Named entity recognition using a new fuzzy support vector machine.

ABSTRACT

Recognizing and extracting exact name entities, like Persons, Locations, Organizations, Dates and Times are very useful to mining information from electronics resources and text. Learning to extract these types of data is called Named Entity Recognition(NER) task. Proper named entity recognition and extraction is important to solve most problems in hot research area such as Question Answering and Summarization Systems, Information Retrieval and Information Extraction, Machine Translation, Video Annotation, Semantic Web Search and Bioinformatics, especially Gene identification, proteins and DNAs names. Nowadays more researchers use three type of approaches namely, Rule-base NER, Machine Learning-base NER and Hybrid NER to identify names. Machine learning method is more famous and applicable than others, because it's more portable and domain independent. Some of the Machine learning algorithms used in NER methods are, support vector machine(SVM), Hidden Markov Model, Maximum Entropy Model (MEM) and Decision Tree. In this paper, we review these methods and compare them based on precision in recognition and also portability using the Message Understanding Conference(MUC) named entity definition and its standard data set to find their strength and weakness of each these methods. We have improved the precision in NER from text using the new proposed method that calls FSVM for NER. In our method we have employed Support Vector Machine as one of the best machine learning algorithm for classification and we contribute a new fuzzy membership function thus removing the Support Vector Machine's weakness points in NER precision and multi classification. The design of our method is a kind of One-Against-All multi classification technique to solve the traditional binary classifier in SVM.