Uncertainty Modelling in Automated Concept Formation

Flaminia Cavallo Simon Colton Alison Pease

Computational Creativity Group, Department of Computing, Imperial College, London,
f.cavallo11@imperial.ac.uk

Abstract: Categorisation and classification are areas that have been well studied in machine learning. However, the use of cognitive theories in psychology as a basis to implement a category formation system designed for creative purposes and based on human behaviour is still largely unexplored. Our aim in this project is to verify how some of the ideas on uncertainty and ambiguity in classification could influence concept classification in an automated theory formation system.

1 Introduction

Our research aim in this project is to investigate how influential psychological theories of human concept formation, such as those described in [8], can be interpreted for the automation of creative acts. In particular, we choose to focus on the automated theory formation system HR developed by Colton et al. [1]. This program works by combining given concepts according to a set of productions rules and heuristic measures to build a theory. Here, concepts are represented by logic predicates (definitions), and by a set of constants (examples) that satisfy these predicates. This representation corresponds to the one proposed by the classical view in the field of conceptualization in cognitive psychology [10]. As many cognitive psychologists point out, this representation is inadequate if we want to represent concepts in a human like manner because:

- There is no way to distinguish between categories' members, and to take into account the typicality of an item with respect to the category.
- It does not take into account in-between categories cases: items that partly belong to more than one category.
- It does not consider how knowledge and high-level perceptions, such as beliefs, goals and context, influence categorization.

Our aim in this project is to extend HR to enable it to operate over real word examples in a human-like way. The final scope is bidirectional: we aim to determine both how ideas suggested by the cognitive psychology community can be used to improve and extend automated concept formation techniques, and also to clarify the notions put forward in the psychology literature research by providing results and analysis from experiments undertaken.

2 HR

HR is an Automated Theory Formation program which takes a theory, conceived as a set of initial concepts, and applies a set of production rules on it in order to construct new concepts. These production rules take as an input the definition of one or two concepts and output the definition for the new concept. For example, the match production rule equates two variables in a definition, and the negate rule negates certain clauses in a definition. Once the new concept definition is created, HR calculates the success set of the definition by collating all tuples of objects which satisfy the definition. The set of positive examples is then used to make conjectures about the new concept, in the form of equivalence conjectures, implication conjectures, or non-existence conjectures. Conjectures are either proved by the OTTER theorem prover [5] or rejected because of a counterexample found by the MACE model generator [6]. The theory is then enriched with either the new theorem or with the newly found counter-examples. HR follows a best-first non-goal-oriented search. This is dictated by an ordered agenda and a set of heuristic rules used to evaluate the interestingness of each concept. The scope of HR is to form interesting clausal theories, starting with some minimal knowledge and enriching it by performing both inductive and deducting reasoning.

3 Project Outline

According to the prototype view in the field of cognitive psychology, humans categorize an item by comparing it with each known category's most typical item (real or imaginary), which is also called a prototype [3]. The similarity between each pair is used to determine the new object's typicality with respect to every category and can be interpreted as a measure of how much the item belongs to the respective categories. Prototypes are represented as schemata reporting the features which we gradually learn are the most frequent and relevant to each category. Prototypes hence are flexible entities and they are influenced by the categorization process itself. We have decided to explore how these observations can influence category creation and theory formation by including a similar notion into HR. To do so, we will assign a degree of membership to every tuple of objects constituting an example of a concept. This parameter will represent the percentage with respect to which the example belongs to the concept, and hence will be directly proportional to its typicality. Note that this implies that an item does not need to fully belong to just one category - in-between categories cases are allowed. The idea is similar to the one used in fuzzy logic, where the degree

of membership of an item is determined by a membership function. However, in our case, this membership function will be flexible over time, depending on the already classified category members and modified every time a new item is classified within a category.

4 Calculation of Typicality

Dunmore [2] observed how concept definitions are chosen and developed according to their use by presenting a study on the definition of prime numbers. This concept was initially defined as "a number which is only divisible by 1 and itself". The number 1 satisfies this definition, and hence it is considered a positive example. However, 1 constitutes a counterexample to many conjectures about primes, for example the Fundamental Theorem of Arithmetic, stating that every natural number is either a prime or can be represented uniquely as a product of primes. In the current version of HR, the above counterexample would be enough to make this conjecture false. A different approach is to review the concept definition of prime numbers itself, for example by considering a different definition such as "number with exactly two divisors". Another example has been proposed by Lakatos [4] who, in the attempt to prove Euler's conjecture, reported 5 different definitions for polyhedra.

In our system, we will allow a concept to have multiple definitions. In the examples above, prime numbers would have two definitions and polyhedra would have 5 definitions. The typicality of an item with respect to a concept will then be calculated according to three measures:

- The number of concepts' definitions that the item satisfies. In the prime number case, 1 satisfies 50% of the definitions, 3 satisfies all definitions and 4 satisfies no definition. In the polyhedra case, the only polyhedra that satisfy all five definition are the regular polyhedra.

- The amount of tweaking that each definition would need in order to include the item. In the prime number example above, the definition "number with exactly two divisors" could be modified to "number with exactly one divisor" or "number with maximum two divisors" in order to include 1. The amount of tweaking will be measured by the number of HR's production rules that would be involved in modifying the definition. The salience of each part of the definition will also be taken into consideration. To calculate it we will take inspiration from psychological studies that underline the importance of both the relevance and the familiarity of attributes in a similarity task [9, 7].

- The number of conjectures about the concept that the item supports, in a similar way to what Lakatos suggested in [4]. In the example above, the typicality of 1 will decrease as we discover that 1 is a counterexample of the Fundamental Theorem of Arithmetic.

The typicality measures over a category success set would help us recognize that conjectures like the Fundamental Theorem of Arithmetic are probably true, as they are true for most typical examples. Moreover, these observations will change our beliefs on what the correct definition of a prime number is. The uncertainty over these beliefs would then be used as new properties about the concept are discovered, and as new more complicated concepts are constructed.

5 Conclusion

This project is still at a preliminary stage. However, we can see it leading to results that can be applied in different areas. For example, the program could be used as a learning and data-mining system on large datasets representing human behaviours which are classifiable, whose properties are not known, and to which a computer could actively contribute in a creative way. A possible application follows the lines of the example given above: the study of how mathematical concepts, conjectures and proofs gradually evolve as more things are discovered about them.

References

[1] S. Colton. *Automated Theory Formation in Pure Mathematics*. Springer-Verlag New York, 2002.

[2] C. Dunmore. *Meta-level revolutions in mathematics*. In Gillies, D., editor, Revolutions in Mathematics, pages 209225. Clarendon Press, Oxford., 1992.

[3] Rosch E. Natural categories. *Cognitive Psychology*, 4(3):328 – 350, 1973.

[4] I. Lakatos. *Proofs and Refutations*. CUP, Cambridge, UK, 1976.

[5] W. McCune. *The OTTER user's guide*. Technical Report ANL/90/9, Argonne National Laboratories, 1990.

[6] W. McCune. *A Davis-Putnam program and its application to finite first-order model search*. Technical Report ANL/MCS-TM-194, Argonne National Laboratories, 1994.

[7] D. L. Medin and J. G. Bettger. Presentation order and recognition of categorically related examples. *Psychonomic Bulletin and Review*, 1:250–254, 1994.

[8] G. Murphy. *The Big Book of Concepts*. MIT Press, 2002.

[9] A. Ortony. Beyond literal similarity. *Psychological Review*, 86(3):161–180, 1979.

[10] E. E. Smith and D. L Medin. *Categories and concepts*. Cambridge, MA: Harvard University Press, 1981.