

Information Extraction from Biomedical Texts: Learning Models with Limited Supervision

Marie-Francine Moens
KULeuven, Leuven, Belgium
sien.moens@cs.kuleuven.be

Abstract

Among the application domains of information extraction, the biomedical domain is one of the most important ones. This is due to the large amount of biomedical text sources including the vast scientific literature and collections of patient reports written in natural language. These sources contain a wealth of crucial knowledge that needs to be mined. Typical mining tasks regard entity recognition, entity-relation extraction, and event and event participant recognition. Recently we witness an interest in the recognition of spatial relationships between entities and of temporal relationships between events. One of the most important problems in information extraction regards dealing with a limited amount of examples that are manually annotated by experts and that can be used for training the extraction models.

In this talk we discuss how we can leverage knowledge contained in unlabelled texts and ontological knowledge about known relationships between the output labels used for the extractions. The former aspect especially focuses on how to automatically create novel training examples from the unlabelled data, the latter on how to integrate the relationships in models for structured machine learning during training and testing of the extraction models in the most efficient way. We show promising results and point to directions of future research.