Optimizing Annotation Effort Using Active Learning Strategies: A Sentiment Analysis Case Study in Persian

Seyed Arad Ashrafi Asli, Behnam Sabeti, Zahra Majdabadi, Preni Golazizian,
Reza Fahmi and Omid Momenzadeh

Miras Technologies International
Number 92, Movahed Danesh St., Tehran, Iran
{arad, behnam, zahra, preni, reza, omid}@miras-tech.com

Abstract

Deep learning models are the current State-of-the-art methodologies towards many real-world problems. However, they need a substantial amount of labeled data to be trained appropriately. Acquiring labeled data can be so challenging in some particular domains or less-resourced languages. There are some practical solutions regarding these issues, such as Active Learning and Transfer Learning. Active learning's idea is simple: let the model choose the samples for annotation instead of labeling the whole dataset. This method leads to a more efficient annotation process. Active Learning models can achieve the baseline performance (the accuracy of the model trained on the whole dataset), with a considerably lower amount of labeled data. Several active learning approaches are tested in this work, and their compatibility with Persian is examined using a brand-new sentiment analysis dataset that is also introduced in this work. *MirasOpinion*, which to our knowledge is the largest Persian sentiment analysis dataset, is crawled from a Persian e-commerce website and annotated using a crowd-sourcing policy. *LDA sampling*, which is an efficient Active Learning strategy using Topic Modeling, is proposed in this research. Active Learning Strategies have shown promising results in the Persian language, and *LDA sampling* showed a competitive performance compared to other approaches.

Keywords: Active Learning, Sentiment Analysis, Dataset Generation, Topic Modeling

1. Introduction

Despite the notable performance, deep learning approaches are data-hungry. They need huge amounts of data to achieve their optimum performance. This data dependency had become a bottleneck in many real-world applications. Many domains require a specialist to annotate the data samples, for instance, the medical domain. Thus, this annotation process can be time-consuming and expensive in many cases. Unfortunately, in less-resourced languages like Persian, the above-mentioned problem is even worse, and there are more limitations in attaining related datasets. So there is an increasing need to make the most out of limited available resources.

In recent years, some researchers tried to overcome this deficiency in the Persian language. For example, MirasText (Sabeti et al., 2018), which is a large Persian corpus with more than 1.4 billion tokens and over 2.8 million documents, is automatically generated. More specifically, in the sentiment analysis domain, there were some efforts in enriching sentiment lexicon for Persian, such as LexiPers (Sabeti et al., 2019), PerSent (Dashtipour et al., 2020), Senti-Fars (Dehkharghani, 2019), and HesNegar (Asgarian et al., 2018). However, these lexicons are mainly used in unsupervised sentiment analysis methods.

Despite those developments, insufficient resources remain as the main difficulty in many domains in Persian. Active learning is one of the effective solutions to this issue. Because it can facilitate the labeling process by lowering the costs and minimizing the annotation effort. The idea behind active learning is to let the learning algorithm choose the data it needs to learn from. This way the model can perform better even with less amount of data compared to traditional algorithms.

Several strategies have been proposed to utilize active learning ideas. These strategies include uncertainty sampling (Lewis and Catlett, 1994; Settles, 2012), Query By Committee (Seung et al., 1992), Expected Model Change (Cai et al., 2013; Käding et al., 2016), Expected Error Reduction (Roy and McCallum, 2001; Guo and Greiner, 2007; Käding et al., 2018), Variance Reduction (Schein and Ungar, 2007).

Active Learning strategies could be applied on numerous classifiers, like SVM classifiers(Tong and Koller, 2001), Neural Networks, and Bayesian approaches (Siddhant and Lipton, 2018).

In this research, we propose to transform the data into a semantic-based latent space using Latent Dirichlet Allocation (LDA) and utilize this new representation for sample selection (Blei et al., 2003). *LDA sampling* utilizes topic modeling to inject representativeness; then, it uses entropy to obtain the most informative instances from each topic. Although *LDA sampling* has a better performance in its first iterations, but all other strategies also show encouraging performance, and it is not possible to choose one policy that overcomes others in all cases.

