**Collaboratively Training Sentiment Classifiers for Multiple Domains**

**Abstract:-**In this project we introduced a collaborative multi-domain sentiment classification approach to train sentiment classifiers for multiple domains simultaneously. In this approach, the sentiment information in different domains is shared to train more accurate and robust sentiment classifiers for each domain when labeled data is scarce. Specifically, we decompose the sentiment classifier of each domain into two components, a global one and a domain-specific one. The global model can capture the general sentiment knowledge and is shared by various domains. The domain-specific model can capture the specific sentiment expressions in each domain. In addition, we extract domain-specific sentiment knowledge from both labeled and unlabeled samples in each domain and use it to enhance the learning of domain-specific sentiment classifiers. Besides, we incorporate the similarities between domains into our approach as regularization over the domain-specific sentiment classifiers to encourage the sharing of sentiment information between similar domains. Two kinds of domain similarity measures are explored, one based on textual content and the other one based on sentiment expressions. Moreover, we introduce two efficient algorithms to solve the model of our approach. Experimental results on benchmark datasets show that our approach can effectively improve the performance of multi-domain sentiment classification and significantly outperform baseline methods.

**Key words —**Sentiment Classification, Multiple Domain, and Multi-Task Learning.

**Existing System:-**In many mainstream sentiment analysis methods, sentiment classification is regarded as a text classification problem. Supervised machine learning techniques, such as SVM, Logistic Regression and CNN, are frequently applied to train sentiment classifiers on labeled datasets and predict the sentiments of unseen texts. These methods have been used to analyze the sentiments of product reviews, micro blogs and so on. However, sentiment classification is widely recognized as a domain-dependent problem. An intuitive solution to this problem is to train a domain specific sentiment classifier for each domain using the labeled samples of this domain. However, the labeled data in many domains is usually scarce. In addition, since there are massive domains involved in online user generated content, it is very costly and time-consuming to annotate enough samples for them. Without sufficient labeled data, it is quite difficult to train an accurate and robust domain-specific sentiment classifier for each domain independently. The motivation of our work is that although each domain has its specific sentiment expressions, different domains also share many common sentiment words. For example, general sentiment words such as “best”, “perfect”, and “worst” convey consistent sentiment polarities in various domains. Thus, training sentiment classifiers for multiple domains simultaneously and exploiting the common sentiment knowledge shared among them can help alleviate the problem of scarce labeled data and help learn more accurate sentiment classifiers for each domain.

**Disadvantage:-**

Sentiment classification is widely recognized as a domain-dependent problem. This is because in different domains different words are used to express sentiments, and the same word may convey different sentiments in different domains. For example, in the domain of electronic product reviews the word “easy” is usually positive, e.g., “this digital camera is easy to use.” However, in the domain of movie reviews, “easy” is frequently used as a negative word. For instance, “the ending of this movie is easy to guess.” Thus, the sentiment classifier trained in one domain may fail to capture the specific sentiment expressions of another domain, and its performance in a different domain is usually unsatisfactory.

**Proposed System:**

In the proposed system we used sentiment classifiers for multiple domains simultaneously in a collaborative way. In our approach, the sentiment classifier of each domain is decomposed into two components, i.e., a global one and a domain specific one. The domain-specific sentiment classifier is trained using labeled samples of one domain and can capture the domain-specific sentiment expressions. The global sentiment classifier is shared by all domains and is trained on the labeled samples from various domains to have better generalization ability. It can capture the general sentiment knowledge consistent in different domains. In addition, we extract prior general sentiment knowledge from general-purpose sentiment lexicons and incorporate it into our approach to guide the learning of the global sentiment classifier. Besides, we propose to extract domain specific sentiment knowledge for each domain from both limited labeled samples and massive unlabeled samples. The domain specific sentiment knowledge is used to enhance the learning of domain-specific sentiment classifiers in our approach. Moreover, since different pairs of domains have different sentiment relatedness, we propose to measure the similarities between domains and incorporate them into our approach to encourage the sharing of sentiment information between similar domains. Two kinds of domain similarity measures are explored, one based on the textual content, and the other one based on the sentiment word distribution. The model of our approach is formulated as a convex optimization problem. In order to solve it efficiently, we introduce an accelerated algorithm based on FISTA. In addition, we propose a parallel algorithm based on ADMM to further improve the efficiency of our approach when domains to be analyzed are massive.

