

Concepts, logic, and cognitive adequacy

Guendalina Righetti¹

¹Free University of Bozen-Bolzano, piazza Università, 1, 39100, Bolzano-Bozen, Italy

Abstract

The project aims to study and develop a computational logic framework inspired and informed by the theories of concepts as they can be found across the disciplines of Cognitive Science and Experimental Psychology. It concentrates on the representation of different phenomena permeating human cognition, focusing mainly on aspects linked to the categorisation task (e.g. typicality effect, exceptions handling, concept vagueness) and the concept combination task (dominance effect, overextension, underextension, alignment of features).

Keywords

Description Logic, Weighted Logic, Cognitive adequacy, Typicality effects, User evaluation

1. Motivation and state of the art

What is a concept? This question is meaningful in philosophy, cognitive sciences, psychology, and in AI, however answers rarely converge. Knowledge Representation (KR) is a relevant goal whether the aim is explaining our cognitive processes or realizing intelligent applications or even understanding what characterizes our knowledge from an ontological or epistemological viewpoint. How to deal with and interpret the notion of ‘concept’ is, however, still an unresolved and hotly debated topic.

Classical Logic mostly interprets the meaning of concepts in terms of their extensions: concepts are represented as sets of objects, and concepts’ extensions are precisely defined. For each object it would be possible to specify whether it falls within the definition of the concept or not. At the same time, the operations on concepts are usually conceived as set theoretic operations, so that the combination of two concepts is usually understood as a simple set intersection. Moreover, Classical Logic is normally characterised by the principle of compositionality, according to which any complex expression, is understood as a function of the parts it is composed of, plus a set of syntactic operations to combine them. Classical logic then comes equipped with the so called *Classical Theory* of concepts. According to the Classical Theory, a concept is definable by a set of individually necessary and jointly sufficient conditions: everything satisfying that set of conditions is considered an instance of that concept; and, vice versa, so that something is considered an instance of a concept, it must satisfy that set of conditions. Category membership is then neatly determined: borderline cases, vagueness and imprecision are excluded.


Psychological evidence, however, demonstrated a more complex and diverse reality. It turned out that human concepts are not characterised by the neatness supposed by the Classical Theory.

CHIItaly 2021 Joint Proceedings of Interactive Experiences and Doctoral Consortium, July 11–13, 2021, Bolzano, Italy

 guendalina.righetti@stud-inf.unibz.it (G. Righetti)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Some experiments showed that it is not always clear whether an object belongs to a category or not, and that the distinction between members and non-members of a category forms essentially a continuum of choices. The important work of Eleanor Rosch [1] demonstrated that concepts show *typicality effects*: some class members are more representative than others (e.g. cats are more typical *Pet* than iguanas). It was also shown that the spectrum of concept combinations is much wider than what is expressible simply in terms of intersections of sets [2]. Early KR systems oriented towards conceptual modeling tried to take into account indications from psychological research. In the 1970s, both the semantic networks of Quillian [3] and Minsky's frame systems [4] were developed. Both can be considered associative models definable in terms of network structures. These systems were extremely simple and allowed the representation of prototypical information associated with ordinary concepts. Nevertheless, their lack of formal semantics made the study of their properties increasingly difficult [5].

From the attempt to give a formal semantical grounding to frames and semantic networks, the field of Description Logics developed, which is to a large extent the study of well-behaved subsets of first-order predicate logic associated with the usual Tarskian semantics for first-order logic. Being decidable fragments of first-order logic, however, DLs are again fragments of purely extensional, classical logic. The use of truth-conditional semantics also involves an implicit admission of the principle of compositionality, which, according to the well-known argument of Pet-Fish proposed by Fodor¹ [6], contrasts with the representation of typicality effects. Description Logics face, just as Classical Logic, a problem of lack of expressiveness. Several extensions to Description Logic have been proposed to overcome the described restrictions: from logics for probabilistic reasoning [7], to fuzzy logics [8]. Extensions of DLs dealing with typicality effects and reasoning with exceptions can be divided into two (overlapping) groups. Many works make use of the notion of *defeasible subsumption* [9, 10, 11], having their roots in the seminal work of Kraus, Lehmann and Magidor [12] on non-monotonic reasoning. Other works, instead, explicitly introduce a "typicality operator" within the language, in order to select the most typical instances of a concept [13, 14, 15]. In both cases, the notion of typicality is normally simply taken from granted and assumed within the model. Also, both kinds of approaches implies the use of a preferential semantics, so that typicality is reduced to an order relation over the elements of the domain - a quite counter-intuitive way of treating typicality.