In order to evaluate previous and proposed strategies, we also introduce the largest sentiment analysis dataset in Persian. *MirasOpinion* raw data is crawled from Digikala[1] comments section. The gathered comments are then annotated using a crowd-sourcing policy.

For the sentiment analysis architecture, two different models are considered: LSTM-based and CNN-based architectures.

Our contributions in this research are as follows:

- Introducing the largest Persian sentiment analysis dataset.

---

[1] www.Digikala.com

- Verification of active learning baseline approaches compatibility with Persian.

- Proposing a novel active learning strategy-based on topic modeling and entropy.

The remainder of the paper is organized as follows. Section 2. reviews related works. Section 3. presents The Sentiment Analysis model architecture used in this work. Active Learning strategies alongside our proposed method are discussed in detail in Section 4.. In Section 5., details of *MirasOpinion* dataset is presented; followed by the sentiment analysis model's evaluation process. Results are further discussed in section 6. Finally, we conclude the paper in Section 7.

2.   Related Work

Development of active learning methodologies, or more precisely pool-based active learning(Settles, 2012), gives rise to various methods for label sampling. One category of these methods is uncertainty-based strategies such as Least Confident, Margin, and Entropy Sampling(Lewis and Catlett, 1994; Settles, 2012). Their basic premise is that the learner can avoid querying the instances it is already confident about (Settles, 2012). Uncertainty-based strategies are the most convenient way for querying labels. Despite the simple nature of these methods, they have shown high performance in all sorts of tasks (even higher performance in some tasks compared to other complicated approaches).

Besides uncertainty-based methods, there are other strategies such as Query By Committee (Seung et al., 1992), Expected Model Change (Cai et al., 2013; Käding et al., 2016), Expected Error Reduction (Roy and McCallum, 2001; Guo and Greiner, 2007; Käding et al., 2018), Variance Reduction (Schein and Ungar, 2007) and clustering-based Methods (Xu et al., 2003; Nguyen and Smeulders, 2004) for querying new samples (Settles, 2012; Olsson, 2009). Most of these heuristics can be divided into two general categories (or a combination of these two): informativeness and representativeness. Informativeness strategies focus on gaining knowledge by decreasing the uncertainty of the statistical model (Du et al., 2019; Yang and Loog, 2017). Query by Committee, Uncertainty Sampling, Expected Error Reduction, and Expected Model change belong to informativeness strategies. On the other hand, representativeness strategies consider distribution of data in sampling, such as Density-Weighted and Variance Minimization approaches. Some try to take both of these into account and create models to query new samples, which are both informative and representative(Du et al., 2017; Huang et al., 2014; Tang and Huang, 2019).

There are also multi-label active learning models. They work in situations where each instance has more than one label(Yang et al., 2009; Reyes et al., 2018; Huang et al., 2015). These methods, however, are not discussed in this paper as they are beyond the scope of our research.

There are a few attempts of employing active learning methods in Persian. For example, (Ghayoomi and Kuhn, 2013) examined entropy sampling and query by committee approaches for parsing and treebanking.

We have examined several methods from these strategies and tracked their integrity with Persian language's semantics. Our proposed method, *LDA sampling*, uses LDA (Blei et al., 2003) for the documents' representativeness, which is somehow close to representative strategies, like clustering. Then it uses entropy to rank the most informative options in each topic to be queried. This method is explored in detail in section 4. (Active Learning Strategies).

3.   Sentiment Analysis Model

In this research, Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Convolutional Neural Network (CNN) have been selected as classifiers due to their acceptable performance in Persian sentiment analysis (Roshanfekr et al., 2017).

In order to further analyze sequential models, we have also tested bidirectional long short-term memory (Bi-LSTM) (Graves and Schmidhuber, 2005) and Gated Recurrent Unit (GRU) (Chung et al., 2015). Bi-LSTM is examined to explore documents from both ends, and GRU tested in order to decrease the model's parameters. However, they didn't make any notable changes in our results. All of them (LSTM, Bi-LSTM, GRU) showed a satisfying and similar performance as a multi-class classifier; based on the examinations, the model's behavior with an LSTM classifier is quite extensible on a Bi-LSTM or a GRU (with our dataset). Thus, active learning strategies are explored on an LSTM-based architecture as well as the CNN-based architecture. Model Architecture is explained in Figure 1.

