RELATION EXTRACTION OF MEDICAL CORPORA AND SENTIMENT ANALYSIS

N.Ravinder, Kallam Bhargavi, Vattikuti Prathyusha, Kalluri Subramanyam

Assistant Professor, Department of Computer Science and Engineering

Koneru Lakshmaiah Educational Foundation,Vaddeswaram, Guntur-522502

Abstract : Information extraction and categorization is the key to comprehend corpus-based knowledge. Relations are defined as grammatical and semantic connections between entities. It becomes challenging to manually annotate the complete text in large structured corpora. However, automatic information extraction would improve decision-making process handier for expertise group of medical practitioners and for patients. Generally, most previous models used NLP (Natural Language Processing) tools and manually designed kernels. This paper proposes the usage of deep machine learning techniques namely conventional neural networks for capturing the relations. The system classifies the entity's relationships into one of the available six categories of relations through a group of defined features that indulges the assigning category and sentiment analysis for structured corpora. This model also provides a solution to composite semantic relation classification through sequence classification and distribution navigational algorithm.

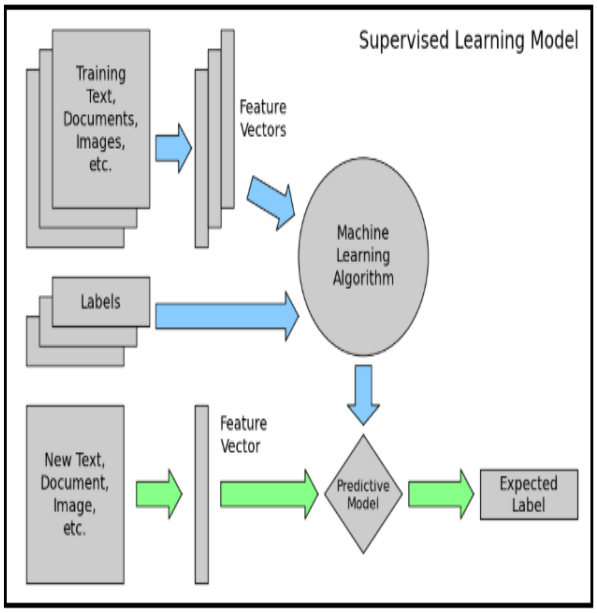
INTRODUCTION

Classification of semantic relationship between a pair of target entities in a text plays crucial role in Natural Language Processing (NLP), information extraction and question-answering. Encapsulation of semantic relation between terms acts as an elemental procedure for various semantic Interpretation tasks. Common-sense based knowledge understanding and extraction of relations from gigantic documents such as product reviews, business articles, management reports, technical manuals would pose challenges which may not be a problem in short texts. As large documents are complex, they can be either structured or semi-structured noisy with multiple formats. This complicates the sectioning of elements of a document. Most of classification model depend on high-level lexical and syntactical features extracted by NLP tools which uses art neural networks may suffer from implicit errors and are computationally expensive. However, classifying direct relationship between two terms is not always possible. In such cases semantic relation is considered as "OTHER" where it defines that tagged entities could not be mapped to any nine available relations.

Conventional neural networks (CNN)[4] helps in capturing the composite semantic relationships between entities. Firstly, the pair of targeted entities are tagged. The tagged pair of entities is the key for discrete classification of terms into six different semantic relations. Sentimental analysis of categorized relations has three categories of sentiments which are positive (1), neutral (0) and negative (-1).

