Argumentative Zoning - An Industry Practitioner View

Karan Praharaj

January 21st

There are obvious problems of lack of organization for research papers based on their content. The content of research is multivariate and multi-faceted. Despite this, there are no concrete ways (apart from citation lists) to extract a knowledge graph, which researchers might use to make themselves aware of similar, related or contrasting work. One of the two proposals we came up with as industry practitioners was to design a research traversing system built on top of a multi-relational graph by incorporating the argumentative zoning approach advanced by the paper discussed this week [1]. The idea is to use the annotation scheme to build a knowledge-base segmented by different categories, and then subsequently generating graph edges between nodes that have a similarity score over a particular preset threshold.

For example, for the category *AIM*, we would want to generate bidirectional edges for work that has the same research objective, or is trying to focus on the same problem space. But the most fundamental and obvious question to this idea would be to ask what "same" entails in this context? It is very fuzzy to say that two problem statements are the same or similar, without defining "sameness" in quantitative terms. But first of all, to even evaluate similarity in any meaningful *non-manual* way, we need to convert the semantic information of the content into machine interpretable information, like continuous vector representations. This generation of sentence representations can be done by means of skip-thought vectors, or recursive auto-encoders. The representations are subsequently compared against one another, possibly by being fed as input into a neural network for classification. The neural network would make decisions about similarity based off

of the pre-defined similarity threshold. In this way, we would have multiple edge links between different research papers (*nodes*). Research works with the same aims and objectives would be connected by an "AIM" bidirectional edge; articles which cite other works for their contrasting ideas or approaches could be connected via a unidirectional edge from the citing paper to the *citee*. Similarly, papers which use similar approaches could also be connected using bidirectional edges.

References

[1] Simone Teufel, Advaith Siddharthan, and Colin Batchelor. Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1493–1502, Singapore, August 2009. Association for Computational Linguistics.