We did not use any pre-trained embeddings and instead employed an embedding layer. Having implemented two mentioned architectures, we trained them on our whole dataset and used their performance as our baseline for comparison with different strategies. Several active learning approaches were then tested on these architectures to check if they improve the baseline performance.

LSTM-based Sentiment Analysis model's main parameters and Architecture are summarized in Table 1

Model Parameter	Amount
Sequence Length	128
Embedding Size	100
Units	128
Activation	tanh
Recurrent Activation	hard sigmoid
Recurrent Initializer	Orthogonal
Kernel Initializer	Glorot Uniform
Recurrent Dropout	0.3

Table 1: LSTM Model Parameters

CNN-based Sentiment Analysis model's main parameters and Architecture are also summarized in Table 2:

4.   Active Learning Strategies

Several available active learning strategies are applied to both sentiment analysis models introduced in the previous

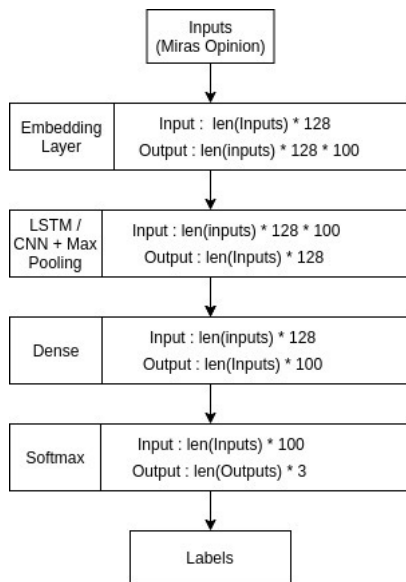| Model Parameter	Amount
Sequence Length	128
Embedding Size	100
kernel size	128
strides	10
Activation	ReLU
Kernel Initializer	Glorot Uniform

Table 2: CNN Model Parameters

Figure 1: Model Architecture

section: LSTM-based and CNN-based. Sentiment analysis models are trained with limited labeled samples and then active learner layer is applied to the classification distribution to query new samples.

The idea behind available strategies and their challenges inspired our proposed method. All of the examined strategies will be explained in detail in this section.

4.1. Traditional Strategies

Informative strategies had been examined first. Uncertainty samplings are chosen among them due to their perfect performance and simplicity.

Entropy sampling, Margin sampling, and Least confident (LC) strategies (Settles, 2012) applied to our sentiment analysis models' classification distribution (softmax output) and determine which samples to annotate next.

Consider $p(y_1)$, $p(y_2)$, $p(y_3)$ as classification prediction's output for unlabeled sample x and they are sorted in decreasing order ($p(y_1)$ the highest probability and $p(y_3)$ the lowest). Different uncertainty algorithms calculate uncertainty of this example as follows:

$$Uncertainty_{entropy} = -\sum_{i=1}^{3} p(y_i)log(p(y_i)) \quad (1)$$

$$Uncertainty_{margin} = 1 - (p(y_1) - p(y_2)) \quad (2)$$

$$Uncertainty_{LC} = 1 - p(y_1) \quad (3)$$

After calculating uncertainties for all unlabeled samples, samples with highest entropy are selected to be queried and labeled.

Then, we tried to take data representativeness into consideration. The last LSTM unit's hidden state or pooling's output in the CNN model were extracted for each document as the document's representation. Then clustering methods such as DBSCAN (Ester et al., 1996) are applied to these vectors. This process was time-consuming since it needs lots of resources to cluster data in high dimensionality. This clustering process also needs to be repeated each time after the model is retrained due to the changes in the hidden state or pooling output, which affects the embedding of each document in our pool. This complex process without any notable contribution seems unnecessary and sparks the idea behind our proposed method: *LDA sampling*.

