Formal Methods Meet XAI: the Tool DEGARI 2.0 for Social Inclusion

Antonio Lieto[1], Gian Luca Pozzato[1], Manuel Striani[1], Stefano Zoia[1] and Rossana Damiano[1]

[1]Dipartimento di Informatica, Università di Torino, Via Pessinetto 12, 10149, Italy

Abstract

We exploit the Description Logic $\mathbf{T}^{\mathrm{CL}}$ in order to develop a diversity-seeking affective recommender system. The tool DEGARI 2.0 (Dynamic Emotion Generator And ReclassIfier) is an explainable, affective-based, art recommender that allows to classify and to suggest, to museum users, cultural items able to evoke not only the very same emotions of already experienced or preferred objects, but also novel items sharing different emotional stances. The system has been tested on the community of deaf people and on the collection of the GAM Museum of Turin, obtaining promising results.

Keywords

Description Logics, Nonmonotonic Reasoning, Explainable AI, Recommender Systems

1. Introduction

In this work, we present DEGARI 2.0, a Description Logic based recommender system, aimed at bridging the differences in the experience of art between different communities, including people with sensory impairments like the Deaf; the latter, indeed, represent the target group of our system and of its evaluation. Our system, introduced in [1], aims at overcoming the limitations of traditional recommendation approaches by exploiting a novel, publicly available, ontological version of Plutchik's model of emotions [2], equipped with opposition and similarity relations between (basic and complex) emotions, as established in the Plutchik's theory. To this aim, we exploit the Description Logic $\mathbf{T}^{\mathrm{CL}}$, a nonmonotonic extension of the typicality logic $\mathcal{ALC} + \mathbf{T_R}$ introduced in [3] for tackling the problem of conceptual combination.

In practice, DEGARI 2.0 employs such ontological structure to suggest museum items not only labeled with the same emotions, but - as mentioned - also to group and recommend artworks evoking *similar* (but not exactly the same) emotions or *opposite* emotions. This kind of alternation in the content suggestion mechanism aims at leading to more comprehensive exploration and fruition of museum collections. Indeed, suggesting museum items evoking

OVERLAY 2022: 4th Workshop on Artificial Intelligence and Formal Verification, Logic, Automata, and Synthesis, November 28, 2022, Udine, Italy
antonio.lieto@unito.it (A. Lieto); gianluca.pozzato@unito.it (G. L. Pozzato); manuel.striani@unito.it (M. Striani); stefano.zoia@unito.it (S. Zoia); rossana.damiano@unito.it (R. Damiano)
https://www.antoniolieto.net (A. Lieto); http://www.di.unito.it/~pozzato/ (G. L. Pozzato); http://www.di.unito.it/~rossana/ (R. Damiano)
0000-0002-8323-8764 (A. Lieto); 0000-0002-3952-4624 (G. L. Pozzato); 0000-0002-7600-576X (M. Striani); 0000-0002-5797-9542 (S. Zoia); 0000-0001-9866-2843 (R. Damiano)

different emotions from the ones already experienced via the fruition of other artworks, is based on the notion of perspective taking [4], i.e. seeing the world (e.g. an exhibition in this case) from other perspectives. Since this approach is used to promote empathy, cohesion and inclusion across social groups, reaching this goal would represent a huge advancement with respect to the current technologies (e.g. like social media or standard recommender systems) that often lead people toward content that fits their own viewpoint, promoting fragmentation and fostering confirmation biases, instead of cohesion, inclusive reflection, and critical thinking.

2. The $\mathbf{T}^{\text{CL}}$ Logic for Combining Prototypes

The core component of DEGARI 2.0 relies on a probabilistic extension of a Description Logic called $\mathbf{T}^{\text{CL}}$ (Typicality-based Compositional Logic), introduced in [5, 3]. This framework allows one to describe and reason upon an ontology with commonsense (i.e. *prototypical*) descriptions of concepts, as well as to dynamically generate novel prototypical concepts in a knowledge base as the result of a human-like recombination of the existing ones [6].

