

Preface

Recently, a new field has emerged taking benefit of both domains: Data Mining (DM) and Natural Language Processing (NLP). Indeed, statistical and machine learning methods hold a predominant position in NLP research¹, advanced methods such as recurrent neural networks, Bayesian networks and kernel based methods are extensively researched, and "may have been too successful (...) as there is no longer much room for anything else"². They have proved their effectiveness for some tasks but one major drawback is that they do not provide human readable models. By contrast, symbolic machine learning methods are known to provide more human-readable model that could be an end in itself (e.g., for stylistics) or improve, by combination, further methods including numerical ones. Research in Data Mining has progressed significantly in the last decades, through the development of advanced algorithms and techniques to extract knowledge from data in different forms. In particular, for two decades Pattern Mining has been one of the most active field in Knowledge Discovery.

This volume contains the papers presented at the ECML/PKDD 2014 workshop: DMNLP'14, held on September 15, 2014 in Nancy. DMNLP'14 (Workshop on Interactions between Data Mining and Natural Language Processing) is the first workshop dedicated to Data Mining and Natural Language Processing cross-fertilization, *i.e.* a workshop where NLP brings new challenges to DM, and where DM gives future prospects to NLP. It is well-known that texts provide a very challenging context to both NLP and DM with a huge volume of low-structured, complex, domain-dependent and task-dependent data. The objective of DMNLP is thus to provide a forum to discuss how Data Mining can be interesting for NLP tasks, providing symbolic knowledge, but also how NLP can enhance data mining approaches by providing richer and/or more complex information to mine and by integrating linguistic knowledge directly in the mining process.

Out of 23 submitted papers, 9 were accepted as regular papers amounting to an acceptance rate of 39%. In addition to regular contributions, two less mature works, which were still considered valuable for discussion, were accepted as posters and appear as extended abstract in this volume.

The high quality of the program of the workshop was ensured by the much-appreciated work of the authors and the Program Committee members. Finally, we wish to thank the local organization team of ECML/PKDD 2014, and more specifically Amedeo Napoli and Chedy Raïssy, and the ECML/PKDD 2014 workshop chairs Bettina Berendt and Patrick Gallinari.

September 2014

Peggy Cellier, Thierry Charnois
Andreas Hotho, Stan Matwin
Marie-Francine Moens, Yannick Toussaint

¹ D. Hall, D. Jurafsky, and C. M. Manning. Studying the History of Ideas Using Topic Models. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pp. 363–371, 2008

² K. Church. A Pendulum Swung Too Far. Linguistic Issues in Language Technology, Vol. 6, CSLI publications, 2011.