WORD SENSE DISAMBIGUATION FOR MALAYALAM IN A CONDITIONAL RANDOM FIELD FRAMEWORK

Junaida M K
IT Education Centre
Thalassery Campus
Kannur University
junaidashukoor@gmail.com

Jisha P Jayan
Centre for Development of Imaging Technology
Thiruvanathapuram
jishapjayan@gmail.com

Elizabeth Sherly
Indian Institute of Information Technology and Management-Kerala
Thiruvanathapuram
sherly@iiitmk.ac.in

Abstract

Word Sense Disambiguation (WSD) or Lexical Ambiguity Resolution is one of the pressing problems in Natural Language Processing (NLP), which identifies the correct sense of an ambiguous word in the specific context in a given sentence. WSD is considered as a harder problem as it depends on a set of classes, which vary depending on the context. This paper describes two algorithms Conditional Random Field (CRF) and Margin Infused Relaxed (MIRA) in a CRF framework for Malayalam WSD. This framework makes use of the contextual feature information along with the parts of speech tag feature in order to predict the various WSD classes. For training set, number of ambiguous words has been annotated with 25 WSD classes. The experimental results of the 10 fold cross validation shows the appropriateness of the proposed CRF based Malayalam word sense tagger.

1 Introduction

Word Sense Disambiguation is an intermediate task which is necessary at one level or another to accomplish in most natural language processing tasks. The development of an automatic Word Sense Disambiguation requires either a sense inventory, usually obtained from a dictionary, thesaurus or a large annotated corpus. The significant amount of information about the word and its neighbours of a particular word in a context gives a sense of a particular word which can be useful in a language model for different speech and text processing applications. In English ’line’ (cord, division, formation, phone, product, text), ’hard’ (difficult to achieve, intense, intense, for surfaces, things), interest (stake, involvement, interestingness, pastime, the thing that is important, charge for borrowing money) are the words with multiple meanings and such words are called polysemy.

WSD is a major subtask of Machine Translation, have relevant significance in almost every application of language technology, including information retrieval, lexicography, knowledge mining/acquisition and semantic interpretation, and is becoming increasingly important in new research fields such as the cognitive Computing, semantic web, bioinformatics etc.

Automatic WSD systems are available for many languages like English, Spanish, Chinese and some Indian languages. Malayalam being an unstructured language, faces a severe problem in the work on automatic WSD. Malayalam is an agglutinating language that exhibits very rich system of morphology and many senses, which is challenging. The basic components required for developing good WSD is the availability of Malayalam dictionary/thesaurus and labelled text corpus. For example,Consider the following sentence from Malayalam: ഞാൻ അവന്റെ കരം പിടിച്ചു (njan avanTe karaM piTiccu) with the meaning I took his hand and രാമു അവന്റെ വീടിന്റെ കരം അടച്ചു (raamu avanTe veeTinTe karaM aTaccu) with the meaning ramu paid his tax.The word കരം (karaM) have different meanings Tax or Hand. Here the distinction of the sense of the word കരം (karaM) is complex due to the lack of capitalization information and free word order of the language.

This paper is organized into different sections. First section dealt with the introduction part. The second section explains the major works carried out in this area. The third and fourth sections describe the complexity of Malayalam language and the Machine learning approach using CRF framework based on two algorithms CRF and MIRA. The next two sections explain the proposed work and implementation. The seventh section includes the experimental results obtained. The eighth sec-

S Bandyopadhyay, D S Sharma and R Sangal. Proc. of the 14th Intl. Conference on Natural Language Processing, pages 495–502, Kolkata, India. December 2017.

tion concludes the paper with future works that can be done as an outcome of this work.