EXSISTING SYSTEM

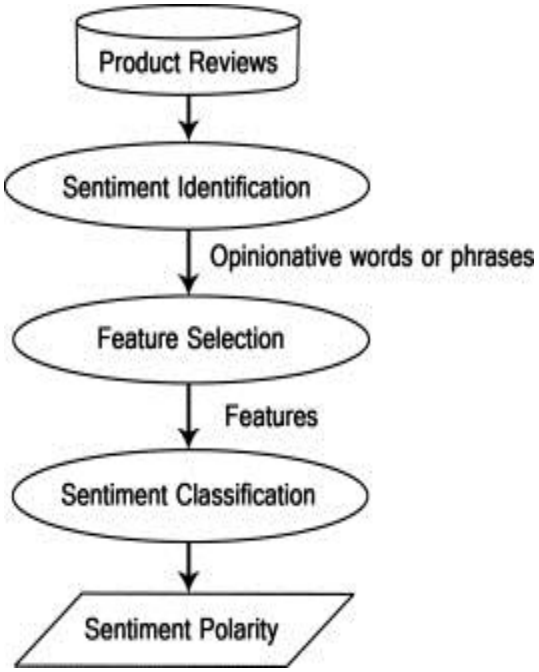
Most of existing systems for semantic relation classification involves unsupervised relation discovery and supervised relation classification. Various methods have been implemented recently for semantic relation classification using machine learning techniques[2] in completely different ways.



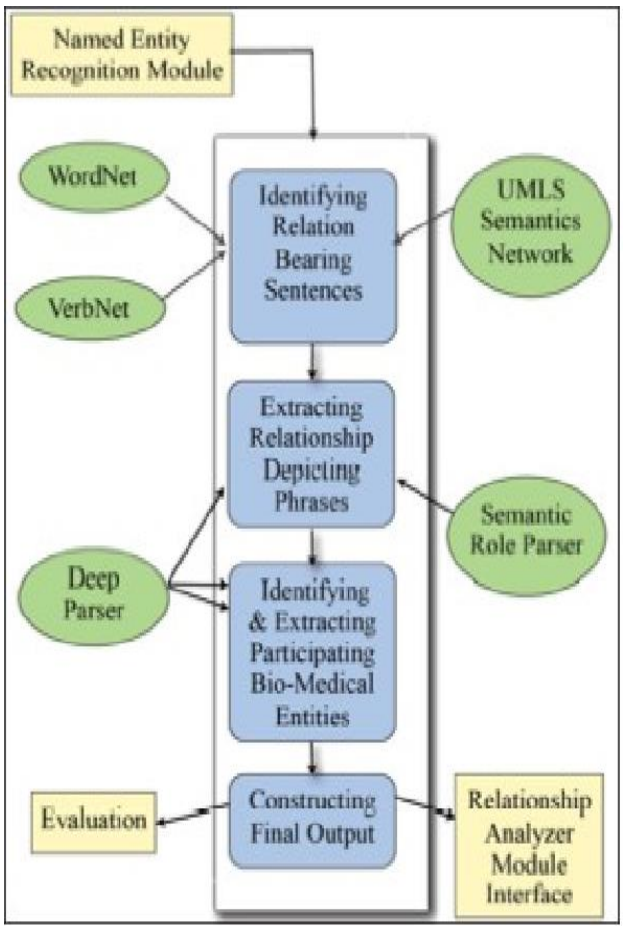
In unsupervised methods, defined features are captured using context features. A pair of words which occur in similar context are assumed to have same relations. Hierarchical clustering methods are used for context of nominal words which represents relation based on most frequent words in the context. Whereas in supervised

methods, approaches use either feature-based or manual kernel-based models.

For sentimental analysis, category features are used to calculate gravity score, affinity score and polarity score. WME 2.0 lexicons, wordnets[5], lexicons and POS (Parts-Of-Speech) resource to human language at higher-level serving as key foundation of high-cognitive models under Natural Language Processing [1]. Sentiments can have two values positive or negative. Where the conceptual knowledge indicates positive for strong relation which judges the condition of behavior of the context.



The document-level sentiment analysis aims sentiments based on the whole sentence which leads to the expressing opinion[6] of the complete document. The approches such as machine learning and lexicon model uses opinion mining[6] whre decision trees, linear classifiers,Rule-based classifiers are implemented.yet, these work efficienty only on short documents.Most recently used effective form of machine learning techique for classification are clustering methods[11].



