Thesis advisor(s)
Mike Rosner
Dr Tim Baldwin

Author
Spandana Gella

Word usages and patterns in social media

Abstract

Many words in any natural language have more than one meaning. Most of the work related to understanding word meaning in context is based on word sense disambiguation (WSD). Traditional WSD approaches relied on a lexical resource or a sense inventory where each sense is mapped to one of the best fitting senses defined in the resource or sense inventory. In this thesis we investigate and develop an approach for understanding word meanings in contexts over social media texts. In particular we deviate from traditional word sense disambiguation and we target usage similarity, an alternative to WSD for understanding meaning over social media texts. We investigate usage similarity using a topic-modeling based approach in which each usage of a word in a context is represented as a multinomial distribution of topics learned from background collection of documents. We create a gold-standard dataset to evaluate our approach. After evaluating the results over multiple background collections we conclude that it is possible to estimate usage similarity over social media texts. We execute a pilot sense tagging task and analyse sense patterns observed over social media texts. We show future directions which if followed might increase the performance of proposed usage similarity approach.