4.2. Proposed Method

Topic Modeling is an efficient way to analyze large text datasets. LDA is the common and most frequently used algorithm for topic modeling (Deerwester et al., 1990; Blei et al., 2003). LDA is a three-level hierarchical Bayesian model. Its intuition could be simplified as: 1. clustering words into topics (each topic is a probability distribution over words), 2. clustering documents as a combination of extracted topics (Using Bayesian Inference to assign each document a probability distribution over topics).

Using the idea of LDA topic modeling, we create a topic model for *MirasOpinion* corpus. Then we assign each document to the most probable topic in the corresponding probability distribution in order to cluster documents in different groups. Employing topic modeling instead of distance-based clustering methods lead to a significant decrease in time and resources. Unlike clustering methods which measure the distance between data, LDA will only take an initial time to infer topics purely based on word counts and co-occurrences, based on the bag-of-words (Harris, 1954) representation of documents.

Using LDA solely and query samples only based on topics is a good baseline; however, it can still be improved. Choosing instances to query in each topic without the help of any other algorithms is highly unlikely to result in the optimum performance. As stated before, clustering with LDA does not carry the notion of distance explicitly. In the distance-based clustering, we could use some insight and make some distance-based rules, such as querying ones in the boundary and the center of the cluster before others. However, none of those are applicable to our case. We could not make similar assumptions and take the instances with the highest scores in each topic or choosing ones with the lowest ranking in their cluster or even sample from both sides.

In this study, the Entropy Sampling strategy (1) was used to query the most informative data within each topic (due to entropy's negligible predominancy between uncertainty

sampling strategies, which is discussed more in the result section). So, now the same procedure of entropy sampling will be applied to each topic, and each topic has a share of total new queries corresponding to its length. The more members a topic possess, the more sampling share will be devoted to that topic. Then the selected instances from all topics will be aggregated and sent to the oracle to be labeled.

Thus, this way, we measure both the informativeness, by Entropy sampling, and the representativeness, with topic modeling inside our strategy. Algorithm 1 provides the general idea behind *LDA sampling*.

Initial Time:
$TopicModels \leftarrow LDA(Dataset)$
$Labeled \leftarrow random(Dataset, InitialSamplesNumber)$
//randomly selecting initial samples to train model
Active Learning Cycle:
$SentimentModel.train(Labeled)$
$SamplesToQuery \leftarrow [\,]$
//Selected samples to query will be appended to this array $for\ topic\ in$
  $TopicModels$:
  $TopicShare \leftarrow len(topic)/len(unlabeld) * totalNewSamples$
  //Current topic's share in new queries proportional to its size
  $Infos \leftarrow [\,]$
  //Informative Samples in each topic
  $for\ doc\ in\ topic$:
    $Info \leftarrow Entropy(SentimentModel.predict(doc))$
    //Calculating entropy of samples' predicted distribution in each topic
    $Infos.append(Info)$
  $Selected \leftarrow HighestEntropy(Infos, TopicShare)$
  //Selecting the most informative samples proportional to topic's share
  $SamplesToQuery.append(Selected)$
  $topic.remove(Selected)$
  //Removing selected samples from unlabeled pool
$Labeled.append(Query(SamplesToQuery))$
//Appending newly queried samples to Labeled ones

Algorithm 1: LDA sampling pseudo code

5. Experiments

5.1. Dataset

MirasOpinion, our dataset, is crawled from the Digikala website, one of the largest e-commerce websites in Iran. 2.5 million comments have been crawled, and after some pre-processing, we reduce its size to one million comments. Then the corpus had been labeled using crowd-sourcing; A telegram bot is used to send the unlabeled data to several users. Our bot asks them to label the represented document as positive, negative, or neutral. Table 3 provides a summary of our dataset statistics.

Total Documents	93868
Max Length	1434
Min Length	3
Mean Length	38.15
Positive Comments	49515
Negative Comments	14882
Neutral Comments	29471

Table 3: *MirasOpinion* Dataset Statistics

5.2. Evaluation Process

Before starting the training process, filtering was applied to our dictionary. Words with less than five occurrences or appearing in more than 40 percent of documents are pruned.

