ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 5, Issue 2, February 2018

Machine Learning Based Approaches for Natural Language Processing

[1] Himanshu Sharma, [2] Rohit Agrawal
[1][2] GLA University Mathura

***Abstract* - Machine Learning can play a vital role in many applications such as data mining, natural language processing, image recognition and expert systems. In the development of natural language system, the corpus based machine learning techniques are widely applied. In this paper, machine learning methods such as classifiers, structured models and unsupervised learning methods are discussed that are applied to natural language processing tasks such as document classification, disambiguation, parsing ,tagging, extraction etc. This paper also covers different levels of linguistic analysis: Lexical Analysis, Parsing, Semantic Analysis, Part-of –Speech Tagging and Discourse Knowledge. The aim of this is to provide valuable information for further research.**

Index Terms — Machine Learning, Corpus, Tagging, Parsing, Discourse.

I. INTRODUCTION

Machine Learning is science that provides computers the ability to learn without being openly programmed. We use machine learning many times a day without knowing it. It brings together computer science and statistics to exploit the predictive power. On the other hand Natural Language Pro-cessing is a artificial intelligence method to communicate with computers using natural language. Speech and Text worked as an input and output of an Natural Language Pro-cessing System. We can broadly divide NLP system into two parts as Natural Language Understanding (NLU) and Natural Language Generation (NLG).NLU refers transforming a given input into some useful representation and also analyzes different aspects of lnguage.NLG refers to generate meaning-ful sentence structures in some natural language from some internal representation.

Most of the approaches that exist for NLP are mainly focused on machine learning that is a type artificial intelligence that examines and apply on patterns in data to develop a program's own understanding. In the early 90s, the application of machine learning techniques for natural language learning problems has drawn attention of NLP researchers. As a human being, we understand our natural language easily and really don’t care about what actually understanding involves. Corpus-based language acquisition techniques that are used by NLP researchers are based on statistics and information theory. In this paper, we are discussing machine learning techniques that are applied to major NLP task such as POS Tagging, Parsing , Semantic analysis Word Sense Disambiguation, Text Categorization, Text Summarization and Information Extraction. Different linguistic levels that are applicable open to Machine Learning approaches are also covered in this paper; after that we explain the different types of machine learning approaches which are apply to natural language application tasks.

II. LEVELS OF LANGAUGE ANALYSIS

The following are different form of knowledge in relation with natural language understanding.

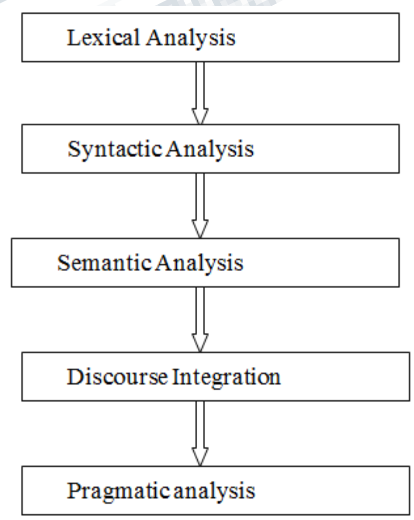

Fig.1 Linguistic levels

A. Lexical Analysis

A lexicon is a gathering of information about the words of a language and the lexical categories to which a word belongs. A lexicon is usually structured as a collection of lexical en-tries, like ("bank" N). Thus Lexical Analysis refers to identify and analyze structure of words. It is dividing whole text into paragraphs, sentences and words. A morphological analyzer identifies a word in a sentence

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

and calls it a token. The tokens identified are classified according to their use i.e. with identify it lexical category (grammatical class).".

