SENTENCE-LEVEL EVENT CLASSIFICATION IN UNSTRUCTURED CONTEXTS

Bekhbudieva Parvina Shokhrukhovna
Teacher of Samarkand state
institute of foreign languages
Lutfieva Amina Safoiddinovna
Master student of Samarkand state
institute of foreign languages

Annotation

In this article, we investigate the use of statistical methods for identifying the sentences in a document that describe one or more instance of a specified event type. We treat this task as a text classification problem where each sentence is either classified as one that contains an instance of the target event or one that does not. Event classification at a sentence level is a very challenging task. We estimate a series of uni-gram new models using three essential smoothing approaches, and investigate their overall behavior on classification performance.

Key words: Support Vector Machine method, probabilistic language modeling, On-Event Sentence, Off-Event Sentence, Natural Language Processing method.

КЛАССИФИКАЦИЯ СОБЫТИЙ НА УРОВНЕ ПРЕДЛОЖЕНИЯ В НЕСТРУКТУРИРОВАННЫХ КОНТЕКСТАХ

Бехбудиева Парвина Шохруховна
Учитель Самаркандского гос.
институт иностранных языков
Лутфиева Амина Сафоиддиновна
Магистрант Самаркандского гос.
институт иностранных языков

Аннотация

В этой статье мы исследуем использование статистических методов для выявления предложений в документе, описывающих один или несколько экземпляров определенного типа события. Мы рассматриваем эту задачу как проблему классификации текста, где каждое предложение классифицируется либо как содержащее экземпляр целевого события, либо как не содержащее его. Классификация событий на уровне предложений является очень сложной задачей. Мы оцениваем ряд новых моделей унограмм, используя три основных

подхода к сглаживанию, и исследуем их общее влияние на эффективность классификации.

Ключевые слова: метод опорных векторов, вероятностное языковое моделирование, предложение по событию, предложение без события, метод обработки естественного языка.

Classifying sentences that describes events is an important task for many natural language applications such as Question Answering and Text Summarisation. In this article, we treat event detection as a sentence level text classification problem.

Event detection is a core Natural Language Processing (NLP) task that focuses on the automatic identification and classification of various event types in text. This task has applications in automatic Question Answering (QA), Text Summarisation and more recently in the context of Semantic Web Retrieval. For example, event recognition is a core task in QA since the majority of web user questions have been found to relate to events and situations in the world

Three event detection approaches are explored in this article. Firstly, we train a Support Vector Machine (SVM) using a variety of term, lexical and additional event based features to encode each training or test instance. Secondly, we adopt a probabilistic language modeling approach that captures how descriptions of event instances in text are likely to be generated. One advantage of language modeling for text classification is that instead of explicitly pre-computing features and selecting a subset based on arbitrary decisions (as is often the case with standard classification learning approaches), the language modeling approach considers all terms occurring in the text as candidate features, and implicitly estimates the contribution of each feature in the final model. For example, if the target event is **Die**, we want our system to extract sentences such as “5 people were killed in the explosion” and “A young boy and his mother were found dead on Wednesday evening”. However, that classifier must also be able to detect complex cases such as: “An ambulance rushed the soldier to hospital, but efforts to save him failed” and reject instances such as “Fragmentation mines have a killing range of 100 feet”. A naıve system that selects only sentences that contain terms connected with death such as **‘kill’, ‘die’** or **‘execute’** as positive instances may catch many positive instances.

The aim of the third event detection system investigated in this article is to evaluate the effectiveness of such a shallow NLP approach by developing a manual rule-based system, which finds sentences connected to a target event type using a hand crafted list of ‘trigger’ terms. This system is compared with the SVM and uni-gram language models in order to investigate how the performance of such a manual approach compares against more sophisticated supervised techniques.

We investigated the task of determining whether a given sentence describes an instance of the target event as a binary text classification task where it is assigned one of the following classes:

• On-Event Sentence: a sentence that contains one or more instances of the target event type.

• Off-Event Sentence: a sentence that does not contain any instances of the target event type.

Sentence-level event extraction aims to identify events in a long text with pre-specified types and corresponding event-specific argument roles. Generating sentence-level events is beneficial for a variety of natural language processing downstream tasks, such as knowledge base construction, article summarization and question answering, since it can produce valuable structured information. However, the complex logic structures in long documents have made it a more challenging task than sentence-level event extraction that extracts the event from the sentence.

References:

1. Martina Naughton, Nicole Stokes. 2010, April. Sentence-level event classification in unstructured texts. Information Retrieval Journal.
2. Ahn, David. 2006. The stages of event extraction. In Proceedings of the ACL Workshop on Annotating and Reasoning about Time and Events, Pp 1–8, Sydney, Australia, July.
3. Allan, James, Jaime Carbonell, George Doddington, Jonathon Yamron, and Yiming Yang. 1998. Topic detection and tracking pilot study. final report.
4. Cohen, Jacob. 1960. A coeficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1): Pp37–46.
5. Curran, James, Stephen Clark, and Johan Bos. 2007. Linguistically motivated large-scale nlp with c & c and boxer. In Proceedings of the ACL 2007 Demonstrations Session (ACL-07 demo), pp 29–32.
6. Filatova, Elena and Vasileios Hatzivassiloglou. 2004. Eventbased extractive summarization. In In Proceedings of ACL Workshop on Summarization, pp 104 – 111.
7. Joachims, Thorsten. 1998. Text categorization with support vector machines: learning with many relevant features. In Nedellec, Claire and C ´ eline Rouveirol, editors, ´ Proceedings of the 10th ECML, pages 137–142,
8. Chemnitz, DE. Springer Verlag, Heidelberg, DE. Kraaij, Wessel and Martijn Spitters. 2003. Language models for topic tracking.