7.1. Conclusions

These days, machine learning techniques are being widely used to solve real-world problems by storing, manipulating, extracting and retrieving data from large sources. Supervised machine learning techniques have been widely adopted however these techniques prove to be very expensive when the systems are implemented over wide range of data. This is due to the fact that significant amount of effort and cost is involved because of obtaining large labeled data sets. Thus active learning provides a way to reduce the labeling costs by labeling only the most useful instances for learning.

Chapter 2 discusses current developments and applications in NLP and literature survey of various machine learning techniques. We identified the different circumstances in which the learner may ask queries and different querying strategies. Chapter 3 discusses the basic concepts of supervised learning, active learning and learning for complex models. We presented an example of learning pipeline models. We concluded that machine learning strategies that take into consideration the informativeness or the relevance of instances can perform better with fewer labeled examples as compared to other learning approaches. Chapter 4 examines a pipelined approach for information extraction with respect to active machine learning. Machine learning problems solved using a pipeline model show better results. Chapter 5 presents an evaluation of state-of-the-art machine learning algorithms on the basis of efficiency, for the task of classification. Chapter 6 presents a combined approach for the design of a learner that aims at increasing the efficiency of the learning tasks. Machine learning algorithms perform more efficiently for a classification task when they are combined together. For the prediction of the correct output class, combined learner selects the class to which highest probability has been assigned among all the learners. Further we conclude that feature selection is important but only as long as it does not decrease the efficiency of the learners by discarding too many attributes on the basis of their relevance.

7.2. Future Work

The combined approach that we presented in this work has some limitations. Although we have used it on state-of-the-art machine learning algorithms, however, we have evaluated its results on only classification tasks. It can be extended to be used for other important problems e.g.
regression and clustering. Moreover, we theoretically showed how active learning can be applied to part-of-speech tagging and included into the pipeline. In future we intend to show its empirical implementation and performance evaluation using various evaluation metrics. In field of active learning future work involves combining active learning with a subfield of machine learning called transfer learning [Torrey and Shavlik, 2009]. It is applicable in situations when we have a training set available for one problem but not for another similar problem. It involves transferring knowledge from one domain to another to speed up learning.