2 Related Works

There are many approaches used for identifying WSD. The two main approaches are Dictionary based and Corpus based. The Dictionary-based method, uses external knowledge resources, which define explicit sense distinctions for assigning the correct sense of a word in context. In corpus-based methods, machine-learning techniques are used to induce models of word usages from large collections of text examples. Both knowledge-based and corpus -based methods have their own benefits and drawbacks. The former approach mainly uses external lexical resources like dictionaries, thesaurus, WordNet etc. They are easy to implement as it requires only simple look up of knowledge resources like machine readable dictionary. The corpus based methods use techniques from statistics and machine learning to induce language models. Learning can be done with supervised or unsupervised methods, which learns sense classifiers from annotated data with minimal or partial human supervision respectively. Many standard machine learning techniques have been applied, including Naive Bayes (NB), Maximum Entropy (ME), Exemplar-based (kNN), Decision Lists (DL), Support Vector Machines (SVM) etc. Naive Bayes algorithm is one of the simplest algorithm, which uses Bayes rule and given the class labels conditional independence of the features are assumed. It has been applied to many experiments in Natural Language Processing as well as WSD with considerable success (Yuret, 2004).

The information theory in particular,The Maximum Entropy approach provides a flexible way to combine statistical evidences from many sources. It has been applied to many NLP problems and also appears as alternative in WSD (Suarez, 2002). Chatterjee (2009) presented a trainable model applies the information theory for Word Sense Disambiguation (WSD) for resolving the ambiguity of English words. Decision Lists were used for lexical ambiguity resolution in Spanish and French accent restoration (Yarowsky, 1994) and used in other work for WSD (Yarowsky, 1995). Parameswarappa (2011) described the machine learning techniques with naive bayes classifier for Kannada target word sense disambiguation using compound words clue and syntactic features in a local context.

Lesk (1986) was one of the first researchers who tried to disambiguate Machine Readable Dictionaries (MRD) using Simplified Lesk algorithms. His algorithm became well-known among WSD researchers. His algorithm was primarily an overlap based algorithm which suffers from overlap scarcity. These methods, highly rely on lexical resources such as machine readable dictionaries, thesaurus etc. For English, this method achieved 50-70% accuracy in correctly disambiguating the words. The work (Sinhar, 2004) mainly focused on Hindi. They used contextual overlap between sentential context and extended sense definitions from Hindi Word Net. Sense bag was created by extracting words from synonyms, glosses, example sentences, hyponyms, and glosses of hyponyms, example sentences of hyponyms, hypernyms, and glosses of hypernyms, example sentences of hypernyms, meronyms, glosses of meronyms, and example sentences of meronyms. A context bag was created by extracting words in the neighborhood i.e. one sentence before and after, of the polysemous word to be disambiguated. The sense which maximized the overlap was assigned as winner sense. By using word co- occurrences of the gloss and the context (Gaona, 2009) presented a measure for sense assignment useful for the simple Lesk algorithm. Based on domain information and WordNet hierarchy (Kolte, 2009) proposed unsupervised approach to WSD. The words in the sentence contribute to determine the domain of the sentence. The availability of WordNet domains makes the domain-oriented text analysis possible. The domain of the target word can be fixed based on the domains of the content words in the local context. This approach can be effectively used to disambiguate nouns.

In Malayalam only a few works have been published. A knowledge based approach to Malayalam WSD (Haroon, 2010) has been done. It is based on a hand devised knowledge source and uses the Lesks and Walkers algorithm and also using the concept of conceptual density with Malayalam WordNet as the Lexical resource. The knowledge based system will result in poor accuracies because of the dependency of the algorithm on the stored tag words within the knowledge source.

