MURDOCH UNIVERSITY

SCHOOL OF INFORMATION TECHNOLOGY

Application of the Recommendation Architecture Model for Text Mining

Uditha Ratnayake B.Sc. (Eng.) (Hons)

This thesis is presented for the degree of Doctor of Philosophy of Murdoch University

October 2003

Declaration

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

Uditha Ratnayake

October 2003

Acknowledgement

I am very grateful to Prof. Tamás (Tom) Gedeon and Dr. Graham Mann, my principal supervisors, for their constant support and inspiring guidance. Tom's expertise and encouragement throughout the research process and the development of the thesis were invaluable. Graham's constructive feedback, enthusiasm and insight made a huge difference to the progress of the thesis. I also thank Dr. Nalin Wickramarachchi, my supervisor in Sri Lanka, for his advice while I worked in Sri Lanka.

My heartfelt gratitude extends to Andrew Coward for his constructive comments on my work in various phases and for many stimulating discussions. His patience in explaining various concepts of the Recommendation Architecture and the help given for programming the prototype are truly appreciated.

My thanks also extend to my colleagues Alex and Kevin for providing useful advice and moral support during my stay at Murdoch. I am indebted to Madu, my husband, for his continuous support, encouragement and patience. Finally, I would like to thank my mother, father and family members for their encouragement and assistance.

List of Publications

The following publications were derived from this research in applying the Recommendation Architecture for the domain of text mining.

Refereed Journal papers

1. U. Ratnayake, T. D. Gedeon, "Extending The Recommendation Architecture Model for Text Mining", *International Journal of Knowledge-Based Intelligent Engineering Systems*, Vol 7, 3, pp. 139-148, July 2003.

2. U. Ratnayake, T. D. Gedeon, N. Wickramarachchi, "Application of the Recommendation Architecture Model for Text Mining", *Australian Journal of Intelligent Processing Systems*, (under review).

Refereed Conference Papers

1. U. Ratnayake, T. D. Gedeon, "Application of the Recommendation Architecture Model for Document Classification", *Proceedings of the 2nd WSEAS International Conference on Scientific Computation and Soft Computing*, pp. 326-331, Crete, 2002.

2. U. Ratnayake, T. D. Gedeon, "Application of the Recommendation Architecture Model for Discovering Associative Similarities in Text", *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP 2002)*, pp. 2059-2063, Singapore, 2002.

3. U. Ratnayake, T. D. Gedeon, "Extending The Recommendation Architecture Model For Effective Text Classification", *Proceedings of The Sixth Australia-Japan Joint Workshop on Intelligent and Evolutionary Systems,* pp. 185-191*,* Canberra, Australia, 2002

4. U. Ratnayake, T. D. Gedeon, N. Wickramarachchi, "Document Classification with the Recommendation Architecture: Extensions for Feature Intensity Recognition and Column Labelling", *Proceedings of the 7th Australasian Document Computing Symposium*, pp. 31-37, Sydney, Australia, 2002.

Abstract

The Recommendation Architecture (RA) model is a new connectionist approach simulating some aspects of the human brain. Application of the RA to a real world problem is a novel research problem and has not been previously addressed in literature. Research conducted with simulated data has shown much promise for the Recommendation Architecture model's ability in pattern discovery and pattern recognition. This thesis investigates the application of the RA model for text mining where pattern discovery and recognition play an important role.

The clustering system of the RA model is examined in detail and a formal notation for representing the fundamental components and algorithms is proposed for clarity of understanding. A software simulation of the clustering system of the RA model is built for empirical studies. In the argument that the RA model is applicable for text mining the following aspects of the model are examined. With its pattern recognition ability the clustering system of the RA is adapted for text classification and text organization. As the core of the RA model is concerned with pattern discovery or identification of associative similarities in input, it is also used to discover unsuspected relationships within the content of documents. How the RA model can be applied to the problems of pattern discovery in text and classification of text is addressed demonstrating results from a series of experiments. The difficulties in applying the RA model to real life data are described and several extensions to the RA model for optimal performance are proposed from the insights obtained from experiments. Furthermore, the RA model can be extended to provide user-friendly interpretation of results. This research shows that with the proposed extensions the

RA model can be successfully applied to the problem of text mining to a large extent. Some limitations exist when the RA model is applied to very noisy data, which are also demonstrated here.

Table of Contents

List of Figures

List of Tables