The logic $\mathbf{T}^{\text{CL}}$ is the result of the integration of two main features: (i) the extension of a nonmonotonic Description Logic of typicality $\mathcal{ALC}+\mathbf{T_R}$, introduced in [7, 8], with a distributed semantics based on the DISPONTE semantics of [9] and restricted to typicality inclusions; (ii) the adoption of a well established heuristics inspired by cognitive semantics for concept combination and generation [10] where, in order to formalize a dominance effect between the concepts to be combined, for every combination we distinguish: a HEAD, representing the stronger element of the combination (i.e. the one from which we want to inherit more properties in the final output of the combination), and one or more MODIFIERS. In the logic $\mathbf{T}^{\text{CL}}$, typical properties can be directly specified by means of a *typicality operator* $\mathbf{T}$ enriching the underlying Description Logic, and a knowledge base can contain inclusions of the form $p\ ::\ \mathbf{T}(C) \sqsubseteq D$ to represent that "typical Cs are also Ds", where p is a real number between 0.5 and 1, representing the probability of finding elements of C being also D. The resulting knowledge base is a triple $\langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{R}$ contains standard, *rigid* inclusions of the form $C \sqsubseteq D$ (all Cs are also Ds), $\mathcal{T}$ contains typicality inclusions $p\ ::\ \mathbf{T}(C) \sqsubseteq D$ and $\mathcal{A}$ is the ABox containing facts about individuals, e.g. $C(a)$ (a is a member of concept C). From a semantic point of view, we consider models equipped by a preference relation among domain elements as in [7], where $x < y$ means that x is "more normal" than y, and that the typical members of a concept C are the minimal elements of C with respect to this relation. An element x is a *typical instance* of a given concept C if x belongs to the extension of the concept C, written $x \in C^{\mathcal{I}}$, and there is no element in $C^{\mathcal{I}}$ "more normal" than x. In order to perform useful nonmonotonic inferences, we consider the stronger semantics introduced in [7], where entailment is restricted to a class of *minimal canonical* models, intuitively those minimizing the atypical instances of concepts. The resulting logic corresponds to a notion of *rational closure* built on the top of $\mathcal{ALC} + \mathbf{T_R}$.

The logic $\mathbf{T}^{\text{CL}}$ extends $\mathcal{ALC} + \mathbf{T_R}$ with the distribution semantics known as DISPONTE [9], which is able to deal with probabilities equipping inclusions and allowing us to describe the notion of *scenario* [3]: intuitively, a scenario is a knowledge base obtained by considering all rigid properties in $\mathcal{R}$ as well as all ABox facts in $\mathcal{A}$, but only a subset of typicality properties in $\mathcal{T}$. The idea is to assume that each typicality inclusion is independent from each other in order to define a probability distribution over *scenarios*: roughly speaking, a scenario is obtained by choosing, for each typicality inclusion of $\mathcal{T}$, whether it is considered as true of false. Reasoning can then be restricted to either all or some scenarios. We also equip each scenario with a probability, easily obtained as the product, for each typicality inclusion, of the probability p in case the inclusion is involved, $(1-p)$ otherwise. It immediately follows that the probability of a scenario introduces a probability distribution over scenarios, that is to say the sum of the probabilities of all scenarios is 1.

In the logic $\mathbf{T}^{\text{CL}}$, in order to deal with the problem of combining prototypical descriptions of concepts as in [3], we adopt typicality inclusions in order to formalize typical properties for both the HEAD and the MODIFIERS concepts, and then to exploit the DISPONTE semantics in order to select *only* some typical properties belonging to them characterizing the combined concept. The preferential semantics underlying the logic $\mathbf{T}^{\text{CL}}$, together with the HEAD-MODIFIER heuristics, are able to tackle the problem of conflicting properties.

Formally, given a knowledge base $\mathcal{K} = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$ and given two concepts C_H and C_M occurring in $\mathcal{K}$, our logic allows one to define the compound concept C as the combination of the HEAD C_H and the MODIFIER C_M, where $C \sqsubseteq C_H \sqcap C_M$ and the typical properties of the form $\mathbf{T}(C) \sqsubseteq D$ to ascribe to the concept C are obtained in the set of scenarios that: 1. are consistent; 2. are not *trivial*, in the sense that the scenarios considering *all* typical properties of the HEAD that can be consistently ascribed to C are discarded; 3. are those giving preference to the typical properties of the HEAD C_H (with respect to those of the MODIFIER C_M) with the highest probability. The set of scenarios remaining are those selected by the logic $\mathbf{T}^{\text{CL}}$ as the result of the procedure. The knowledge base obtained as the result of combining concepts C_H and C_M into the compound concept C is called C-*revised* knowledge base:

$$\mathcal{K}_C = \langle \mathcal{R}, \mathcal{T} \cup \{p \,:\, \mathbf{T}(C) \sqsubseteq D\}, \mathcal{A} \rangle,$$

for all D such that $\mathbf{T}(C) \sqsubseteq D$ belongs to the selected scenario(s).