3 Complexity of Malayalam

This section introduces the linguistic preliminaries of Malayalam language and complexities involved in the Malayalam Word Sense Disambiguation. The world languages are classified into fixed word order and free word order. In fixed word order the words constituting a sentence can be positioned in a sentence according to grammatical rules in some standard ways. On the other hand, in the free word order no fixed ordering is imposed on the sequence of words in a sentence. The English language is example of fixed word order language and Sanskrit is pure free word order language. Generally Malayalam is a free word order language and agglutinating language and exhibits very rich system of morphology. Morphology includes inflection, conflation (sandhi), and derivation. Word Sense Disambiguation is a difficult task in Natural Language Processing, In addition to the difficulties involved in Word Sense Disambiguation, the complexity level is even more in unstructured language like Malayalam. Here we will briefly describe the complexities involved in our work. For example, consider a sentence അവൻ നടന്നു With the meaning He walk and യോഗം നടന്നു With the meaning Meeting executed. Here the distinction of the sense of the word നട is very complex due to the lacks of capitalization information and free word order of the language. Applying stochastic models to the WSD problem requires large amounts of annotated data in order to achieve reasonable performance. Stochastic models have been applied successfully to English, German and other European languages for which large sets of labeled data are available. The problem remains difficult for Indian languages (ILs) due to the lack of such large annotated corpora. This is due to the fact that many different encoding standards are being used. Also, the number of Malayalam documents are available in the web is comparatively quite limited. Malayalam word sense disambiguation is of interest due to a number of applications like machine translation, text summarization, information retrieval.

To begin with, this experiment requires a sense tagged corpus in -order to achieve considerable accuracy for disambiguation. Developing corpus is a tedious and very time consuming task. The next issue involved in this work is the unavailability of sense inventory which will decide appropriate senses to the specific word in a context. The most appropriate meaning of a word is selected from a predefined set of possibilities, usually known as sense inventories. An efficient POS tagger in Malayalam is required to extract Word Sense Disambiguation, which also requires large corpus for training.

4 Machine Learning Using CRF

Statistical methods work by employing a probabilistic model containing features of the data. Features of the data, that can be understood as rules set for the probabilistic model, are created by learning the resulting corpora with properly marked tags. The probabilistic model then uses the features to calculate and determine the foremost probable tags. As such, if the annotated features of the data are correct and reliable, the model would have a high likelihood to find almost all the tags within a text.

CRF has found its application in many domains that may deal with the structured data. They are considered to be state of the art techniques for many applications in NLP. CRFs are a probabilistic framework (Wallachi, 2004) that is used for labeling and segmenting structured data, such as sequences, trees and lattices. CRFs bring together the best of generative and classification models. These are mainly undirected graphical models (Zhang, 2013). The underlying idea is that of defining a probability distribution which is conditional over label sequences given a particular observation sequence, rather than a joint distribution over both label and observation sequences. A key advantage of CRFs is their great flexibility to include a wide variety of arbitrary, non - independent features of the input (McCallum, 2002). The primary advantage of CRFs over HMMs is their conditional nature, which result in the relaxation of independent assumptions required by HMMs in order to ensure tractable inference. Additionally, CRFs avoid the label bias problem (Lafferty, 2001).

Margin Infused Relaxed Algorithm (MIRA) is a machine learning algorithm for multi-class classification problems. It has been introduced (Crammer, 2003). It learns set of parameters (vector or matrix) by processing all the given training examples one at a time, according to each training example parameters are updated. So that the current training example is classified correctly with a margin against incorrect classifications at least as large as their loss. The change of the parameters is kept as small as possible. A two-class version called binary MIRA is not requiring the solution of a quadratic programming problem, so it is simple. Binary MIRA can be used in an onevs-all configuration, it can be extended to a multiclass learner that approximates full MIRA, but may be faster to train. The flow of the algorithm looks as follows:

Algorithm MIRA

Input: Training examples $T = \{x_i, y_i\}$

Output: Set of parameters w

i ? 0, $w^{(0)}$? 0

for n ? 1 to N

  for t ? 1 to $|T|$

    $w^{(i+1)}$? update $w^{(i)}$ according to $\{x_t, y_t\}$

    i ? $i + 1$

  end for

end for

return $\frac{\sum_{j=1}^{N \times |T|} w^{(j)}}{N \times |T|}$

Figure 1: Algorithm MIRA

In the present work, we propose a machine approach using two different algorithms namely CRF and MIRA of Conditional Random Field framework for unrestricted Malayalam text WSD. The main steps involved are corpus collection, preprocessing, tagging, training and analysis. The template for training the CRF engine is defined. A lot of work is being done in the fields of corpus building, creating an efficient POS tagger and subject identification in Malayalam language.

