

ABSTRACT

Lexical semantics focuses on semantic understanding of lexical units like words, sub-words, phrases, etc. Dealing with lexical semantics is one of the basic steps towards natural language understanding. Being motivated by the idea of using networks (graphs) to represent the language units and their relationships, researchers have applied network theory in NLP tasks like evaluating machine generated summaries, detection of ambiguity in a text, part-of-speech (POS) induction, word sense induction, etc. In this thesis, our focus is to explore the scope of utilizing networks for improving semantic understanding of lexical units. We attempt this investigation using two types of networks: i) external networks like Wikipedia network, citation network, etc., ii) networks prepared from the given corpus like distributional thesaurus network, hierarchical hyponym-hypernym network, etc.

In our exploration with the external networks, we attempt to use inherent information from networks for tasks like Wikification and Wikipedia enrichment. For Wikification of scientific articles we show the utilization of metapath based approach for bringing in extra information from citation network, author publication network, etc. whereas for enrichment of scientific wikipedia articles we show the usefulness of wikilink network for borrowing references.

Next, we show the utilization of complex network measures obtained from the distributional thesaurus network for lexical relation detection. Specifically, we come up with complex network measures to build supervised models for co-hyponymy detection and novel sense detection. We focus our study on discriminating co-hyponymy relation from other lexical relations like hypernymy, meronymy, etc. For novel sense detection, given two corpora from different time periods, we detect if a word has got a new sense in the later time point. We use the changes in

the network features, captured from the underlying distributional networks for these time points, to precisely capture if a word has got a new sense in the later time point.

Finally, we explore the use of network embedding methods for word representation as well as compositionality detection in noun phrases. We propose a pathway to improve the state-of-the-art word representation by combining with representations obtained from distributional thesaurus network, via network embedding methods. We also devise a simple, efficient approach for combining distributional (Distributional Semantic Model) and hypernymy information (Poincaré model) for the task of compositionality detection of noun phrases to achieve the state-of-the-art performance.

In summary, this thesis is an endeavour towards pushing the boundaries of state-of-the-art models dealing with lexical semantics tasks by making use of the networks in various forms.

Keywords: Lexical Semantics, Wikification, Wikipedia Enrichment, Lexical Relation Detection, Word Representation, Compositionality Detection, Wikilink Network, Citation Network, Distributional Thesaurus network, Complex Network Measure, Network Embedding, Poincaré Embedding.