2. Tooth operator and cognitive adequacy

It therefore remains an important goal to study and develop a computational logic framework informed by the theories of concepts as they can be found across the disciplines of Cognitive Science and Experimental Psychology. We concentrate here on the representation of different phenomena permeating human cognition, focusing mainly on aspects linked to the categorisation task and the concept combination task. Categorization and concept combination are indeed considered basic cornerstones of cognition, according to which the validity of theories of concepts is tested [16]. Providing a formal and computational model for categorisation

¹The concept of "pet-fish" result from the composition of the concepts "pet" and "fish". The prototype of the concept "pet" may be furry, the one of "fish" is probably greyish, while the prototype of "pet-fish" is neither furry, nor greyish.

and concept combination able to take into account cognitive phenomena would thus improve the connections between experimental science and Artificial Intelligence. Part of the work has been focused on extending Description Logics to model cognitively relevant features of classification and concept combination. In order to do that, we introduced, in a series of papers, a new logical operator, the so-called “Tooth operator”, which allows introducing weights into the standard representation language. This choice was guided by the intuition that different properties, or features, may have different importance in the definition of a concept, and then also in identifying instances and their typicality. For instance, in defining the concept of an elephant (and in classifying instances as ‘elephants’), having a trunk may be considered more important than having a tail. This idea was first proposed in [17], where we introduced a family of operators within \mathcal{ALC} , one of the most widely used Description Logic formalism. These operators, in particular, apply to sets of concept descriptions and return a composed concept whose instances are those that satisfy “enough” of the listed concept descriptions. To provide a formal meaning of “enough”, the operator takes a list of weighted concepts as argument, as well as a threshold. The combined concept applies to every instance whose sum of the weights of the concepts it satisfies meets the threshold. The study of the formal properties of the operator was further carried out in [18, 19], where a link between this formalism and linear classification models was showed.

The design of the Tooth operator was inspired by the Prototype Theory [1]. For this reason, it is said to be *cognitively grounded*. This alone, however, has little to say about its *cognitive adequacy*. The notion of cognitive adequacy is a very diversified one, as it applies to distinct contexts with a slightly different meaning. In cognitive modelling, a first notion of cognitive adequacy relates to “phenomenon adequacy”: the focus is on the ability of a system to replicate experimental data and cognitive phenomena, as they are observed and studied within the field of experimental psychology and cognitive science. In the context of categorisation, a paradigmatic example is the one of typicality, on which, as mentioned above, a large number of logical and formal approaches have been tested. Rather obviously, typicality is however just one of many phenomena observed, and according to which the adequacy of a modelling framework may be tested. Overextension in conjunction and attributes emergence [2], situational effects [20], dominance effect [2] are all additional phenomena that hardly reconcile with compositionality, and integrating them in a logic-based framework is an open challenge.

In [17], the tooth operator was shown to be able to represent the typicality effect in the classification task, as well as to represent fine-grained dependencies among the attributes that define a concept. In [21, 22], the usefulness of the proposed approach to model cognitively relevant features of categorization and concept combination was further analysed, proposing an analysis and representation of some of the cognitive phenomena linked to concept combination previously mentioned (namely overextension, underextension and dominance [2]).

3. Further directions:

There is also another nuance for the notion of cognitive adequacy, which pairs well with the one of understandability - namely the capability of a representation framework to be intuitive and readable also for non-experts, making it then more interpretable and less error-prone.

Understandability and cognitive adequacy are related because if a representation framework is similar to the representation system in the human mind, using and understanding it should be consequently easier.

The notion of understandability (or comprehensibility) has gained popularity in recent years, and this is also due to the increasing interest in Explainable AI. How to precisely characterise understandability is however far from being obvious, and there is, in general, no consensus on a precise definition. It is further normally assumed that measuring understandability implies the use of “human-grounded metrics” [23]. To operationalise this idea, different strategies have been adopted in the literature. In order to measure a system’s understandability, subjects are usually presented with (at least) two different representation frameworks and asked to perform the same task (often, a classification task) in the two different ‘environments’. Across the different studies, the evaluation metrics can then differ, but normally range between 3 parameters, namely accuracy (how many times did the subjects reply correctly?), speed (how fast they were?), and confidence (how confident did they feel in the reply?). In [24], for instance, the understandability of decision tables, binary decision trees, and propositional rules is tested, and the evaluation is carried out combining the metrics of accuracy, speed, and confidence of the interpretation. [25] focus on the interpretability of decision tree models and rule-based models, using *perceived understandability* as the only metric for the evaluation (which is somehow similar to what is elsewhere called confidence). [26] compare different propositional theories and evaluate their interpretability in terms of accuracy, confidence, and speed. The same parameters are also taken into account in [27], to measure the understandability of decision trees. Following this line of research, we plan to experimentally compare the understandability of Tooth-formulas and disjunctive normal form (DNF) formulas², adopting a setting analogous to the one proposed in [27], and evaluating subjects performances on accuracy, speed, and confidence in the reply. We plan two kinds of experiments. The first one addressed to logic-experts, to directly present and evaluate the two formalisms. The second one directed at a more general audience, translating both DNF and Tooth expressions into natural language: this would allow the evaluation of the *algorithm* of classification behind the two formalisms. We argue that such a study will provide a new perspective, and a well-rounded description, on the cognitive adequacy of the proposed approach.