3. The tool DEGARI 2.0

DEGARI 2.0 exploits the logic $\mathbf{T}^{\text{CL}}$ in order to provide an ontological formalization of the circumplex theory of emotions devised by the cognitive psychologist Robert Plutchik [11], [2]. According to this theory, emotions, and their interconnections, can be represented on a wheel, in which the affective distance between different emotional states is a function of their radial distance. The Plutchik's ontology, formalizing such a theory, encodes emotional categories in a taxonomy, representing: basic or primary emotions; complex (or compound) emotions; opposition between emotions; similarity between emotions. In particular, by following Plutchik's account, complex emotion are considered as resulting from the composition of two basic emotions (where the pair of basic emotions involved in the composition is called a dyad). The compositions occurring between similar emotions (adjacent on the wheel) are called primary dyads. Pairs of less similar emotions are called secondary dyads (if the radial distance between them is 2) or tertiary dyads (if the distance is 3), while opposites cannot be combined.

The information about the emotional concepts and their corresponding features to combine via $\mathbf{T}^{\text{CL}}$ are extracted from the NRC Emotion Intensity Lexicon [12]: such lexicon provides a list of English words, each with real-values representing intensity scores for the eight basic emotions of Plutchik's theory. The intensity scores were obtained via crowd-sourcing, using best-worst scaling annotation scheme. This lexicon associates words to emotional concepts in descending order of emotional intensity and, for our purposes, we considered the most intensively associated terms for each basic emotion as typical features of such emotion. In this way, the prototypes of the basic emotions were formed, and the $\mathbf{T}^{\text{CL}}$ reasoning framework is used to generate the compound emotions. Such prototypes of basic emotions are formalized by means of a $\mathbf{T}^{\text{CL}}$ knowledge base, whose TBox contains both *rigid* inclusions of the form

$$BasicEmotion \sqsubseteq Concept,$$

in order to express essential desiderata but also constraints, as an example $Joy \sqsubseteq PositiveEmotion$ as well as *prototypical* properties of the form

$$p \ :: \ \mathbf{T}(BasicEmotion) \sqsubseteq TypicalConcept,$$

representing typical concepts of a given emotion, where p is a real number in the range $(0.5, 1]$, expressing the frequency of such a concept in items belonging to that emotion: for instance, $0.72 \ :: \ \mathbf{T}(Surprise) \sqsubseteq Delight$ is used to express that the typical feature of being surprised contains/refers to the emotional concept *Delight* with a probability/degree of belief of the 72%.

Once the association of lexical features to the emotional concepts in the Plutchik's ontology is obtained and the compound emotions are generated via the logic $\mathbf{T}^{\text{CL}}$, the system is able to reclassify the cultural items in the novel formed emotional categories. Intuitively, an item belongs to the new generated emotion if its metadata (name, description, title) contain all the rigid properties as well as at least the 30% of the typical properties of such a derived emotion. The 30% threshold was empirically determined: i.e., it is the percentage that provides the better trade-off between over-categorization and missed categorizations [13, 14].

4. Conclusions

We have tested DEGARI 2.0 with members of Istituto dei Sordi and on the collection of the GAM Museum of Turin. The experiments provided in [1] show that the effort of tackling diversity-seeking, affective-based and explainable museum recommendations received a moderate, improvable, acceptance from the deaf community. This is an encouraging result considering the challenge of the cognitive barriers involved in the process of the accepting suggestions that do not fit one's own preferences and viewpoints.

Experiments concerning the perceived explainability of the provided categorization lead to some key elements emerged as guidelines to design and improve the next generation of inclusive and transparent AI systems, potentially going beyond the specific needs of the deaf community. In this regard, it is important to point out how state of the art neural systems and language models, like SenticNet 7, do not have, as a built-in, this feature. It represents, however, one of the major requirements for modern AI systems interacting with the humans (see the recent General Data Protection Regulation (GDPR) that emphasized the users' right to explanation [15]).

