

A brief introduction to event extraction

Huiling You

University of Oslo

huiliny@ifi.uio.no

Media
Futures.

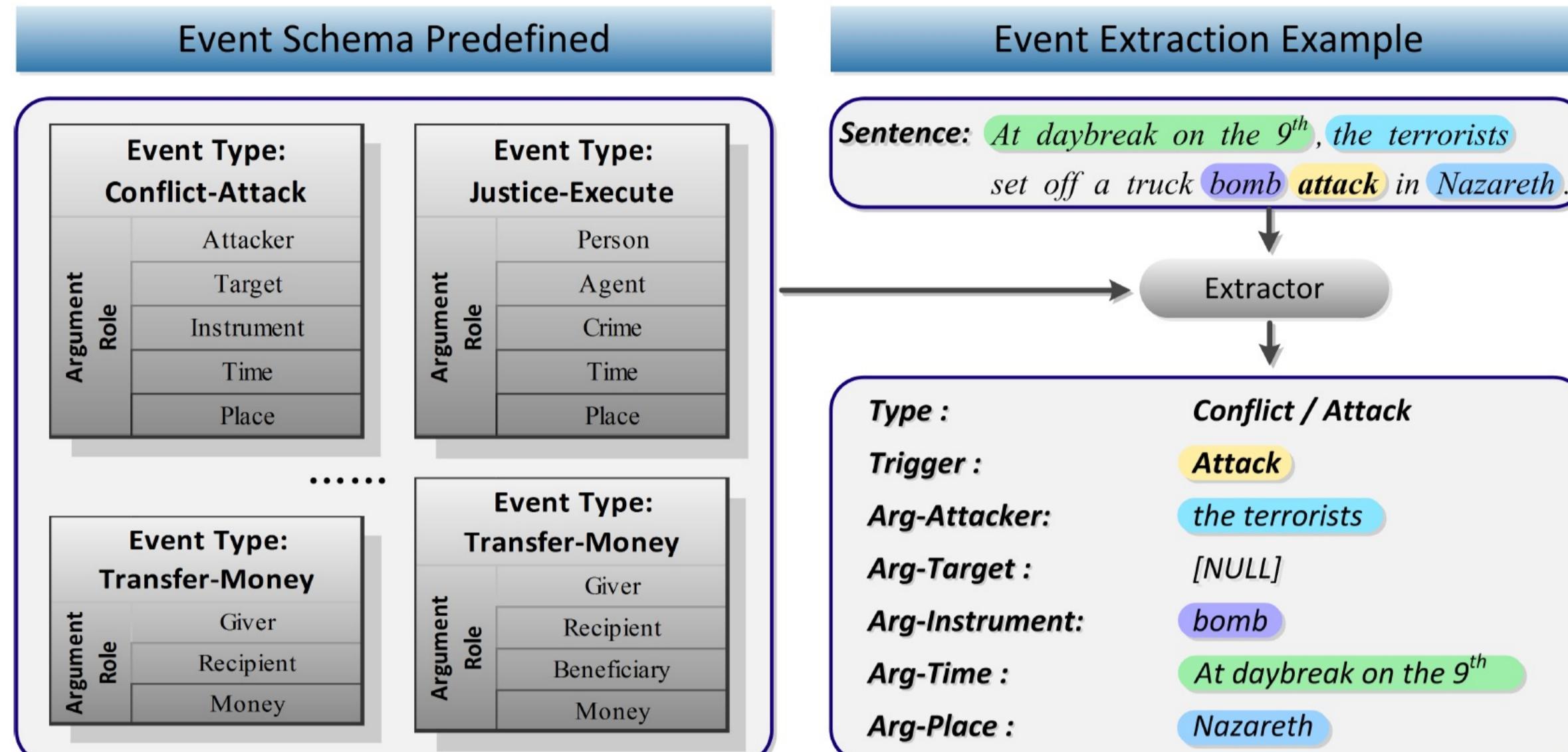


Figure 1: An example of closed-domain event extraction, taken from Figure 1 by Xiang and Wang (2019)¹

Motivation

In natural language processing (NLP), *event extraction* aims at detecting event instance(s) in texts and identifying the event type together with its participants and attributes, if existing. The obtained structured representations of events can further be used in diverse tasks, such as to expand existing knowledge base, to monitor social events, and so on.

Event extraction tasks

There are two main tasks in event extraction.

- **Closed-domain** event extraction uses predefined event schema to detect and extract desired event types from text.
- **Open-domain** event extraction aims at detecting events from text, and in many cases, clustering similar events via extracted event keywords.