5 Implementation

For WSD implementation, is used CRF++ and the experiment was carried out on different Malayalam ambiguous words. The template file contains the features specified for training and testing. The template file has multiple lines, each corresponds to a particular composite feature. It helps in generating n-gram features from the feature columns. The variables U and B are used to represent the features which denotes uni-gram and bi- gram respectively. The template line that starts with U predicts the current label generating n weights for n different labels. The template line which starts with B defines the current and previous labels generating n*n weights in the model. The composite feature is expressed by %x[i,j] with respect to the current labels. The template for CRF is defined as follows:

```
# Unigram
Unigram U00:%x[-2,0]
U01:%x[-1,0]
U02:%x[0,0]
U03:%x[1,0]
U04:%x[2,0]
U05:%x[-1,0]/%x[0,0]
U06:%x[0,0]/%x[1,0]
U10:%x[-2,1]
U11:%x[-1,1]
U12:%x[0,1]
U13:%x[1,1]
U14:%x[2,1]
U15:%x[-2,1]/%x[-1,1]
U16:%x[-1,1]/%x[0,1]
U17:%x[0,1]/%x[1,1]
U18:%x[1,1]/%x[2,1]
U20:%x[-2,1]/%x[-1,1]/%x[0,1]
U21:%x[-1,1]/%x[0,1]/%x[1,1]
U22:%x[0,1]/%x[1,1]/%x[2,1]
# Bigram
B
```

In order to accommodate common words and senses, we have used manually collected sentence from various Malayalam newspapers, Wikipedia articles, blogs, books, novels etc. Table 1 shows these words and sense.

In order to avoid inconsistencies present in spelling, spacing and punctuation, preprocessing is done by thoroughly checking the database. Then manual tagging of polysemous words and parts of speech tagging were carried out.

Feature selection plays an important role in machine learning. The experiments have been carried out using the basic context information and Parts of speech tag combination of word and tag context. The features are binary valued functions which associate a tag with various elements of the context. The experiments used two groups of features: word and word + part-of-speech bigrams. Following are the details of the features that have been applied to WSD task. Word features are lexical features, unique words that occur in the training set in a specific window range. Word feature contains the following attributes. w-2, w-1, w, w+1, w+2, (w- 2,w-1,w), (w-1,w,w+1), (w,w+1,w+2) , where the last three correspond to collocations of three consecutive words. Word + POS features are lexico-syntactic features combining POS information in a predefined range of

Word	Senses (classes)
രസം (rasaM)	താൽപര്യം, കറി, ഇഷ്ടം, രുചി, മെർക്കുറി (taal˜paryaM, kaRi, ishTaM, ruci, mer˜kkuRi)
നട (naTa)	ക്രിയ, നടക്കുക , പടി (kRIya , naTakkuka , paTi)
അടി (aTi)	ചുവടളവ്, പാദം, തല്ല് (cuvaTaLav , paadaM , tall)
വാനം (vaanaM)	മാനം, അഭിമാനം (maanaM , abhimaanaM)
ഉത്തരം (uttaraM)	മറുപടി, ചോദ്യോത്തരം, താങ്ങ് (maRupaTi, coodyoottaraM, taangng)

Table 1: Ambiguous Words and Senses (Classes)