Before training sentiment analysis mode, an initial state of the model was saved; then, this initial model was used in each iteration of the active learner to train the model from scratch.

Training started with a limited number of random samples (1000). After the initial training, the model decides the samples it wanted and begins to query them. In each iteration model queries 1000 new instances that are selected based on the chosen strategy.

Also, in each iteration, recently queried samples will be appended to the beginning of the labeled data, instead of its end; the intuition behind this idea is to update the model's weights with the most informative and representative instances in the last batches, which shows a slight improvement.

Table 4 provides some of the training process's hyperparameters.

Hyperparameters	Amount
Epochs	10
Batch Size	512
Dropout	0.3
Train Ratio	0.8

Table 4: Model Hyperparameters

LDA clustering parameters are presented in Table 5. The number of clusters had been chosen by trial and error. Persian stopwords for topic modeling [2] were used to filter vocabulary before applying LDA clustering.

Parameters	Amount
Number of Topics	10
Max Iteration	5
Learning Method	Online
Learning offset	50
Random State	0
learning decay	0.7

Table 5: LDA parameters

6. Results

Before applying active learning strategies, we measured accuracy, Precision, Recall, and F1 score for both CNN and LSTM models, which were trained on the whole dataset. Although they show close outcomes in all of these metrics, the CNN training process takes a significantly lower time compared to LSTM. CNN trains four times faster than

---

[2]https://github.com/kharazi/persian-stopwords/blob/master/persian

LSTM. This difference in time could be important in the real-world scenarios which we have some oracles waiting for active learner queries to annotate them.

A comparison between CNN-based and LSTM-based architectures' performance is shown in table 6.

Model	Precision	Recall	F1	Accuracy
CNN	78.4	78.2	78.3	78.2
LSTM	79.1	78.7	78.9	78.7

Table 6: Comparison between CNN and LSTM models' performance on *MirasOpinion* Dataset

Testing entropy, margin, and least confident, we observed that all of them reach the baseline accuracy (accuracy of the model trained on the whole data) on the CNN architecture with less than 16 percent of total data. Results on the LSTM-based architecture are similar to the CNN-based one; thus, we only provide the results on the CNN model. Figure 2 gives a complete visual comparison between Uncertainty strategies.
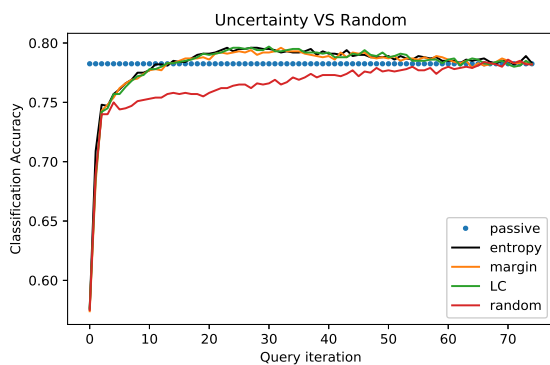

Figure 2: Uncertainty Sampling Comparison

As it is shown in the figure 2, none of the uncertainty-based strategies outperform others significantly (it seems entropy strategy is a little bit better, but the difference is quite negligible).

LDA sampling performance is truly competitive with other strategies. Entropy sampling is selected as the representer of uncertainty strategies. Figure 3 compares *LDA sampling* with Entropy sampling.

Density-based approaches have also been evaluated. However, due to their high computational needs, after testing them with a limited portion of data and observing no notable contribution, we decided not to explore them furthermore.