References

[1] A. Lieto, G. L. Pozzato, M. Striani, S. Zoia, R. Damiano, Degari 2.0: A diversity-seeking, explainable, and affective art recommender for social inclusion, Cognitive Systems Research 77 (2023) 1–17. URL: https://www.sciencedirect.com/science/article/pii/S1389041722000456. doi:https://doi.org/10.1016/j.cogsys.2022.10.001.

[2] R. Plutchik, The nature of emotions, American scientist 89 (2001) 344–350.

[3] A. Lieto, G. L. Pozzato, A description logic framework for commonsense conceptual combination integrating typicality, probabilities and cognitive heuristics, Journal of Experimental and Theoretical Artificial Intelligence 32 (2020) 769–804. doi:10.1080/0952813X.2019.1672799.

[4] T. Pedersen, A. Wecker, T. Kuflik, P. Mulholland, B. Diaz-Agudo, Introducing empathy into recommender systems as a tool for promoting social cohesion, in: Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station, USA, CEUR Workshop Proceedings, 2021.

[5] A. Lieto, G. L. Pozzato, A description logic of typicality for conceptual combination, in: M. Ceci, N. Japkowicz, J. Liu, G. A. Papadopoulos, Z. W. Ras (Eds.), Foundations of Intelligent Systems - 24th International Symposium, ISMIS 2018, Limassol, Cyprus, October 29-31, 2018, Proceedings, volume 11177 of *Lecture Notes in Computer Science*, Springer, 2018, pp. 189–199. URL: https://doi.org/10.1007/978-3-030-01851-1_19. doi:10.1007/978-3-030-01851-1_19.

[6] A. Lieto, F. Perrone, G. L. Pozzato, E. Chiodino, Beyond subgoaling: A dynamic knowledge generation framework for creative problem solving in cognitive architectures, Cognitive Systems Research 58 (2019) 305–316.

[7] L. Giordano, V. Gliozzi, N. Olivetti, G. L. Pozzato, Semantic characterization of rational closure: From propositional logic to description logics, Artificial Intelligence 226 (2015) 1–33. URL: https://doi.org/10.1016/j.artint.2015.05.001. doi:10.1016/j.artint.2015.05.001.

[8] L. Giordano, V. Gliozzi, N. Olivetti, G. L. Pozzato, Rational closure in SHIQ, in: M. Bienvenu, M. Ortiz, R. Rosati, M. Simkus (Eds.), Informal Proceedings of the 27th International Workshop on Description Logics, Vienna, Austria, July 17-20, 2014, volume 1193 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2014, pp. 543–555. URL: http://ceur-ws.org/Vol-1193/paper_20.pdf.

[9] F. Riguzzi, E. Bellodi, E. Lamma, R. Zese, Reasoning with probabilistic ontologies, in: Q. Yang, M. J. Wooldridge (Eds.), Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, AAAI Press, 2015, pp. 4310–4316. URL: http://ijcai.org/Abstract/15/613.

[10] J. A. Hampton, Inheritance of attributes in natural concept conjunctions, Memory & Cognition 15 (1987) 55–71.

[11] R. Plutchik, A general psychoevolutionary theory of emotion, in: Theories of emotion, Elsevier, 1980, pp. 3–33.

[12] S. Mohammad, Word affect intensities, in: N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, T. Tokunaga (Eds.), Proceedings of the Eleventh International Conference

on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018, European Language Resources Association (ELRA), 2018. URL: http://www.lrec-conf.org/proceedings/lrec2018/summaries/329.html.

[13] E. Chiodino, D. Di Luccio, A. Lieto, A. Messina, G. L. Pozzato, D. Rubinetti, A knowledge-based system for the dynamic generation and classification of novel contents in multimedia broadcasting, in: G. De Giacomo, A. Catalá, B. Dilkina, M. Milano, S. Barro, A. Bugarín, J. Lang (Eds.), ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020, volume 325 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2020, pp. 680–687. URL: https://doi.org/10.3233/FAIA200154. doi:10.3233/FAIA200154.

[14] A. Lieto, G. L. Pozzato, S. Zoia, V. Patti, R. Damiano, A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification, Knowledge Based Systems 227 (2021) 107166. doi:10.1016/j.knosys.2021.107166.

[15] B. Goodman, S. Flaxman, European union regulations on algorithmic decision-making and a “right to explanation”, AI magazine 38 (2017) 50–57.