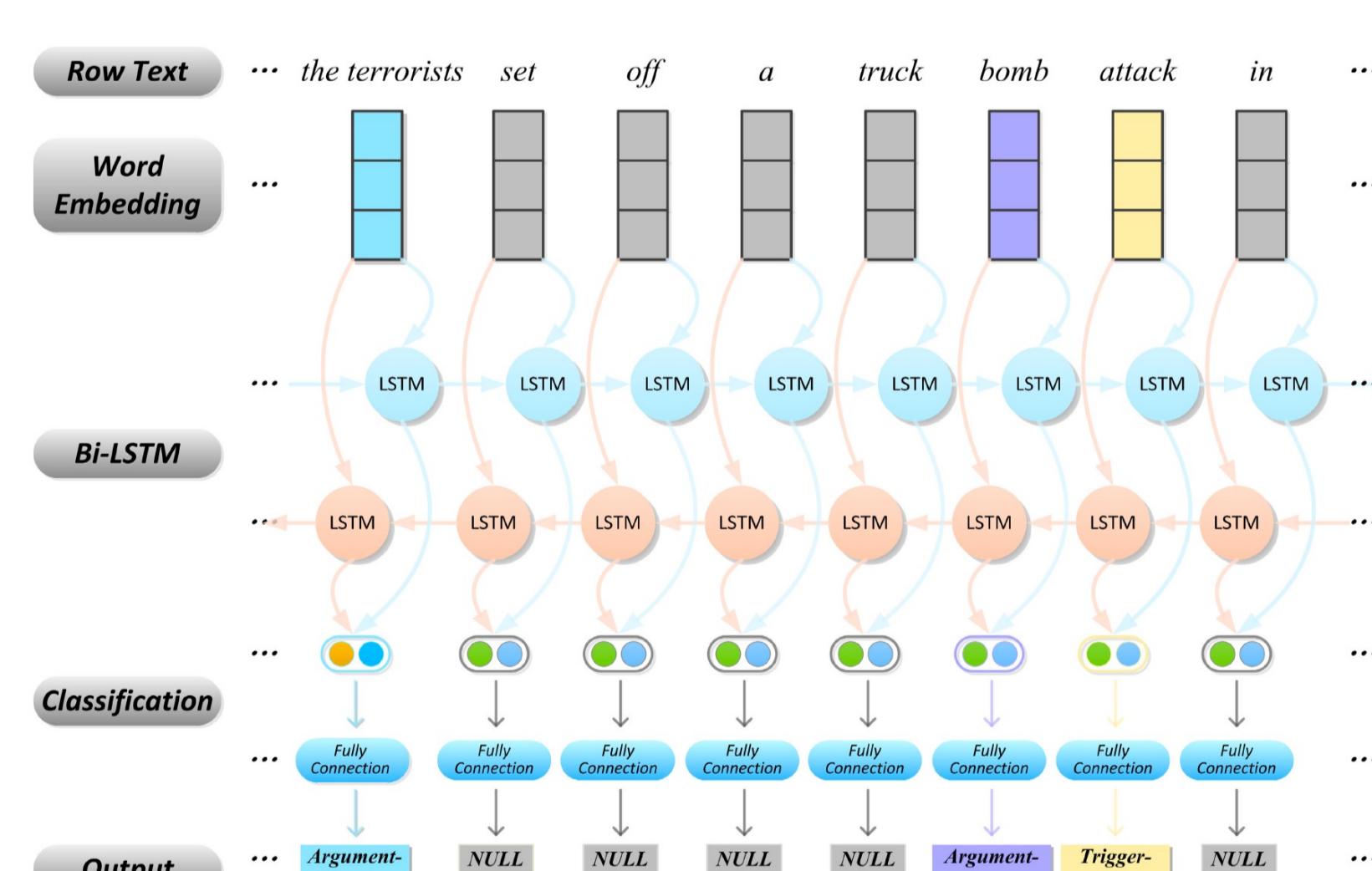


Figure 2: An example of event extraction based on machine learning, taken from Figure 4 by Xiang and Wang (2019)¹

1. Xiang Wei and Bang Wang. A survey of event extraction from text. *IEEE Access*, 7:173111–173137, 2019.

Corpora

The public evaluation programs provide various corpora accompanied by corresponding evaluation tasks, and each corpus is often manually annotated by professionals and experts according to task definition.

- **THE ACE EVENT CORPUS** is annotated for various extraction tasks, including entity, time, value, relation and event extraction.
- **THE TAC-KBP CORPUS** provides the event nugget detection task to discover explicit mentions of event of particular types and subtypes.
- **THE TDT CORPUS** is a useful source for open-domain event extraction, with news stories annotated with hundreds of topics (events) collected from different sources.
- There are also many other domain-specific corpora.

Methodologies

The methodologies of event extraction fall into five groups.

- The earlier approaches are highly based on **pattern matching**, with either manual or automatic pattern construction.
- The **machine learning** approaches adopt classic machine learning algorithms, like support vector machine (SVM), maximum entropy (ME), and others.
- The **deep learning** approaches resort to different neural networks, such as convolutional neural network (CNN), recurrent neural networks (RNN), graph neural networks (GNNs), or neural models based on attention mechanism.
- The **semi-supervised** learning methods make use of a small set of labeled gold data for wider coverage.
- The **unsupervised learning** approaches are adopted for open-domain event extraction, for trigger and arguments detection from unlabeled corpus.

My project

The aim of this PhD project is to:

- Develop neural models for closed-domain event extraction from English text.
- Propose a novel formulation of event extraction as neural dependency graph parsing.
- Adapt neural event extraction models for English to Norwegian.
- Develop and make available state-of-the-art models for event extraction from Norwegian news text.

PARTNERS

amedia

Bergens Tidende

fonn group

HIGHSOFT

IBM

NRK

Schibsted

2

VIMOND

VIZRT



NORCE

HOST



UNIVERSITY OF BERGEN

FUNDER

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskningsrådet

sfi