**Advantages:**

1. Besides the single-node version algorithm for solving the model of our approach, in this paper we propose a parallel version algorithm, which is more efficient when there are a large number of domains to be analyzed.

2. To extract domain-specific sentiment knowledge by combining limited labeled samples with massive unlabeled samples, which is not considered in previous work.

3. Our approach can learn accurate sentiment classifiers for multiple domains simultaneously in a collaborative way and handle the problem of insufficient labeled data by exploiting the sentiment relatedness between different domains.

**Modules:**

1. **Multi-Domain Sentiment Classification**
2. **Multi-Task Learning**
3. **DOMAIN-SPECIFIC SENTIMENT KNOWLEDGE**

**AND DOMAIN SIMILARITY**

**Modules Description:**

1. **Multi-Domain Sentiment Classification:-**Sentiment classification has been widely known as a highly domain-dependent problem. Different domains have different ways to express sentiments, and a sentiment classifier trained in one domain usually perform not very well in another domain. For example, “easy” is a positive word in Kitchen domain (e.g., “this fryer is very easy to use”). However, it is frequently used as a negative word in Movie domain (e.g., “the ending of this film is easy to guess”). Thus, the sentiment classifier trained in Movie domain cannot predict the sentiment of “easy” in Kitchen domain accurately.
2. **Multi-Task Learning:-**The approach proposed in this module is based on multi-task learning. The aim of multi-task learning is to improve the generalization ability and prediction performance by learning multiple related tasks simultaneously and leveraging the common knowledge shared by these tasks appropriately. The main difference between different multi-task learning methods lies in how they model and incorporate the task relatedness.
3. **DOMAIN-SPECIFIC SENTIMENT KNOWLEDGE AND DOMAIN SIMILARITY**

In this module, we introduce two important components that will be used in our collaborative multi-domain sentiment classification approach (CMSC). The first one is the domain-specific sentiment knowledge, which is mined from massive unlabeled samples and a small number of labeled samples. It can provide prior knowledge of the sentiment expressions used in each domain. The second one is domain similarity, which measures whether two domains share similar terms and sentiment expressions.

**Future Scope:**

In this project we proposed to extract domain-specific sentiment knowledge from both labeled and unlabeled samples, and use it to enhance the learning of the domain-specific sentiment classifiers. Besides, we propose to use the prior general sentiment knowledge in general-purpose sentiment lexicons to guide the learning of the global sentiment classifier. In addition, we propose to incorporate the similarities between different domains into our approach as regularization over the domain specific sentiment classifiers to encourage the sharing of sentiment information between similar domains.

**Modified Title:**

1. **A Dynamic Training Sentiment Classifiers for Multiple Domains**
2. **Training Sentiment Classifiers By Using Multi-Task Learning**

# H/W System Configuration:-

# Processor - Pentium –IV

Speed - 3.5 GHz

RAM - 1GB (min)

Hard Disk - 40 GB

# S/W System Configuration:-

Operating System : Windows 7 or above

Application Server : IIS 7.0

Front End : ASP.Net (.Net 2013)

Database : SQL SERVER 2014

Database Connectivity : ADO.Net.

**Conclusion:**

In this project we introduced a collaborative multi-domain sentiment classification approach. Our approach can learn accurate sentiment classifiers for multiple domains simultaneously in a collaborative way and handle the problem of insufficient labeled data by exploiting the sentiment relatedness between different domains. In our approach, the sentiment classifier of each domain is decomposed into two components, a global one and a domain-specific one. The global model can capture the general sentiment knowledge shared by different domains and the domain-specific models are used to capture the specific sentiment expressions of each domain.