

# CHAPTER 7

## Conclusion

This chapter concludes the thesis by summarizing the works, findings and contributions of the thesis. It also presents some directions for future research.

### 7.1 Summary of Works

This thesis investigated the application of a WSD technique for efficient disambiguation of texts belonging to two domains namely Tourism and Health. The research work for the thesis began with forming a well-grounded overview on paradigms of WSD techniques. Each WSD technique is very powerful by itself but when combined with other WSD techniques results in a more effective approach. Chapter 6 of this thesis presented an efficient hybrid WSD algorithm for document clustering.

A human mind can grasp the nuances and subtleties of words, in a context, with little effort but for a computer it becomes an absolutely critical problem. The task of WSD is to obtain the sense of target words or all words against a sense repository like the WordNet or a Thesaurus using the context in which the word appears. The problem of WSD has been studied by researchers for a significant period of time and a number of techniques for tackling this problem have been proposed from time to time. A summary of the contributions of the thesis are presented below:

Chapter 2 presents a literature review on various WSD techniques ,sense tagged document corpus of various languages , artificial Neural network , Hopfield network and the energy function of Hopfield network.

Chapter 3 discussed WordNet ,IndoWordnet and a Sense marking tool for sense marking a corpora at length . In all our experiments the senses of the words were obtained from IndoWordnet . This chapter has discussed in detail various syntactic and semantic relations of WordNet . Some of the semantic relations were of special importance to us as we have used those while proposing our algorithm. Supervised WSD requires sense marked corpora for disambiguation. A sense marking tool for sense marking a corpora has been discussed in this chapter.

Chapter 4 discussed WSD and various techniques of performing WSD. All the broad techniques of WSD namely knowledge based, Supervised , Semi-Supervised and Unsupervised techniques were studied at length to form a well-formed idea as to how this critical task of WSD may be tackled. Comparisons were also done among various algorithms coming under the classes of WSD techniques mentioned above to determine relative merits and demerits of those algorithms.

Chapter 5 discussed a lemmatizer for obtaining root words of a language using trie data structure . The words in a corpora come in morphed form whereas the words in a WordNet appears only in its root form. So an efficient lemmatizer is of utmost importance. This chapter introduced a lemmatizer for languages namely Hindi , Bengali and Nepali. This chapter also presented a study on a type of artificial Neural network called Hopfield network and the energy function of Hopfield network. The energy function has been adapted for use in our proposed algorithm.

Chapter 6 discussed about various parameters of WSD where we have taken the recourse of some semantic relations present in WordNet and proposed a hybrid algorithm for performing WSD in two domains namely Tourism and Health The implementation of Overlap based approach was presented from where we have obtained a parameter called knowledge\_concurrence which is used with corpus and WordNet based parameters in our proposed algorithm. The algorithm proposed combines knowledge from multiple knowledge sources. This chapter showed that the

proposed algorithm resulted in enhanced performance over the standard baseline and the knowledge based algorithm PPR . It was observed from experiments that some amount of infusion of target data with source data during training results in overall improvement of performance.

## **7.2 Summary of Contributions**

- A systematic survey of WSD approaches and techniques, WordNet and its various relations, Indo WordNet and lemmatizers.
- A Hybrid algorithm for WSD experimented on 2 domains, namely, Tourism and Health.
- Encouraging result that if sense marked data from one Domain (for e.g. A) is available, then a WSD system can be built for another Domain( for e.g. B) by infusing some amount of data from Domain-B into Domain-A.

## **7.3 Future Directions**

The proposed algorithm was implemented on two domains namely Tourism and Nepali across two languages namely Bengali and Nepali. We would like to test the algorithm on other domains of some of the scheduled languages of India . We have used some semantic relations namely Hypernymy-Hyponymy, modifies-Noun from WordNet in our work. We would like to use other semantic relations from WordNet to find out their impact on the algorithm proposed.

In the case of lemmatizer tool, when the lemmatizer cannot determine the exact root word of an inflected word it should ideally show the list of root words in the trie, very close to the inflected word, to the person using the lemmatizer so that the person using his domain knowledge of the language concerned can choose the appropriate root word. We envisage to include this functionality in the tool in the near future.