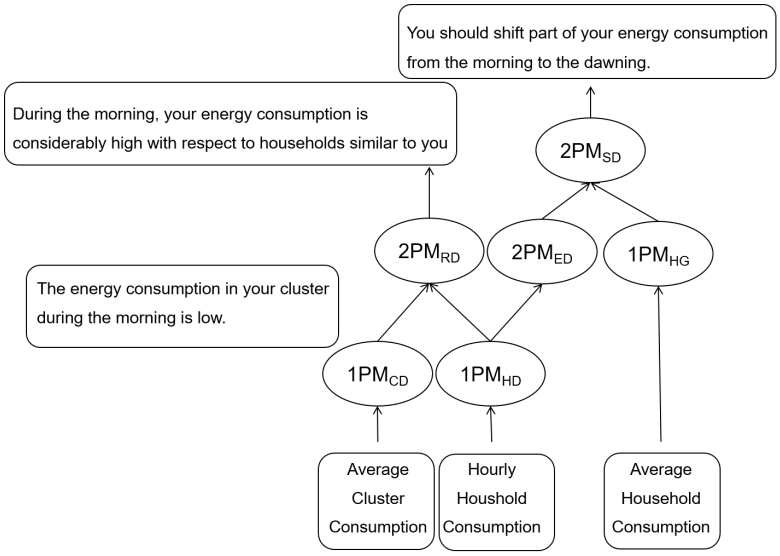

Fig. 3. Example of IFS/GLMP for NatConsumers project.

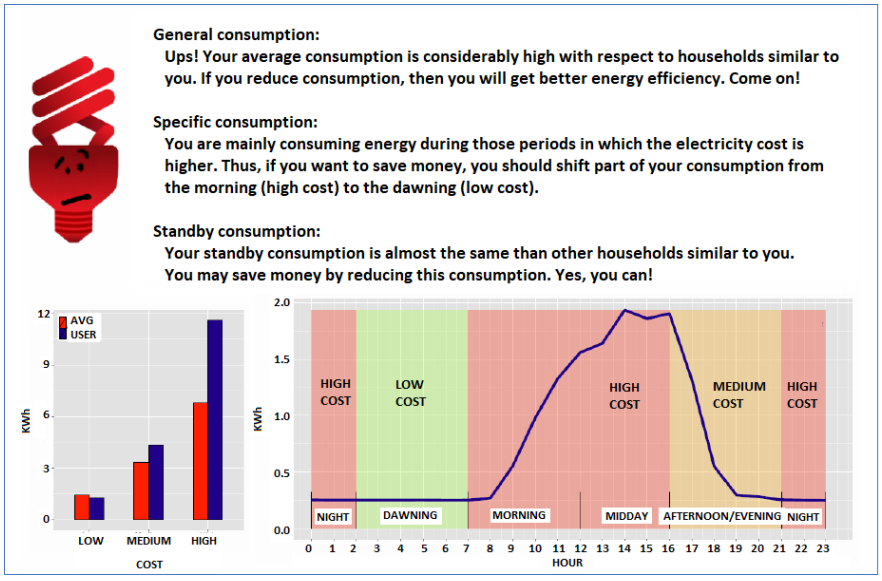

Fig. 4. Illustrative example of customized report for one of the anonymous households involved in the NatConsumers project.

4 Conclusions and Prospects

Taking profit of our previous background as designers of interpretable fuzzy systems, we have gone a step forward in the generation of explainable AI systems.

They are ready to convey citizens with valuable knowledge (represented by NL texts) which is automatically extracted from data. Nevertheless, we are aware this is only a first step and a lot of work remains to do. For example, cost-effectiveness, scalability or interface with other AI systems will turn up as key issues to address with the aim of applying our proposal universally in a cognitive city. In the short term, we plan to extend our framework with interactive dialog systems (endowed with argumentation theory approaches) which are likely to enhance the human-machine interaction capability of our explainable AI systems.

Results on applied research can be profitably used in theoretical research on interpretability. Roughly speaking, the results achieved on studying interpretability of fuzzy systems are mostly based on common-sense arguments and heuristic approaches [24]. This recently gave rise to some critical positions on the state-of-art [25], which stimulates new directions of investigation. From a theoretical viewpoint, research on comprehensibility can be cast into the field of communications between granular worlds, where agents are endowed with a granular representation of knowledge and want to communicate some information. In this abstract setting, several levels of representation of information can be established, from numerical (representing signals) to symbolic (representing terms), possibly passing through intermediate, hybrid levels. In this sense, Granular Computing could be intended as the cognitive bridge from the numerical reality to the symbolic world, where abstract relations emerge from specific dependencies. Namely, in analyzing the main theme of communication between granular worlds, the main properties of semantic communication could be investigated by using the tools offered by Granular Computing. Examples of these properties are imprecision, uncertainty, fuzziness or vagueness. All these properties are intimately connected with NL, and, traditionally, systematically removed in view of communicating information to machines and between machines. A lot of work has already been done in literature, by the definition of theories like Possibility Theory [26], Precisiated NL [27] or the theory of Z-numbers [28]. Yet, there is large room for further research, especially if aimed at giving computational solutions, i.e., solutions that can be translated into computer programs.