Also, it is shown in Figure 2 and Figure 3 the model could reach the baseline approach with less than 16 percent of labeled data. *LDA sampling* had the best performance; sentiment analysis models (LSTM-based, CNN-based), which used *LDA sampling*, reaches the baseline approach's accuracy with 14 percent of total data. The integration of the uncertainty strategies with the sentiment analysis models resulted in a quite close performance; they reach the
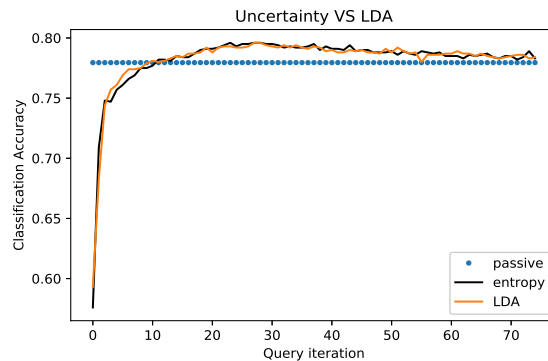

Figure 3: LDA sampling VS Uncertainty sampling(entropy)

baseline accuracy with 16 percent of labeled data. Table 7 provides a brief overview of *LDA sampling* performance (best strategy) that have been examined on *MirasOpinion* dataset. Performance is compared based on the F1 score. Model A is the first time that the model reaches the baseline, as explained before in this section. Model B is the best performance of the sentiment analysis model.

	baseline	Model A	Model B
F1 Score	78%	78%	**80%**
Used Data	100%	**14%**	39%

Table 7: Active Learning strategy (LDA sampling) Performance at a glance

After labeling 39% of the data, the model's performance began to decrease. After reaching this point, the model began to query less valuable data, and overfitting occurred. It is possible to detect this condition and apply a stopping criterion; for example, the variance of the acquired F1-scores begins to rise because of its fluctuations (Ghayoomi, 2010).

7. Conclusion and Future Works

In this paper, we introduced *MirasOpinion*, which is the largest Persian sentiment analysis dataset. Two different models for Sentiment Analysis were examined on this dataset: LSTM and CNN. Both of these models reach approximately 80 percent in the F1 score. It is worth mentioning that CNN's training process takes a substantially less amount of time (approximately 25 percent of our LSTM's training time).

We also proposed a novel Active Learning strategy for text-based classification tasks called *LDA sampling*. This method uses Latent Dirichlet Allocation with the aim of gaining an overview of the documents' representation by clustering them into different topics. Then, the model uses entropy sampling to query the highest informative instances inside each clustered topic. Our method shows a competitive performance compared to the other uncertainty-based approaches, especially in the early stages.

Both of the sentiment analysis models reach and pass their baseline accuracy with less than 16 percent of labeled data,

which were chosen by the active learner.

We have only tested our proposed approach to the sentiment analysis task; however, this method needs to be further evaluated on other tasks with different datasets as well.

8. Acknowledgements

9. References

Asgarian, E., Kahani, M., and Sharifi, S. (2018). Hesnegar: Persian sentiment wordnet. *Signal and Data Processing*, 15:71–86, 06.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March.

Cai, W., Zhang, Y., and Zhou, J. (2013). Maximizing expected model change for active learning in regression. *2013 IEEE 13th International Conference on Data Mining*, pages 51–60.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2015). Gated Feedback Recurrent Neural Networks. *arXiv e-prints*, page arXiv:1502.02367, Feb.

Dashtipour, K., Raza, A., Gelbukh, A., Zhang, R., Cambria, E., and Hussain, A., (2020). *PerSent 2.0: Persian Sentiment Lexicon Enriched with Domain-Specific Words*, pages 497–509. 02.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, 41(6):391–407.

Dehkharghani, R. (2019). Sentifars: A persian polarity lexicon for sentiment analysis. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 19(2), September.

Du, B., Wang, Z., Zhang, L., Zhang, L., Liu, W., Shen, J., and Tao, D. (2017). Exploring representativeness and informativeness for active learning. *IEEE Transactions on Cybernetics*, 47(1):14–26, Jan.

Du, B., Wang, Z., Zhang, L., Zhang, L., Liu, W., Shen, J., and Tao, D. (2019). Exploring Representativeness and Informativeness for Active Learning. *arXiv e-prints*, page arXiv:1904.06685, Apr.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pages 226–231. AAAI Press.

Ghayoomi, M. and Kuhn, J. (2013). Sampling methods in active learning for treebanking. In *Proceedings of the 12th Workshop on Treebanks and Linguistic Theories (TLTâ12)*, pages 49–60.