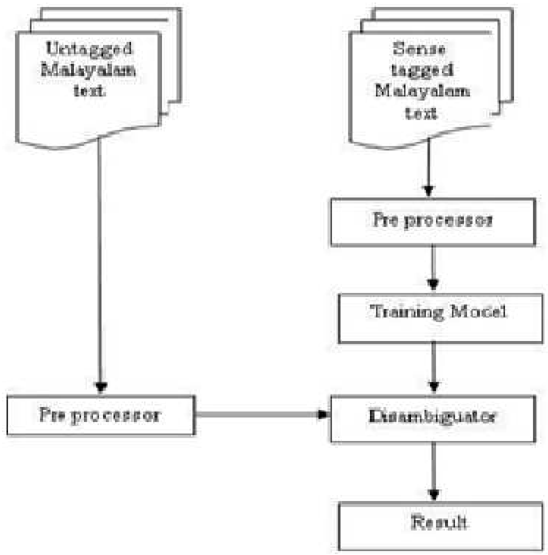

Figure 2: Block Diagram

the particular word. Word + POS feature contains w-2,w-1, w, w+1, w+2, with parts of speech information p, (w,p), (p-1,p). The tagging was performed using BIS tagset. Four taggers have been implemented based on the CRF and MIRA model. The first tagger (Word) makes use of the simple contextual feature, whereas the second tagger (Word+POS) uses parts of speech information features along with the simple contextual features. Each tagger is trained and tested using both the models, CRF based stochastic tagging scheme and MIRA. The same training corpus has been used to estimate the parameters for all the models.

6 Experimental Results

For the evaluation of this experiment we have used n-fold crossvalidation method due to the lack of huge amount of corpus.Usually data is split in to 70% for training and 30% for testing or in some cases 80% for training and 20% for testing. Although this distribution is commonly used for large datasets, it presents a challenge for smaller datasets and it might lead to problem of representativeness of the training or testing data. For these experiments, the method of n-fold cross validation is used divided in ten sets, each set containing 10% of the total data, therefore a ten-fold cross validation. The 10-fold was chosen mainly because the amount of data used for the experiments is not considered to be big as in most other applications. Because of that, fewer partitions were employed in order to ensure that a reasonable number of instances are included in each partition.Therefore it is necessary to ensure that random sampling is done in a way that guarantees that each class in the data set is properly represented in both the training and test sets.

The evaluation of the CRF and MIRA based models has been done using evaluation matrices. We have implemented two CRF and MIRA based models using Word feature and Word + POS feature.The classification is performed for skewed and highly imbalanced data, accuracy is very high and it does not reflect exactly the performance of the classifier. For this reason, precision (P), recall (R) and F-measure (F) scores are reported, which shows how precise and complete the classification is on the positive class. The TP, TN, FP and FN refer to true positives, true negatives, false positives and false negatives respectively. In a binary class based classification context, the terms positive and negative used in these definitions are associated with membership to one of the two semantic classes involved inc the classification (senses).

```
ഉരുക്കിന്റെ N_NN      NULL
ആവിർഭാവത്തോടെ V_VM_VNF          NULL
വലിയ    JJ       NULL
വലിയ    JJ       NULL
ഉത്തരങ്ങൾ          N_NN     താങ്ങ്
നിഛ്പ്രയാസം         RB      NULL
നിർമിക്കാമെന്നായി V_VM_VNF          NULL
.          RD_PUNC NULL
ചോദ്യം   N_NN     NULL
വായിച്ച്   V_VM_VNF           NULL
കണ്ണിൽ   N_NN     NULL
ഇരുട്ട്    N_NN     NULL
കയറുമ്പോൾ         N_NNP    NULL
ഉത്തരം   N_NN     ചോദ്യോത്തരം
കിട്ടാതെ  V_VM_VNF           NULL
വട്ടംകറങ്ങുമ്പോൾ  V_VM_VNF          NULL
തലയിൽ  N_NN     NULL
മിന്നുന്ന   JJ       NULL
ക്രിയേറ്റിവ്         N_NN     NULL
ഐഡിയാസ്          N_NN     NULL
എഴുതി   V_VM_VNF           NULL
വയ്ക്കാൻ V_VM_VNF           NULL
കൂടിയുള്ളതല്ലേ       N_NN     NULL
.          RD_PUNC NULL

ഒരു       QT_QTC   NULL
ഗ്ലാസ്സ്     N_NN     NULL
രസം N_NN  കറി
കുടിക്കാൻ          V_VM_VNF          NULL
കൊതിയാവുന്നു      V_VM_VF  NULL
.          RD_PUNC NULL

മിനി     N_NN     NULL
സ്കേർട്ടിട്ടാൽ      V_VM_VNF          NULL
മാനം     N_NN     അഭിമാനം
പോകുമോV_VM_VINF          NULL
```

Figure 3: Sample Tagged Corpus

For example, where disambiguation involves the classes മാനം and അഭിമാനം, TP (TN) refers to the മാനം (അഭിമാനം) test occurrences correctly classified as such by the system. Likewise, FP (FN) refers to those അഭിമാനം (മാനം) test occurrences that have been misclassified by the system as belonging to class മാനം (അഭിമാനം).