Relation extraction system

SEMATIC RELATIONS

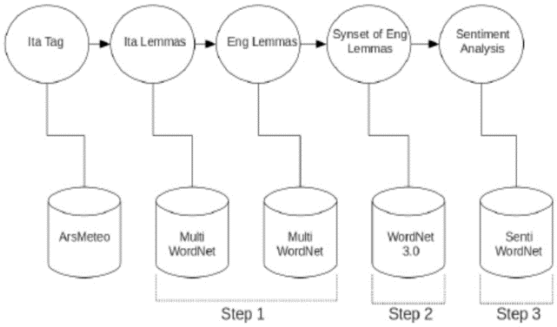
Semantic relations are associations which exist between the meanings at word-level or between the phrases. In many cases, the semantic relations are mapped between the nine available direct linguistic relationships. The relationship is named “others” if no direct relation can be mapped within the current knowledge path graph or semantic networks.[7]

Relation	Frequency
Cause-Effect	1331 (12.4%)
Component-Whole	1253 (11.7%)
Entity-Destination	1137 (10.6%)
Entity-Origin	974 (9.1%)
Product-Producer	948 (8.8%)
Member-Collection	923 (8.6%)
Message-Topic	895 (8.4%)
Content-Container	732 (6.8%)
Instrument-Agency	660 (6.2%)
Other	1864 (17.4%)

SENTIMENT ANALYSIS

For sentiment analysis document- level[10] classification there is a square measure of four completely different levels which are

- Sentence level
- Document level
- Phrase level
- Word level



The key task is to classify whether the text is subjective or objective. Objective expressions are considered as facts. Whereas, subjective are well-defined and specified opinions.

For example,

Subjective-

This fruit tastes good.

Objective-

This fruit is in red color.

Sentiment Analysis	Subjectivity Analysis
Positive	Subjective
Negative	
Neutral	Objective

CLSSIFICATION OF SENTIMENT ANALYSIS

PRAPOSED SYSTEM

The components of the speech is numerous indicator of sentiment expression and works on judgement detection that classify the relationship between presence of adjectives and sentence subjectivity. But several experimental results shown that victimization solely adjectives as options ends up in worse performance.

In the model, information is employed as a part of feature set that does sentiment classification on dataset.

The main aim of this work is to provide approach to semantic relation classification[8] using one/more relations between terms. Semantic classification is identifying the underlying abstract relation in sentence. Let us say the sentence S with annotated pair of target words e1 and e2 is provided. Then classification system aims at capturing relation between e1 and e2[7].

For example, Cause – Effect (e_1, e_2)

<e_1>Air pollution</e_1> is responsible for <e_2>global warming</e_2>.

Challenge in composite semantic relation extraction is ruled out by combination of sequence-based machined learning models, distributed semantic models and commonsense relational knowledge bases.

The model combines following techniques:

- Using structured commonsense knowledge base to define direct linguistic semantic relations.
- Uses pre-filtering method on distributional navigational algorithm.
- Using sequence-based neural networks namely conventional neural networks[4] to quantify sequence frequencies of semantic relations.

Modeling the entity relation extraction mainly concentrates on how different entities are related. This model uses multi-step process

Step – 1

Applies deep syntactic parsing techniques of sentence and identify the dependencies between terms.