Acknowledgements

This work was supported by TIN2014-56633-C3-3-R (ABS4SOWproject) from the Spanish "Ministerio de Economía y Competitividad". Financial support from the Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2016-2019) and the European Union (European Regional Development Fund - ERDF), is gratefully acknowledged.

References

1. Zadeh, L.A.: From computing with numbers to computing with words - From manipulation of measurements to manipulation of perceptions. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications **46**(1) (1999) 105–119

2. Zadeh, L.A.: A new direction in AI: toward a computational theory of perceptions. Artificial Intelligence Magazine **22**(1) (2001) 73–84
3. Zadeh, L.A.: Fuzzy sets. Information and Control **8**(3) (1965) 338–353
4. Alonso, J.M., Castiello, C., Lucarelli, M., Mencar, C.: Modeling interpretable fuzzy rule-based classifiers for medical decision support. In: Medical Applications of Intelligent Data Analysis: Research Avancements. IGI GLOBAL (2012) 255–272
5. Trillas, E., Eciolaza, L.: Fuzzy Logic: An Introductory Course for Engineering Students. Springer (2015)
6. Bargiela, A., Pedrycz, W.: Human-Centric Information Processing Through Granular Modelling. Studies in Computational Intelligence. Springer, Berlin, Heidelberg (2009)
7. Alonso, J.M., Ramos-Soto, A., Reiter, E., van Deemter, K.: An exploratory study on the benefits of using natural language for explaining fuzzy rule-based systems. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, Italy (2017) 1–6 http://dx.doi.org/10.1109/FUZZ-IEEE.2017.8015489.
8. Gunning, D.: Explainable Artificial Intelligence (XAI). Technical report, Defense Advanced Research Projects Agency (DARPA), Arlington, USA (2016) DARPA-BAA-16-53.
9. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016) http://www.deeplearningbook.org.
10. Finger, M., Portmann, E.: What Are Cognitive Cities? In Portmann, E., Finger, M., eds.: Towards Cognitive Cities. Volume 63 of Studies in Systems, Decision and Control. Springer International Publishing (2016) 1–11
11. Halliday, M.A.K., Matthiessen, M.I.M.: Construing Experience through Meaning: A Language-based Approach to Cognition. Continuum (1999)
12. Michalski, R.S.: A theory and methodology of inductive learning. Artificial Intelligence **20**(2) (1983) 111–161
13. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems **90**(2) (1997) 111–127
14. Bargiela, A., Pedrycz, W.: Granular computing: an introduction. Kluwer Academic Publishers (2003)
15. Zadeh, L.A.: Is there a need for fuzzy logic? Information Sciences **178**(13) (2008) 2751–2779
16. Novak, M.: Cognitive Cities. In: Intelligent Environments. Elsevier (1997) 386–420
17. Mostashari, A., Arnold, F., Mansouri, M., Finger, M.: Cognitive cities and intelligent urban governance. Network Industries Quarterly **13**(3) (2011) 4–7
18. Epstein, S.L.: Wanted: Collaborative intelligence. Artificial Intelligence **221** (4 2015) 36–45
19. Alonso, J.M., Castiello, C., Mencar, C.: The role of interpretable fuzzy systems in designing cognitive cities. In Portmann, E., Seising, R., Tabachi, M., eds.: Designing Cognitive Cities: Linking Citizens to Computational Intelligence to Make Efficient, Sustainable and Resilient Cities a Reality. Studies in Systems, Decision and Control. Springer Verlag (2017) 1–21
20. Trivino, G., Sugeno, M.: Towards linguistic descriptions of phenomena. International Journal of Approximate Reasoning **54** (2013) 22–34
21. Reiter, E., Dale, R.: Building natural language generation systems. Cambridge University Press (2000)
22. Conde-Clemente, P., Alonso, J.M., Trivino, G.: Towards automatic generation of linguistic advice for saving energy at home. Soft Computing (2016) 1–15 http://dx.doi.org/10.1007/s00500-016-2430-5.

23. Conde-Clemente, P., Alonso, J.M., Trivino, G.: rLDCP: R package for text generation from data. In: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, Italy (2017) 1–6 http://dx.doi.org/10.1109/FUZZ-IEEE.2017.8015487.
24. Alonso, J.M., Castiello, C., Mencar, C.: Interpretability of Fuzzy Systems: Current Research Trends and Prospects. In Kacprzyk, J., Pedrycz, W., eds.: Springer Handbook of Computational Intelligence. Springer Berlin / Heidelberg (2015) 219–237
25. Hüllermeier, E.: Does machine learning need fuzzy logic? Fuzzy Sets and Systems **281** (2015) 292–299
26. Dubois, D., Prade, H.: Possibility theory and its applications: Where do we stand? In Kacprzyk, J., Pedrycz, W., eds.: Springer Handbook of Computational Intelligence. Springer Berlin / Heidelberg (2015) 31–60
27. Zadeh, L.A.: Toward human level machine intelligence - Is it achievable? the need for a paradigm shift. IEEE Computational Intelligence Magazine **3**(3) (2008) 11–22
28. Zadeh, L.A.: A Note on Z-numbers. Information Sciences **181**(14) (2011) 2923–2932