References

- [1] E. Rosch, Principles of categorization, Rosch, E., et al., 1978, pp. 27–48.
- [2] J. A. Hampton, Inheritance of attributes in natural concept conjunctions, *MemoryCognition* 15 (1987) 55–71.
- [3] M. R. Quillian, Word concepts: A theory and simulation of some basic semantic capabilities, *Behavioral Science* 5 (1967) 410–430.
- [4] M. Minsky, *A Framework for Representing Knowledge*, Winston P. H., 1975.
- [5] H. L. R. Brachman, *Readings in Knowledge Representation*, Los Altos, CA: Morgan Kaufmann, 1985.

²DNF is conventionally considered as a standard/benchmark in term of both expressivity and interpretability of logic-based knowledge representation.

- [6] J. Fodor, The present status of the innateness controversy, J. Fodor, 1981.
- [7] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann Pub., 2014.
- [8] R. R. Yager, L. A. Zadeh, An Introduction to Fuzzy Logic Applications in Intelligent Systems, Springer Science+Business Media, 2012.
- [9] K. Britz, J. Heidema, T. Meyer, Modelling object typicality in description logics, in: *Description Logics*, volume 477 of *CEUR Workshop Proc.*, 2009.
- [10] G. Casini, U. Straccia, Rational closure for defeasible description logics, in: *Logics in Artificial Intelligence, JELIA 2010*, volume 6341, 2010, pp. 77–90.
- [11] K. Britz, I. J. Varzinczak, Rationality and context in defeasible subsumption, in: *FoIKS 2018*, volume 10833 of *Lecture Notes in Computer Science*, 2018, pp. 114–132.
- [12] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic Reasoning, Preferential Models and Cumulative Logics, *Artificial Intelligence* 44 (1990) 167–207.
- [13] L. Giordano, V. Gliozzi, N. Olivetti, G. L. Pozzato, Preferential description logics, in: *Logic for Programming, AI, and Reasoning, LPAR 2007*, volume 4790, 2007, pp. 257–272.
- [14] K. Britz, T. Meyer, I. J. Varzinczak, Semantic foundation for preferential description logics, in: *AI 2011: Advances in Artificial Intelligence*, 2011, pp. 491–500.
- [15] A. Lieto, G. L. Pozzato, A description logic of typicality for conceptual combination, in: *Foundations of Intelligent Systems - ISMIS*, 2018, pp. 189–199.
- [16] G. L. Murphy, *The Big Book of Concepts*, Cambridge, MA: The MIT Press, 2002.
- [17] D. Porello, O. Kutz, G. Righetti, N. Troquard, P. Galliani, C. Masolo, A toothful of concepts: Towards a theory of weighted concept combination, in: *Proc. of DL Workshop*, 2019.
- [18] P. Galliani, O. Kutz, D. Porello, G. Righetti, N. Troquard, On knowledge dependence in weighted description logic, in: *GCAI 2019, Proc.*, volume 65, 2019, pp. 68–80.
- [19] P. Galliani, G. Righetti, O. Kutz, D. Porello, N. Troquard, Perceptron connectives in knowledge representation, in: *Knowledge Engineering and Knowledge Management, EKAW 2020*, Springer, 2020, pp. 183–193.
- [20] L. Barsalou, Grounded Cognition, *Annual Review of Psychology* 59 (2008) 617–645.
- [21] G. Righetti, D. Porello, O. Kutz, N. Troquard, C. Masolo, Pink panthers and toothless tigers: three problems in classification, in: *Proc. of AIC Workshop*, 2019, pp. 39–53.
- [22] G. Righetti, P. Galliani, O. Kutz, D. Porello, C. Masolo, N. Troquard, Weighted description logic for classification problems, in: *GCAI 2019 Proc.*, 2019, pp. 108–112.
- [23] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, 2017.
- [24] J. Huysmans, K. Dejaeger, C. Mues, J. Vanthienen, B. Baesens, An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models, *Decis. Support Syst.* 51 (2011) 141–154.
- [25] H. Allahyari, N. Lavesson, User-oriented assessment of classification model understandability, in: *SCAI 2011 Proc.*, volume 227, IOS Press, 2011, pp. 11–19.
- [26] S. Booth, C. Muise, J. Shah, Evaluating the interpretability of the knowledge compilation map, in: S. Kraus (Ed.), *Proc. of IJCAI*, 2019, pp. 5801–5807.
- [27] R. Confalonieri, T. Weyde, T. R. Besold, F. M. del Prado Martín, TREPAN reloaded: A knowledge-driven approach to explaining black-box models, in: *ECAI 2020*, IOS Press, 2020, pp. 2457–2464.