Ghayoomi, M. (2010). Using variance as a stopping criterion for active learning of frame assignment. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, ALNLP â10, page 1â9, USA. Association for Computational Linguistics.

Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm networks. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 4, pages 2047–2052 vol. 4, July.

Guo, Y. and Greiner, R. (2007). Optimistic active learning using mutual information. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 823–829, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Harris, Z. S. (1954). Distributional structure. ¡i¿WORD¡/i¿, 10(2-3):146–162.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9:1735–80, 12.

Huang, S., Jin, R., and Zhou, Z. (2014). Active learning by querying informative and representative examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10):1936–1949, Oct.

Huang, S.-J., Chen, S., and Zhou, Z.-H. (2015). Multi-label active learning: Query type matters. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 946–952. AAAI Press.

Käding, C., Freytag, A., Rodner, E., Perino, A., and Denzler, J. (2016). Large-scale active learning with approximations of expected model output changes. In Bodo Rosenhahn et al., editors, *Pattern Recognition*, pages 179–191, Cham. Springer International Publishing.

Käding, C., Rodner, E., Freytag, A., Mothes, O., Barz, B., and Denzler, J. (2018). Active learning for regression tasks with expected model output changes. In *BMVC*.

Lewis, D. D. and Catlett, J. (1994). Heterogeneous uncertainty sampling for supervised learning. In *In Proceedings of the Eleventh International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann.

Nguyen, H. T. and Smeulders, A. (2004). Active learning using pre-clustering. In *Proceedings of the Twenty-first International Conference on Machine Learning*, ICML '04, pages 79–, New York, NY, USA. ACM.

Olsson, F. (2009). A literature survey of active machine learning in the context of natural language processing. 05.

Reyes, O., Morell, C., and Ventura, S. (2018). Effective active learning strategy for multi-label learning. *Neurocomputing*, 273:494 – 508.

Roshanfekr, B., Khadivi, S., and Rahmati, M. (2017). Sentiment analysis using deep learning on persian texts. In *2017 Iranian Conference on Electrical Engineering (ICEE)*, pages 1503–1508, May.

Roy, N. and McCallum, A. (2001). Toward optimal active learning through sampling estimation of error reduction. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 441–448, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Sabeti, B., Abedi, H., Janalizadeh Choobbasti, A., Najafabadi, S., and Vaheb, A. (2018). Mirastext an automatically generated text corpus for persian. 05.

Sabeti, B., Hosseini, P., Ghassem-Sani, G., and Mirroshandel, S. A. (2019). Lexipers: An ontology

based sentiment lexicon for persian. *arXiv preprint arXiv:1911.05263*.

Schein, A. I. and Ungar, L. H. (2007). Active learning for logistic regression: an evaluation. *Machine Learning*, 68(3):235–265, Oct.

Settles, B. (2012). *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool.

Seung, H. S., Opper, M., and Sompolinsky, H. (1992). Query by committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, pages 287–294, New York, NY, USA. ACM.

Siddhant, A. and Lipton, Z. C. (2018). Deep bayesian active learning for natural language processing: Results of a large-scale empirical study. *CoRR*, abs/1808.05697.

Tang, Y.-P. and Huang, S.-J. (2019). Self-paced active learning: Query the right thing at the right time. In *AAAI*.

Tong, S. and Koller, D. (2001). Support vector machine active learning with applications to text classification. In *JOURNAL OF MACHINE LEARNING RESEARCH*, pages 999–1006.

Xu, Z., Yu, K., Tresp, V., Xu, X., and Wang, J. (2003). Representative sampling for text classification using support vector machines. In *Proceedings of the 25th European Conference on IR Research*, ECIR'03, pages 393–407, Berlin, Heidelberg. Springer-Verlag.

Yang, Y. and Loog, M. (2017). A Variance Maximization Criterion for Active Learning. *arXiv e-prints*, page arXiv:1706.07642, Jun.

Yang, B., Sun, J.-T., Wang, T., and Chen, Z. (2009). Effective multi-label active learning for text classification. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 917–926, New York, NY, USA. ACM.