

Extracting Biological Processes with Global Constraints

Author 1

XYZ Company
111 Anywhere Street
Mytown, NY 10000, USA
author1@xyz.org

Author 2

ABC University
900 Main Street
Ourcity, PQ, Canada A1A 1T2
author2@abc.ca

Abstract

Reasoning over processes is fundamental for language understanding applications such as Question Answering. In this paper we propose a method for extracting relations between events in a process. We annotate 150 paragraphs describing biological processes and show that by taking advantage of the global structure of a process we can substantially improve performance. In addition, we release our data set.

1 Introduction

Motivation: Being able to reason over processes is crucial for language understanding applications. Consider the question

Some question that can only be answered by having some structure over a process

A human reading the paragraph in Figure ?? would be able to describe that etc. However, traditional state-of-the-art systems that are based on retrieval would fail. In this paper we suggest a method for extracting a process structure that will allow answering of complex questions. For example, knowing that... can help us answer the question.

Relation to previous work: Extracting processes is related to two lines of works in Information Extraction - event extraction and timeline construction. Recent work in event extraction (Riedel and McCallum, 2011; McClosky et al., 2011) is based on BioNLP challenges and focuses on extraction of a closed set of events such as *regulation* and *phosphorylation* from a single sentence and their relations to

proteins. However, a process is typically described over multiple sentences and involves a large number of possible events. Work on timeline construction (Do et al., 2012; McClosky and Manning, 2012) requires partially ordering a set of events that is described in a sequence of sentence. However, fully capturing process structure requires a rich set of relations (*cause*, *super*) that is missing from this line of work.

Emphasizing this work: In this paper, we find the structure of a biological process by extracting the relations between the process events. Properties of our task (a) spans multiple sentences (b) open-ended set of events (c) rich set of relations comparing to timeline construction (d) the nature of the text - it is a textbook rather than abstracts. (e) We do not use domain-specific knowledge. Some sentence that says that by doing this we will be able to answer the complex question from the beginning - linking to language understanding.

Technical contribution Processes have a global structure and we want to take advantage of that when extracting the relations. For example, all events a process description are connected to one another and there are various constraints such as if two event mentions refer to the same event then they must be related in a similar way to a third event. Similar to many recent works in NLP () we model global constraints using ILP, however since many of the constraints can be violated we use soft constraints. We show that by encoding global constraints we can substantially improve performance.

Contributions Three main contributions

- We define the task of process extraction - what

is a process

- We propose a method for process extraction that uses global constraints and show that it improves performance
- We release a set of 150 biological processes, annotated by biologists.

structure Background, Definition of a process, Local classifier (old features and new features), Global model, Experiments and maybe analysis

2 Related Work

BioNLP work

Timeline construction work.

Scripts work - Chambers, Poon 2013.

Work that uses global constraints with ILP or dual decomposition or whatever.

References

- Quang Do, Wei Lu, and Dan Roth. 2012. Joint inference for event timeline construction. In *EMNLP-CoNLL*, pages 677–687.
- David McClosky and Christopher D. Manning. 2012. Learning constraints for consistent timeline extraction. In *EMNLP-CoNLL*, pages 873–882.
- David McClosky, Mihai Surdeanu, and Christopher D. Manning. 2011. Event extraction as dependency parsing. In *ACL*, pages 1626–1635.
- Sebastian Riedel and Andrew McCallum. 2011. Fast and robust joint models for biomedical event extraction. In *Proceedings of the Conference on Empirical methods in natural language processing (EMNLP '11)*.