B. Syntactic Parsing

The most important goal of parsing is to construct a parse tree for a sentence from a given grammar. We check whether a sentence is well formed or not. Thus we analyze words in a sentence and arrange words in a manner that shows relation-ship between these words in sentence. For example "Ramgoes to school" is valid sentence and "Ram school the goes" is rejected by a syntactic analyzer. Consider a grammar G as:

S → NP VP   NP → ART N   NP →ART ADJ N
VP →V   VP →V NP

Here S is a sentence, NP is a noun phrase, VP is verb phrase, ART is article, ADJ is adjective and V is verb.
Parsing can be broadly classified into two types:(1)Top-Down Parser and (2)Bottom Up Parser. In Top down parsing, the parsing starts with the start symbol S given in the gram-mar and applies the production rules so that it changes into a sequence of terminal symbols that matches the input sen-tence that is to be parsed. If it matches with the input sen-tence then the parsing is successful. If not, the process is started over again and we apply other production rules. The parsing is repeated till a particular rule is found which ex-plains the structure of the sentence. In Bottom up parsing method we start with the sentence and tries to reach at the start symbol S by replacing right hand side of a rule with it corresponding left hand side.

C. Semantic Analysis

The main purpose of semantic analysis is to generate the partial meaning of a sentence from its syntactic structure. The sentence is analyzed for its meaningfulness. There are various approaches of performing semantic analysis. First approach is sytnta driven approach. In this approach, we generate the meaning of a sentence from the meaning of its parts. Second approach is Semantic Grammar, In this approach we augment with domain specific semantics and mainly developed for dialogue systems.

D. Discourse Analysis

In real world, the meaning of a sentence is actually may depend upon the meaning of the previous sentence. This is known ad discourse knowledge. For example, interpreting pronouns and interpreting temporal aspects of the infor-mation. Consider a sentence "Ram hits a bike with a stone. It bounces back". Here "it" refers to stones. Discourse structure depends on application: Monologue, Dialogue or Human Computer Interaction. Thus it deals with the study of the relationship between language and its use context.

E. Pragmatic Analysis

In Pragmatic Analysis we have to analyze on how different situations affects use of an sentence and how this use affects the meaning of sentence. It deals with those aspects of sen-tence that require real world knowledge. It focuses on the communicative use of language realized as intentional human action. Thus it a practical usage of sentence: what a sentence means in practice.

III. DIFFICULTIES IN NATURAL LANGUAGE UNDERSTANDING

A. Lexical Ambiguity

A word may be a noun or verb. For example word "bank" can be noun as well as verb. This is the example of lexical ambiguity.

B. Syntactic Ambiguity

A sentence can have more than one parse trees thus this sentence is ambiguous in nature. For example, "Tell all trains on Monday." –This is an example of syntactic ambiguity. We can have two interpretations of the same sentence as shown in the fig.

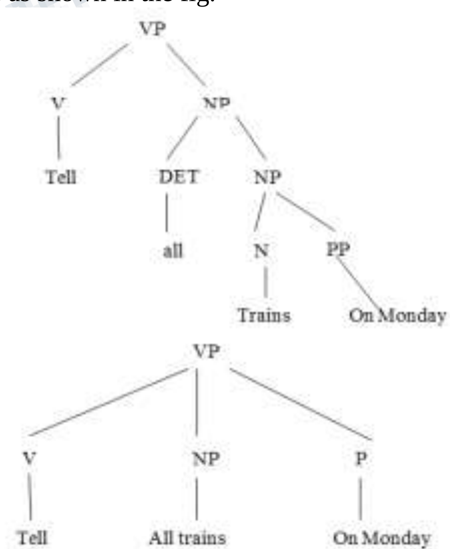

Fig.2 Two structural representations of Tell all trains on Mon-day.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

C. Referential Ambiguity

This type of ambiguity means when a word or a phrase re-fers to two or more properties or things. For example use of pronouns creates a referential ambiguity. Consider the sen-tence:John killed the girl with the knife.Here one interpreta-tion can be"kill the girl having knife" and other can be "kill the girl by knife".

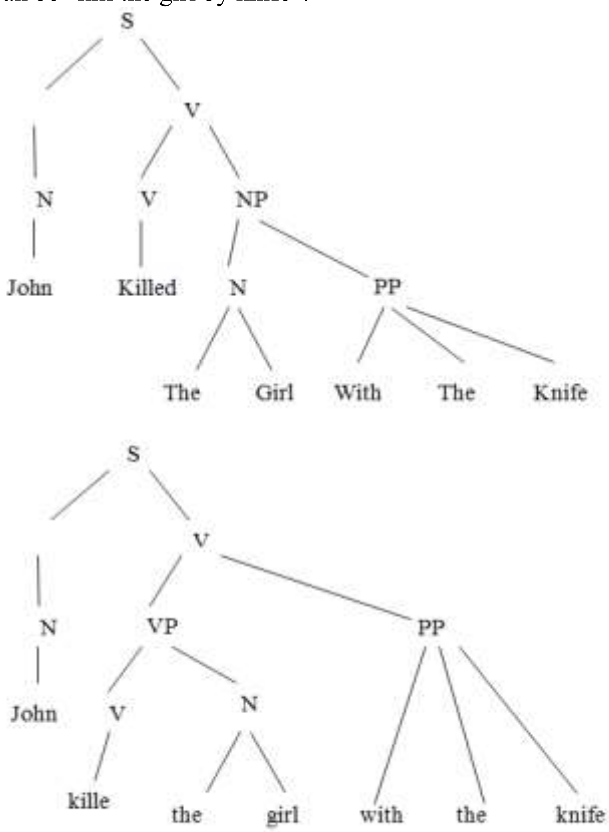

Fig.3 Two interpretations of John killed the girl with the knife

IV. MACHINE LEARNING BASED APPROACHES FOR NATURAL LANGUAGE PROCESSING

A. Naive Bayes Classifier

Naive Bayes is a simple and powerful technique for predic-tive modeling. In this approach, training is very fast because only the probability of each class and the probability of each class when given different input (x) values required to be calculated. We need not to fit any coefficients by optimiza-tion procedures. It is widely used by machine learning and NLP researchers with a great success. This algorithm is ap-plies to many NLP disambiguation tasks such as Part-of-Speech tagging, Word Sense disambiguation and Text Cate-gorization etc. We can do sentiment analysis of tweet on twitter. We will collect the tweets from twitter using Twitter streaming API. The collected tweets can be preprocessed using Natural Language Toolkit methods. Then we will select the features of the tweets based on Chi square and Naïve Bayes classifier is used to classify the tweets as positive and-negative.

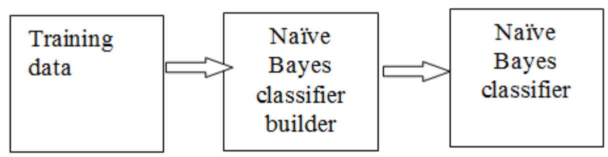

Fig.4 Naïve Bayes classifier builder

B. N-gram and Hidden Markov Model

N-gram model uses the previous N-1 words in a sequence to predict n-th word. Hence we tried to approximate the proba-bility of each word in terms of its context. Suppose a lan-guage has N word types in its lexicon, how likely is word a to follow word b? For simplicity we can have following models:

In Unigram model: Prob (pen)
In Bigram model: Prob (pen | black)
In Trigram model: Prob (pen | your black)

The Markov model/chain is the guesses that the future behavior of a system only depends on limited/narrow history. Thus in Markov Model with ith-order, the next state only depends on the i latest states, hence we conclude that an N-gram model is a (N-1)-order Markov model.

Hidden Markov Models (HMMs) are variations of Markov Models. In HMMs we consider two layers of states: a visible layer that represents to input symbols and a hidden layer which is learnt by the system, describing tobroader categories. HMMs are widely used in language disambiguation tasks such as POS tagging, names entity recognition and classification and extensions of HMMs can also be used for word sense disambiguation tasks. Consider the sentence "flies like a sand". Here flies can be noun(N),verb(V) and so on.We will calculate the probability of flies as a verb, as a noun and so on. Similarly we will calculate the probability of like as verb, noun and so on. Depending upon probability outcome we will provide most appropriate part of speech to each word in the sentence.

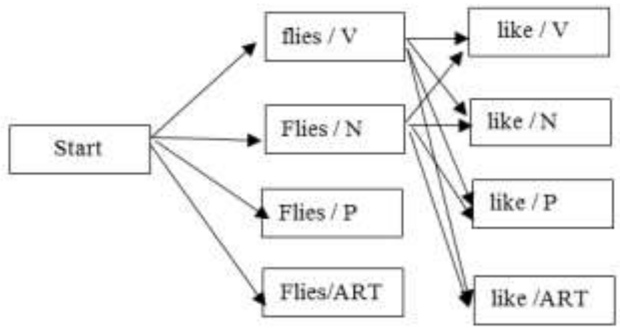

Fig.5 Encoding the possible sequences using the Markov Assumptions

C. Log-Linear Models

Log Linear models have wide applications in NLP classifica-tion tasks. Log Linear regression that is mostly used for bina-ry classification is also used to classify verbs for machine translation purposes. This model is also used for POS tagging.

We have some input domain X, and a finite label set Y. Our aimis to provide a conditional probability P(y / x) for any x ε Xand y ε Y.

A feature is a function f : X x Y →R.Say we have m

features ϕ_k for = 1……m.We also have a parameter

vector W ε R^m.We define

$$P(y / x,W)= e^{W.\phi(x,y)} / \Sigma_{y' \varepsilon Y} e^{W.\phi(x,y')}$$

D. Transformation Based Learning

Transformation based learning is used widely for corpus based natural language learning. This algorithm is based on mistake driven greedy approach that generates set of rules. We iteratively add a rule that best repairs current errors. This algorithm is widely used in many natural language problems such as POS tagging, Word Sense Disambiguation and Pars-ing. Samuel, Carberry and Vijay-Shankar describes a Monte Carlo version of the Transformation based learning algorithm that focuses on the large-scale search problem of selecting restricts from all possible combinations of a pre-selected set of conditions. Every condition is made up of a feature and a distance. The feature refers a property of utterance that may be important for the Dialogue Act Tagging applications, and the distance point to the relative position of the utterance that the feature should be applied to.

E. Decision Trees

Decision Trees are widely used for Classification Problems. They are part of the family of machine learning methods and allow a hierarchical distribution of dataset collection. Using a decision tree algorithm is generated knowledge in the form of a hierarchical tree structure that can be used to classify instances based on a training dataset. An example of the decision tree used in natural language processing is the syntactic tree generated with a grammar. Parts-of-speech are very important in morphology because they can give us a large amount of information about a word and its neighbours and the way the word is pronounced. So, the problem of assigning parts-of-speech to words (part-of-speech tagging) is very important in speech and language processing. Decision trees can also be used in pragmatics. Interpreting dialog act assume that the system must decide whether a given input is a statement, a question, a directive, or an acknowledgement.

We also use decision tree to solve ambiguity problems in natural language at different levels such as text summariza-tion, text categorization and word sense disambiguation.

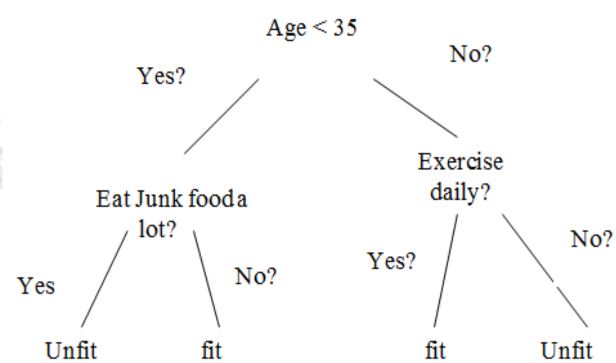

Fig.6 Decision tree to find whether a person is fit or not.

Suppose we like to predict whether a person is fit with the information like age, eating habit, and physical activity, etc given to us. The decision / internal nodes here are questions such as ‘His/Her age?’, ‘Exercise daily or not?’, ‘Does he eat a lot of junk food’? And the leaves represent the outcomes like either ‘fit’, or ‘unfit’. In this example it is a binary classi-fication problem (yes/ no type problem).

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

F .Clustering Algorithms

These algorithms are unsupervised machine learning algo-rithms examples and are used to in various NLP tasks such as semantic classification, syntactic classification, document retrieval and machine translation. The modified versions of these algorithms are used to handle noun and pronoun phrase co-reference resolutions in information extraction.We can include clustering techniques calculated by various means of classification, machine learning, tagging, stemming, and parsing. This will also improve the probability measure of the whole sentence by looking ahead of traditional n-grammodels entirely and developing measures of key, linked pairs of words in the sentence.

G. Support Vector Machines

Support Vector Machines (SVMs) work on principle Struc-tural Risk Minimization. SVMs are widely and efficiently used in Pattern Recognition problems. In the field of Natural Language Processing, SVMs are used in Text Categorization tasks. The SVM based Recursive Feature Elimination algorithm is a important method for feature selection and extraction, used in natural language processing. Support Vector Machine method give good performance on the chunking tasks.

H. Neural Networks

Neural Networks are used in Natural Language Processing filed in many problems such as Speech recognition and syn-thesis, Optical Character Recognition, Part-of-Speech Tag-ging Parsing, sentence analysis, PP-attachment disambigua-tion and text categorization.

In NLP field, words and their nearby contexts are quite im-portant: a word bounded by related context is important, while a word surrounded by unrelated context is not very important. Every word is mapped to a vector that is described in terms of its features (which in turn relate to the word's related context), and thus neural networks concepts can be used to learn which features is important and that maximize a word vector's score.

A very close model to this is the semantic network. In Se-mantic networks we have nodes that represent concepts or logics and connections that represent semantically meaning-ful relationship between these concepts or logics. These net-works are mainly described as associative network models than as neural/brain models. The activation rules that imple-ment data retrieval in these associative networks, often re-ferred to as spreading activation, particularly produces a joint search. Therefore, they are also called "spreadingactivation" models.

I. Genetic Algorithms

Genetic Algorithms are used in Dialogue systems. These systems are computer based systems that are used to com-municate with human in some natural language in form of text or speech. Also Genetic algorithms are used in Language Generation. Language generation is a method that aims to create natu-rallanguage from a data representation such as a knowledge base. These Language generators are often used to create textual form of natural language. Genetic Algorithms are also used in story generation. We can also use these algorithms in recognizing lexical infer-ence - given two terms a and b, predicting whether the mean-ing of b can be inferred from a.

J. Instance Based Learning

IBL methods for modeling real-valued or discrete valued predicted functionsIBL methods initially store the presented training data. These algorithms are widely used in many areas of Artificial Intelligence. These algorithms are used in many NLP problems such as information extraction, lexical, semantic disambiguation of complete sentences, chunking, context sensitive parsing, text categorization and semantic interpretation. Instance based Learning with automatic feature selection is a widely used approach in the Word Sense Disambiguation field.

V. CONCLUSION

In this paper, many machine learning methods are described that are used in natural language processing field. Many NLP tasks such as POS tagging, Syntactic Parsing, Semantic Analysis, Word Sense Disambiguation, Text Categorization, Text Summarization, Information Extraction, Language Generation, Dialogue based Systems are addresses in terms of machine learning approaches. This paper has explored all the main methods used in different linguistic levels. In future work, disambiguation problems must be addressed in detail based on the given Machine Learning Methods.

REFERENCES

[1] Leech, G, Garside, R & Atwell, E, 1983. The Automatic Grammatical Tagging of the LOB CorpusICAME Journal of the International Computer Archive of Modern English Vol.7.

[2] Samuel, K, Carberry, S, and Vijay-Shanker, K, 1998. Dialogue act tagging with transformation-based

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

learning. In Proceedings of COLING/ACL-98, Montreal, volume 2, pp.1150-1156

[3] Brill, E. 1995. Transformation-based error-driven learning and natural language processing: a casestudy in Part-of-Speech tagging. Computational Linguistics, volume 21(4), pages 543-566.

[4] OpenNLP, http://opennlp.sourceforge.net

[5] E.Charnaik,Statistical Language Learning. The MITPress,Cambridge,Massachusetts,1993.

[6] Bishop C. 2006. Pattern Recognition and Machine Learning, Springer