Step – 2

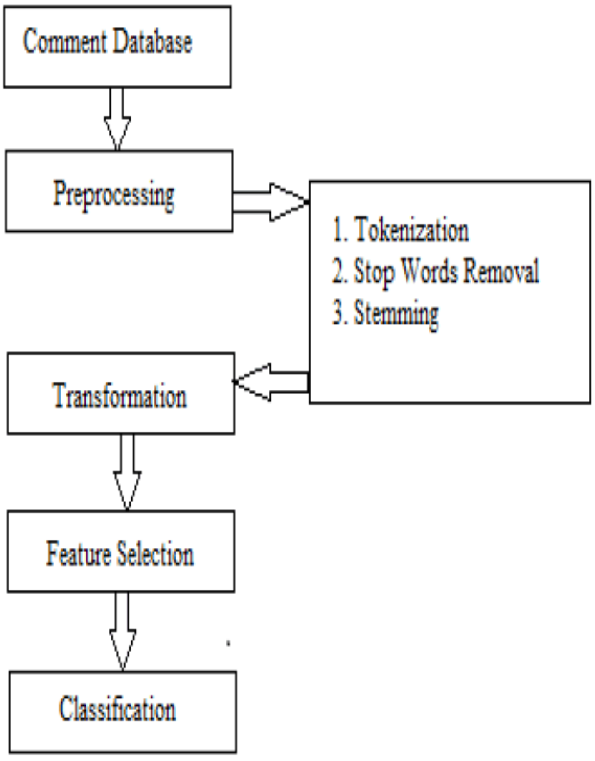
Performs entity extraction and entity linking for resolution of entities.

Step – 3

Semantic relationships are identified using deep convolutional neural networks.

Step – 4

Classification is done using patterns that are pre-defined in categories.



Steps and techniques used in classification.

Sentimental analysis is carried out by using deep neural networks[9] performing POS[3] on the categoriesed data.

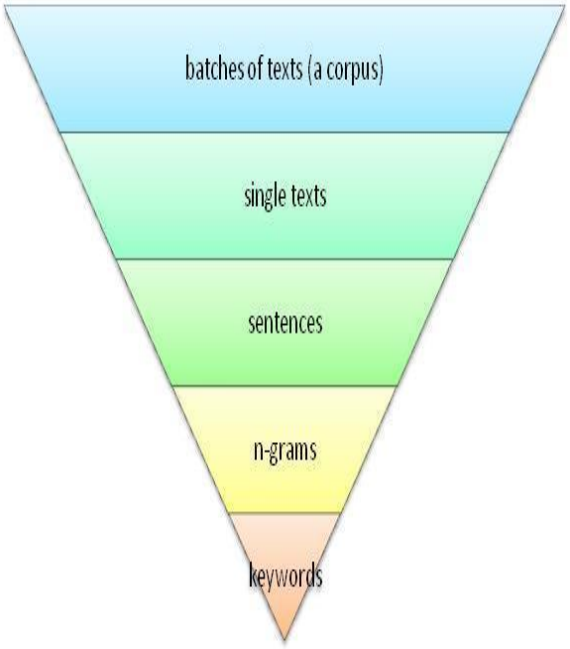
The text is parsed using an NLP tool kit.

PRP: Personal Pronoun

VBP: Verb, 3rd – person singular present

NN: Noun, singular or mass

JJ: Adjective, singular present



Based on the Parts of speech, words are assigned positive or negative score of sentiment using NLP toolkit and sentence sentiment calculator.

The sentiment is referred as follows

If score is between o and 1

POSITIVE

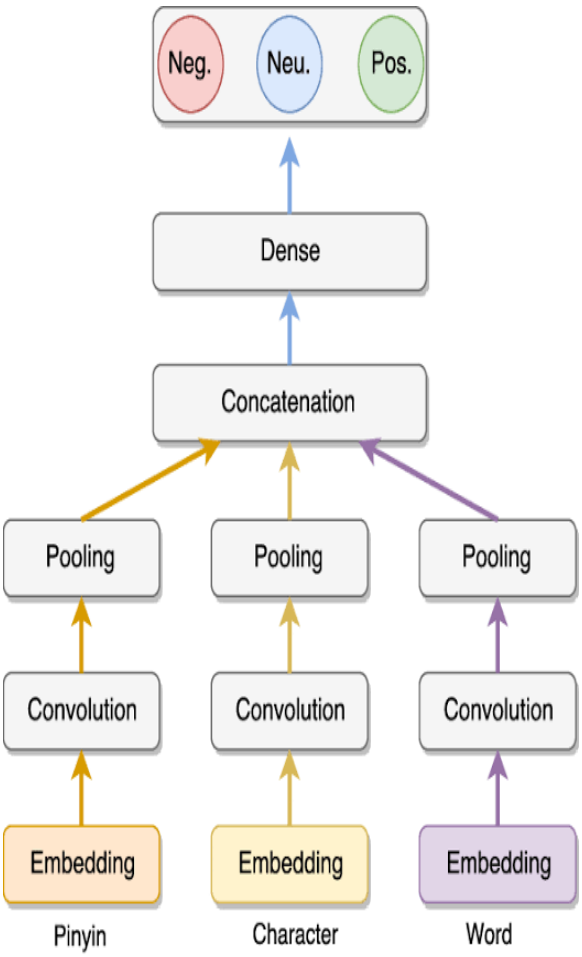
If score is between -1 and 0

NEGATIVE

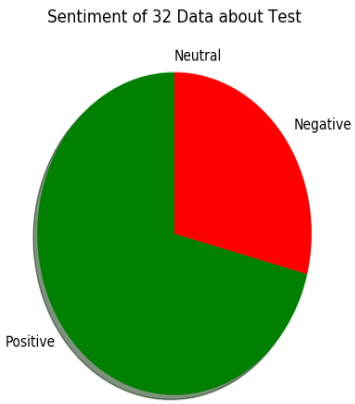
If score is exactly 0

NEUTRAL

Here, in this approach the proximity for neutral occurrence is low.



The result is illustrated in the pie chart which depicts the frequencies of the behavior of the corpus based on the categorization.



CONCLUSION

In healthcare system, clinical decision-making is essential for enhancement of medical treatment. This model is the key for improving the automatic clinical decision handling systems for both expertized doctors and patients.
As numerical decimal scores of sentiments are distinguished either negative or positive, the probability of obtaining neutral result has minimal chances which enriches the quality and accuracy of the system.

REFERENCES

[1] Relation Extraction of Medical Concepts Using Categorization and Sentiment AnalysisMondal, A., Cambria, E., Das, D. et al. Cogn Comput (2018) 10: 670. https://doi.org/10.1007/s12559-018-9567-8

[2] Using Machine Learning Techniques for Sentiment Analysis
https://ddd.uab.cat/pub/tfg/2017/tfg_70824/machine-learning-techniques.pdf

[3] A System for Semantic Relation Extraction and Classification https://www.aclweb.org/anthology/S18-1137.pdf

[4] Dual CNN for Relation Extraction with Knowledge-Based Attention and Word EmbeddingsComputational Intelligence and Neuroscience Volume 2019, Article ID 6789520, 10 pages https://doi.org/10.1155/2019/6789520

[5] Evaluating semantic relations in neural word embeddings with biomedical and general domain knowledge bases From The 2nd International Workshop on Semantics-Powered Data Analytics Kansas City, MO, USA. 13 November 2017

[6] A lexicon model for deep sentiment analysis and opinion mining applications Volume 53, Issue 4, November 2012, Pages 680-688

[7] Using Grammar Rule Clusters for Semantic Relation Classification.
https://www.aclweb.org/anthology/W11-0907.pdf

[8] A Unified Architecture for Semantic Role Labeling and Relation Classificationhttp://people.csail.mit.edu/jiang_guo/papers/coling2016-mtlsrc.pdf

[9] Sentiment Classification Using Negative and Intensive Sentiment Supplement Information Chen, X., Rao, Y., Xie, H. et al. Data Sci. Eng. (2019) 4: 109. https://doi.org/10.1007/s41019-019-0094-8

[10] V. S. Shirsat, R. S. Jagdale, S. N. Deshmukh, "Document Level Sentiment Analysis from News Articles", 2017 International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-4, 2017.

[11] Anitha, R., & Vimal Kumar, D. (2019). Enhanced slope one algorithm using hierarchical clustering and trust based collaborative filtering for ecommerce applications. *International Journal of Engineering and Advanced Technology*, *9*(1), 3006–3013.