The experimental results for the 10-fold cross validation test for the CRF-based Malayalam word sense disambiguation system with Word feature and Word+POS feature are presented in 2 and 3 respectively.

The system has demonstrated overall average precision, recall, F- measure values of 58.688, 53.678, and 52.359 respectively for Word Feature. The result shows the overall average precision, recall, F-measure values are 61.387, 49.454, and 51.75 respectively for Word +POS feature.

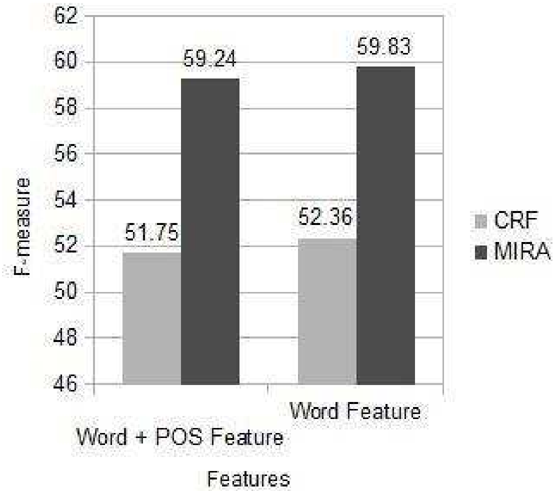

Figure 4: Analysis of F-measure result

The experimental results for the 10-fold cross validation test for the MIRA-based Malayalam word sense disambiguation system with Word feature and Word+POS feature are presented in Table 3. The system has demonstrated overall average precision, recall, F- measure values of 62.598, 63.045, and 59.829 respectively for word feature and 61.387, 49.454, and 59.75 word+POS feature respectively.

The performance evaluation of the models are done using F-measure. Using the value of F-measure the performance result presented in Fig 2 shows that in word feature and word + POS feature, MIRA model outperforms with the CRF model. The use of simple contextual feature give a little improvement for both CRF and MIRA model. Using the F measure, the performance results displayed in the above figure show that regardless of the contextual features or POS information feature the MIRA-based tagger outperforms CRF based framework.

7 Conclusion and Future Directions

7.1 Conclusion

This work addresses CRF based word-sense disambiguation with two different approaches. CRF provides flexibility to include diversity of features. We have used two algorithms in CRF framework which is basic Conditional Random Field algorithm and Margin Infused Relaxed (MIRA) algo-

Test Set	Word Feature			Word + POS Feature		
Sl. no	Precision	Recall	F-measure	Precision	Recall	F-measure
1	57.25	43.74	47.72	58.42	54.2	49.53
2	69.08	56	56.84	67.09	53.99	57.15
3	42.36	46.09	41.29	66.18	50.22	55.13
4	41.34	43.69	40.04	56.69	48.9	49.81
5	68.85	65.27	61.99	61.11	46.2	48.95
6	52.3	56.63	52.22	61.11	50.6	52.88
7	70.05	65.72	62.59	68.39	51.47	55.08
8	66.01	57.69	55.89	68.14	45.83	52.61
9	60.31	52.27	53.94	44.3	40.22	41.47
10	59.33	49.68	51.07	62.44	52.82	54.89
Average	58.68	58.69	52.35	61.387	49.454	51.75

