Symmetric Parallel Class Expression Learning

TRAN, Cong An

School of Engineering and Advanced Technology
Massey University
New Zealand

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

June, 2013

For my grandparents, my parents, my wife and my son

Acknowledgements

I would like to take this opportunity to thank my supervisors, Prof. Hans W. Guesgen, Prof. Jens Dietrich and Prof. Stephen Marsland for their guidance and support. I am thankful that I could benefit from their combined research experience. Prof. Hans W. Guesgen introduced me the world of Ambient Intelligent with its humanistic ideas of helping the elderly to have a better life. Prof. Jens Dietrich attracted me to his colourful world of Software Engineering. His endless source of ideas impressed me that this is the world of the active and creative people. Prof. Stephen Marsland, a 'diverse' professor with the excellent background in mathematics and computer science, encouraged me into his mysterious world of Machine Learning.

I also want to thank Jens Lehmann, the research group leader of Machine Learning and Ontology Engineering (MOLE) Group at the University of Leipzig, as well as the author of the DL-Learner framework, for his kindness to grant me permission to access his DL-Learner framework repository as well as for his valuable advice to evaluate learners. I also highly appreciate Prof. Adrian Paschke, Director of RuleML Inc., for his advice on my research direction.

I would like to thank the Ministry of Education and Training of Vietnam for awarding me the scholarship for my study. Thanks also go to the School of Engineering and Advanced Technology (SEAT) for the financial support that has enabled me to attend international conferences, which not only improved my research but also enriched my life.

Thanks to all members of Massey University Smart Home (MUSE) research group for their valuable discussions and feedbacks on my research. I also

want to thank Michele Wagner for her administrative support throughout my degree.

This thesis would not be possible without the warm and support of my Palmy-based family, Mr. Vo The Truyen and his family, Mr. Nguyen Buu Huan and Mr. Nguyen Van Long, who give me a home away from home. In particular, my special thanks go to Mr. Truyen's for treating me as their little brother and Mr. Huan for checking and giving me many helpful advices on my writing. I could not have asked for more warm and kind friends as them.

Finally, I would like to give all my deepest gratitude and respect to my family. To my grandparents and my parents for their continuous love, support and patience. To my parents-in-law for taking good care of my son in more than four years, which released me from family concerns to concentrate on my research. To my sister and brother for being with my parents during my study. To my beloved wife, Nguyen Thu Huong, and my son, Tran Cong Huan. It is my fortune to have them in my life. They are always by my side to share all the laughers and tears.

Abstract

The growth of both size and complexity of learning problems in description logic applications, such as the Semantic Web, requires fast and scalable description logic learning algorithms. This thesis proposes such algorithms using several related approaches and compares them with existing algorithms. Particular measures used for comparison include computation time and accuracy on a range of learning problems of different sizes and complexities.

The first step is to use parallelisation inspired by the map-reduce framework. The top-down learning approach, coupled with an implicit divide-and-conquer strategy, also facilitates the discovery of solutions for a certain class of complex learning problems. A reduction step aggregates the partial solutions and also provides additional flexibility to customise learnt results.

A symmetric class expression learning algorithm produces separate definitions of positive (true) examples and negative (false) examples (which can be computed in parallel). By treating these two sets of definitions 'symmetrically', it is sometimes possible to reduce the size of the search space significantly. The use of negative example denotions enhances learning problems with exceptions, where the negative examples ('exceptions') follow a few relatively simple patterns.

In general, correctness (true positives) and completeness (true negatives) of a learning algorithm are traded off against each other because these two criteria are normally conflicting. Particular learning algorithms have an inherent bias towards either correctness or completeness. The use of negative definitions enables an approach (called fortification in this thesis) to improve predictive correctness by applying an appropriate level of over-specialisation to the prediction model, while avoiding over-fitting.

The experiments presented in the thesis show that these algorithms have the potential to improve both the computation time and predictive accuracy of description logic learning when compared to existing algorithms.

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES