Survey: Extracting Participant Mention and Event Timeline Creation

Ajay Gupta
Pushpak Bhattacharyya
CFILT,Indian Institute of Technology Bombay, India
{ajaygupta, pb}@cse.iitb.ac.in

Abstract

Information Extraction refers to the automatic extraction of structured information such as entities, relationships between entities, event and temporal relation between events from unstructured sources. Information Extraction is a branch of natural language processing that has a wide range of applications, including question answering, knowledge base population, information retrieval etc. The extraction of structure from noisy, unstructured sources is a challenging task.

1 Introduction

Early systems were rule-based with manually coded rules [(Appelt et al., 1993; Lehnert et al., 1993; Riloff et al., 1993)] . As manual coding of rules became tedious, algorithms for automatically learning rules from examples were developed [(Aitken, 2002)]. As extraction systems were targeted on more noisy unstructured sources, rules were found to be too brittle. Then came the age of statistical learning, where in parallel two kinds of techniques were deployed: generative models based on Hidden Markov Models [(Bikel et al., 1997)] and conditional models based on maximum entropy [(Borthwick et al., 1998; Klein and Manning, 2002)]. Both were superseded by global conditional models, popularly called Conditional Random Fields [(Lafferty et al., 2001)]. As the scope of extraction systems widened to require a more holistic analysis of a document's structure, techniques from grammar construction were developed. In spite of this journey of varied techniques, there is no clear winner. Rule-based methods and statistical methods continue to be used in parallel depending on the nature of the extraction task. There also exist hybrid models that attempt to reap the benefits of both statistical and rule-based methods. Deep learning is the new state-of-the-art paradigm which is widely used in most of the information extraction systems.

2 Participant (Entities) Extraction

The first step in information extraction is to detect the entities or participants in the text. A named entity is, roughly speaking, anything that can be referred to with a proper name: a person, a location, an organization. The term is commonly extended to include things that are not entities, including dates, times, and other kinds of temporal expressions, and even numerical expressions like prices.

Named entity recognition means finding spans of text that constitute proper names and then classifying the type of the entity. Recognition is difficult partly because of the ambiguity of segmentation; we need to decide what's an entity and what is not, and where the boundaries are. The entities are useful in relationship between participants, events and many other information extraction tasks.
To extract this participants we have explored various approaches as explained in below sections:

2.1 Ontology Based Approaches

In the paper [(Buey et al., 2016)] authors are using ontology to extract information from legal documents. They are using 144 Spanish notary acts for extraction of information. The information that they are extracting are grouped into two

- **Document Parameters:** It includes data like title of document, date, location, notary name.

- **Person Parameters:** It includes name, surname, marital state, address, region, country.

The extraction process is guided by an ontology, which stores information about the structure and the content of different types of documents to be processed. Their overall extraction process include "text preprocessing", "text chunking" into different section, "section processing" to extract specific information from these section.

Cristian and Milagro [2017] have developed a resource for the legal domain, by mapping the legal domain ontology called Legal Knowledge Interchange Format (LKIF) and Wikipedia based ontology, YAGO. They have used curriculum learning to train the classifier. They have manually define a mapping between the LKIF and YAGO ontology. Tagged sentence corpus is considered for training and from them only those sentences are considered which have at least one named entity and have more than three mentions on Wikipedia. The curriculum learning is applied as follows, a neural network (one hidden layer smaller than input input) with randomly set weights is trained to distinguish NE vs. non-NE. Once this classifier has converged, the weights obtained are used as the starting point of a classifier with a similar architecture but with more specific classes (Person, Organization, Document, Abstraction, Act, non-NE). Again when this classifier converges, its weights are used for the next level of classification which are LKIF concepts, and finally they classify into YAGO classes.

2.2 Supervised and Unsupervised Approach

In the paper [(Dozier et al., 2009)] authors are basically extracting five types of entities they are judges, attorneys, companies, jurisdiction and courts. For extracting them they are using three different approaches.

1. **Lookup Based:** In this method a gazette is created which contains common names and they are been looked up to recognize the named entities.

2. **Contextual Rules:** In this method some set of rules are created to get instances of named entities. For example one such rule is if the name is preceded by the "Mr." then the word following this would be a name.

3. **Statistical model:** In this approach they just train a model for identifying a named entity and the feature used are similar to contextual rules.

Once the named entities are identified, resolution is performed. Resolution basically mean assigning particular class to extracted names. This they basically do using Support Vector Machine as a classification task.

3 Event Extraction

One facet of information extraction is event extraction (EE): identifying instances of selected types of events appearing in natural language text. For each instance, EE should identify the type of the event, the event trigger (the word or phrase which evokes the event), the participants in the event, and (where possible) the time and place of the event. Various literature that are covered are explained below.

3.1 Dependency Feature Based Event Extraction

McClosky, Mihai and Christopher [2011] proposed a simple approach for the extraction of event structures by taking the tree of event-argument relations and using it directly as the representation in a reranking dependency parser. They have got competitive results and showed that the joint modeling of event structures is beneficial.

In "BioNLP09 Shared Task on Event Extraction" the use of dependency tree is also explored by Buyko [2009]. In the paper [(Miwa et al., 2010)] authors have analyzed how event extraction performance is affected by parser and dependency representation.

3.2 Supervised Approach

Liu and Yubo [2016] pointed out that Frames defined in FrameNet (FN) share highly similar structures with events in ACE event extraction program. An event in ACE is composed of an event trigger and a set of arguments. Analogously, a frame in FN is composed of a lexical unit and a set of frame elements, which play similar roles as triggers and arguments of ACE events respectively. Hence they used FN to get extra event data for training and reported improvement in event detection task.

Nikolaos and Frederique [2010] has proposed a semiautomatic approach for event extraction from legal domain for legal case building and reasoning. They have identified a number of classes of relations among people and organisations that they believe to be of interest to lawyers, during case construction, independently from the litigation domain. Those classes correspond to events or event abstractions and include the following: “is employed by”, “meets”, “says”. They are using Xerox Incremental Parser’s event recognition module with their own logic to extract events. Once “Event Detection”, “Referent Detection” and “Temporal Expression Detection” is done then “Coreference Module” and “Temporal reasoning and normalization module” is used to integrate those information to built “case knowledge model”. This knowledge model now can be consumed by lawyers to query different question and built logical reasoning for case building.

Nguyen and Grishman [2015] studied the event detection problem using convolutional neural networks (CNNs) that overcome the two fundamental limitations of the traditional feature-based approaches to this task: complicated feature engineering for rich feature sets and error propagation from the preceding stages which generate these features. The experimental results show that the CNNs outperform the best reported feature-based systems in the general setting as well as the domain adaptation setting without resorting to extensive external resources.

Nguyen and Kyunghyun [2016] proposed to do event extraction in a joint framework with bidirectional recurrent neural networks, thereby benefiting from the advantages of the two models as well as addressing issues inherent in the existing approaches. They systematically investigate different memory features for the joint model and demonstrate that the proposed model achieves the state-of-the-art performance on the ACE 2005 dataset.

Li, Ji, & Huang [2013] used perceptron model with token-based tagging to jointly extract event triggers and arguments.

4 Temporal Ordering of Events

With both the events and the temporal expressions in a text having been detected, the next logical task is to use this information to fit the events into a complete timeline. Such a timeline would be useful for applications such as question answering and summarization. This ambitious task is the subject of considerable current research but is beyond the capabilities of current systems. A somewhat simpler, but still useful, task is to impose a partial ordering on the events and temporal expressions mentioned in a text. Such an ordering can provide many of the same benefits as a true timeline.

(D’Souza and Ng, 2013) in their paper ”Classifying Temporal Relations with Rich Linguistic Knowledge” has used hybrid approach to classify the Event-Event and Event-time relation into original 14 classes of TimeBank. The rules are data-driven and helps to better handle the skewed distribution of 14 class data of TimeBank. Learning based approach uses various lexical, grammatical, syntactic and semantic based features to train the classification model. Total ____ number of rules are made and out of them only those which have accuracy of 80% were used in one setting and in another all the rules were used. Various accuracies are reported by them with different set of features. Overall with their results they have shown that hybrid systems decreases the error by 15 to 16%.

In (Chambers et al., 2007) paper ”Classifying Temporal Relations Between Events” authors has proposed a two stage event relation classification into six classes. The first stage learns features for individual event. In second stage these these features along with some extra lexical features are used. According to them relation between the events in same sentence follows different distribution then the relation between the inter sentence event. In stage 2 of proposed approach two classification models are build one for intra sentence and another for inter sentence event pair relation. Author uses SVM, Naive Bayes and Maximum Entropy classifiers and report a increment of 3% accuracy on TimeBank corpus.

Conclusion

Participant Extraction, Event Extraction and creating Timeline of Event are some of the most important tasks of Information Extraction. Ontology based, lookup based, contextual rules based and statistical model based literature is covered for Entity Extraction literature. Dependency feature based Event Extraction literature is covered to get some idea about how without any supervision one can extract events. Finally we cover literature for Temporal Event Ordering and try to implement one of the system.

References

James Stuart Aitken. 2002. Learning information extraction rules: An inductive logic programming approach. In *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 355–359. IOS Press.

Douglas E Appelt, Jerry R Hobbs, John Bear, David Israel, and Mabry Tyson. 1993. Fastus: A finite-state processor for information extraction from real-world text. In *IJCAI*, volume 93, pages 1172–1178.

Daniel M Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. 1997. Nymble: a high-performance learning name-finder. In *Proceedings of the fifth conference on Applied natural language processing*, pages 194–201. Association for Computational Linguistics.

Andrew Borthwick, John Sterling, Eugene Agichtein, and Ralph Grishman. 1998. Exploiting diverse knowledge sources via maximum entropy in named entity recognition. In *Sixth Workshop on Very Large Corpora*.

María Granados Buey, Angel Luis Garrido, Carlos Bobed, and Sergio Ilarri. 2016. The ais project: Boosting information extraction from legal documents by using ontologies. In *ICAART (2)*, pages 438–445.

Ekaterina Buyko, Erik Faessler, Joachim Wermter, and Udo Hahn. 2009. Event extraction from trimmed dependency graphs. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, pages 19–27. Association for Computational Linguistics.

Cristian Cardellino, Milagro Teruel, Laura Alonso Alemany, and Serena Villata. 2017. Legal nerc with ontologies, wikipedia and curriculum learning. *EACL 2017*, page 254.

Nathanael Chambers, Shan Wang, and Dan Jurafsky. 2007. Classifying temporal relations between events. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 173–176. Association for Computational Linguistics.

Christopher Dozier, Ravi Kondadadi, Marc Light, Arun Vachher, Sriharsha Veeramachaneni, and Ramdev Wudali. 2009. Named entity recognition and resolution in legal text. In *Semantic Processing of Legal Texts*. Springer Verlag.

Jennifer D'Souza and Vincent Ng. 2013. Classifying temporal relations with rich linguistic knowledge. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 918–927.

Dan Klein and Christopher D Manning. 2002. Conditional structure versus conditional estimation in nlp models. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 9–16. Association for Computational Linguistics.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Nikolaos Lagos, Frederique Segond, Stefania Castellani, and Jacki ONeill. 2010. Event extraction for legal case building and reasoning. In *International Conference on Intelligent Information Processing*, pages 92–101. Springer.

Wendy Lehnert, J McCarthy, S Soderland, E Riloff, C Cardie, J Peterson, F Feng, C Dolan, and S Goldman. 1993. Umass/hughes: Description of the circus system used for tipster text. In *Proceedings of a workshop on held at Fredericksburg, Virginia: September 19-23, 1993*, pages 241–256. Association for Computational Linguistics.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *ACL (1)*, pages 73–82.

Shulin Liu, Yubo Chen, Shizhu He, Kang Liu, and Jun Zhao. 2016. Leveraging framenet to improve automatic event detection. In *ACL (1)*.

David McClosky, Mihai Surdeanu, and Christopher D Manning. 2011. Event extraction as dependency parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1626–1635. Association for Computational Linguistics.

Makoto Miwa, Sampo Pyysalo, Tadayoshi Hara, and Jun'ichi Tsujii. 2010. Evaluating dependency representation for event extraction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 779–787. Association for Computational Linguistics.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *HLT-NAACL*, pages 300–309.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *ACL (2)*, pages 365–371.

Ellen Riloff et al. 1993. Automatically constructing a dictionary for information extraction tasks. In *AAAI*, pages 811–816.