



PhD Position: Argumentation Mining with Deep Learning and Random Forest

Do argumentative patterns influence helpfulness of online reviews?

University of Liege (Belgium)

What is Argumentation Mining?

Argumentative mining (AM) is a relatively novel and very challenging task in Natural Language Processing (NLP). The main aim in AM is to automatically detect argumentative structures within texts, typically in the form of discourse. Examples of argumentative patterns include justification, premise, conclusions and counter-arguments.

Several AM techniques have been proposed in scientific literature. For instance, Ashley and Walker (2013) and Mochales and Moens (2011) developed techniques for mining argumentative patterns from legal documents, while Park and Cardie (2014) applied AM techniques to user comments. Annotated corpora for AM also exist (Reed et al. 2008). These corpora are particularly useful for training supervised learning methods for AM.

The PhD Project

The overall goal of this PhD project is to contribute to and advance the nascent field of AM. The selected candidate will have the opportunity to develop novel AM algorithms and techniques, which will then be applied to online reviews to evaluate the performance of the proposed algorithms and determine whether the presence of argumentative patterns are predictive of the helpfulness of online reviews.

The objective of this PhD project is thus two-fold: (a) scientific, i.e. to develop novel AM techniques based on state-of-the-art machine learning techniques, especially Deep Learning and

Random Forest and (b) business application/managerial, i.e. to investigate the correlation between the presence of argumentative patterns and review helpfulness (are reviews containing a larger proportion of argumentative patterns likelier to be voted as most helpful?)

Machine Learning Techniques and Brief Project Description

The proposed PhD project intends to adopt a machine learning approach to AM rather than a purely symbolic or linguistic approach. In this respect, the selected candidate will be expected to investigate the performance of two recent machine learning paradigms, Deep Learning (with Neural Networks) and Random Forests. The candidate will develop a sound understand of these paradigms and proposed novel Deep Learning and Radom Forest algorithms for AM.

Other recent machine learning paradigms, in particular, minimally-supervised learning and distant supervision, will also be investigated in the context of AM.

The proposed algorithms will be applied to online reviews (see Corpora section below) in order to automatically detect argumentative patterns from the textual contents. The influence of the discovered patterns on the review helpfulness will then be studied. Specifically, the selected candidate will apply regression techniques to determine whether the presence of argumentative patterns influence the helpfulness score of reviews, in other words, whether reviews containing a larger proportion of argumentative patterns are considered more helpful.

Corpora

Several corpora are already available for evaluating and applying the AM algorithms developed in the project, including product and service reviews from Amazon (books, movies), YELP and TripAdvisor. In addition, access to the Penn Discourse TreeBank (PDTB) is also available.

International Collaboration

The selected candidate will have the opportunity to work in close collaboration with other scholars in Japan, The Netherlands, Belgium, and France.

Requirements and Application Procedure

We are looking for highly motivated and relatively independent candidates. The candidates should at the minimum be proficient in:

- Programming: any language of their choice, with a preference for Java, R, Python
- Machine Learning at Masters Level
- Quantitative Methods: time-series, regression, statistics, vector/matrix algebra
- Good command of written and spoken English

Interested candidates are required to send their CV and a letter of motivation via e-mail to Prof. Ashwin Ittoo, ashwin.ittoo@ulg.ac.be by 23rd November 2015. Please clearly indicate 'PhD application' in the subject line.

Only those candidates deemed most suitable for the research position will be contacted. The starting date is negotiable, but we would prefer candidates who could start soonest.

Other Useful Information

This PhD project will be conducted at the HEC Management School of the University of Liège, in Belgium.

Liège (Dutch: Luik, German: Lüttich) is a modern city with a rich historical background. Sitting on the banks of the Meuse River, it is the birthplace of Emperor Charlemagne, the Father of Europe. The city is a cultural hub in the Wallonia region with a unique character, an eclectic mix of architecture from the Middle Ages to the present, a dramatic setting, exciting nightlife, a number of museums, and varied natural surroundings.

Liège is also a centrally located city; it is only a few miles away from Maastricht, , one hour by train from Brussels and Cologne, two hours from Paris, Luxembourg and the Flemish cities of the north, and about three hours from London. The ultra modern Liège-Guillemins train station, designed by the famous architect Santiago Calatrava, is an architectural masterpiece, and a major high-speed train hub in Europe.

References

1. Ashley, K. D., & Walker, V. R. (2013, June). Toward constructing evidence-based legal arguments using legal decision documents and machine learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law* (pp. 176-180). ACM.
2. Mochales, R., & Moens, M. F. (2011). Argumentation mining. *Artificial Intelligence and Law*, 19(1), 1-22.
3. Park, J., & Cardie, C. (2014). Identifying appropriate support for propositions in online user comments. *ACL 2014*, 29.
4. Reed, C., Mochales Palau, R., Rowe, G., & Moens, M. F. (2008). Language resources for studying argument. In *Proceedings of the 6th conference on language resources and evaluation-LREC 2008* (pp. 91-100).