Table 2: RESULTS OF 10 FOLD CROSS VALIDATION USING CRF FOR WORD AND WORD+POS FEATURE

Test Set	Word Feature			Word + POS Feature		
Sl. no	Precision	Recall	F-measure	Precision	Recall	F-measure
1	73.48	71.33	71.05	58.42	54.51	54.88
2	51.9	66.7	56.04	56.37	66.4	57.46
3	73.22	73.71	68.75	60.79	56.07	56.01
4	49.14	54.7	46.13	72.59	63.15	65.55
5	62.1	60.03	58.79	59.49	52.99	52.05
6	67.24	71.2	66.46	70.16	57.22	61.85
7	68.44	58.01	60.62	67.99	57.52	59.8
8	59.5	57.24	55.72	69.05	67.75	65.59
9	70.41	69.83	67.25	62.42	54.87	56.81
10	50.55	47.7	47.48	63.52	63.54	62.44
Average	62.598	63.045	59.829	63.926	59.402	59.244

Table 3: RESULTS OF 10 FOLD CROSS VALIDATION USING MIRA FOR WORD AND WORD+POS FEATURE

rithm. A word sense tagger is created for Malayalam to get an effective word disambiguation using CRF and MIRA. The system is evaluated with manually created words and the accuracy is measured using n-fold cross validation. Results based on the value of F-Measure shows that the performance of MIRA gives the best results with an overall average for word feature precision, recall, F-measure of 62.598, 63.045 and 59.829 respectively for 10-folds. The experimental results are very promising when large amount of annotated corpus was used and handling morphology exhaustively. More words and senses can be added to this so as to increase the accuracy. Other machine learning techniques like Naive Bayes classifier, ME, Neural Networks etc can be applied in this study and the results so obtained can be compared with the existing works.

References

Yuret D. 2004. *Some experiments with a naive bayes wsd system.In Senseval-3: , Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text.* 265-268U.

Suarez A Palomar. 2002. *A maximum entropy-based word sense disambiguation system..*, In*Proceedings of the 19th international conference on Computational linguistics. Association for Computational inguistics* 1 :1-7.

Chatterjee N Misra. 2009. *Word-Sense Disambiguation using maximum entropy model.* , In*Methods and Models in Computer Science ICM2CS 2009, Proceeding of International Conference on. IEEE.* 1-4.

Yarowsky D. 1994. *Decision lists for lexical ambiguity resolution: application to accent restoration in Spanish and French*, In *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics.* 88-95.

Yarowsky D. 1995. . *Unsupervised word sense disambiguation rivaling supervised methods.*, *In Proceedings of the 33rd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics*. 189-196..

Parameswarappa SS Narayana. 2011. *Target word sense disambiguation system for Kannada language.* , In *Advances in Recent Technologies in Communication and Computing (ARTCom 2011), 3rd International Conference on.* IET. 269-273.

Lesk M. 1986. *Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone.* In *Proceedings of the 5th annual international conference on Systems documentation.* ACM.24-26.

Sinha M Kumar M Pande P Kashyap L Bhattacharyya. 2004. *Hindi word sense disambiguation*, In *International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems*

Gaona Gelbukh Bandyopadhyay. 2009. *Web-based variant of the Lesk approach to word sense disambiguation. In Artificial Intelligence* In *International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems.*.

Kolte S G Bhirud S G. 2009. *WordNet: a knowledge source for word sense disambiguation International Journal of Recent Trends in Engineering.*

Haroon R 2010. *Malayalam word sense disambiguation.* In *Computational Intelligence and Computing Research (ICCIC). IEEE International Conference on. IEEE*. 1-4.

Wallach H M 2004. random fields: An introduction.Technical Reports , In *CIS*. 22.

Zhang J Xu J Zhang Y. 2013. *Name Origin Recognition in Chinese Texts Based on Conditional Random Fields.*, In *2013 I nternational Conference on Information Science and ComputerApplications (ISCA 2013). Atlantis Press*.

McCallum. 2002. *Efficiently inducing features of conditional random fields.*, In *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc*. 403-410.

Lafferty J McCallum A Pereira F C. 2001. *Conditional random fields: Probabilistic models for segmenting and labeling sequence data.*,

Crammer K Singer Y. 2003. . *Ultraconservative online algorithms for multiclass problems.*, *The Journal of Machine Learning Research*. 3:951-991.