Text Classification for Intelligent Portfolio Management

Young-Woo Seo, Joseph A. Giampapa, and Katia P. Sycara

The Robotics Institute
Carnegie Mellon University

{ywseo, garof+, katia}@cs.cmu.edu

Abstract

In the application domain of stock portfolio management, software agents that evaluate the risks associated with the individual companies of a portfolio should be able to read electronic news articles that are written to give investors an indication of the financial outlook of a company. There is a positive correlation between news reports on a company's financial outlook and the company's attractiveness as an investment. However, because of the volume of such reports, it is impossible for financial analysts or investors to track and read each one. Therefore, it would be very helpful to have a system that automatically classifies news reports that reflect positively or negatively on a company's financial outlook. To accomplish this task, we treat the understanding of news articles as a text classification problem. In this paper, we propose a text classification method that we call, "Domain Experts" and "Self-Confident" sampling, and compare it with naive Bayes with expectation maximization (EM). We evaluate these learning techniques in terms of how well they improve with unlabeled data after being initially trained on a small number of human-labeled articles and how well they classify the latest financial news articles. The significance of this work lies in the new classification method that we propose and in the sampling technique we used for improving classification accuracy.

1 Introduction

In the application domain of stock portfolio management, there is a large volume of information about a company and its financial performance for humans to effectively attend to and manage while making decisions. To address this problem, we proposed a multi-agent system, called *Warren*[1] [4], [15] that helps the user track information on a portfolio of companies of interest. Warren is composed of different agents that help the user track the stock price, performance history, earnings summaries, and Beta value (risk) associated with the individual holdings in their stock portfolio, and to proactively advise the user whenever the portfolio may be too risky for the user's preferred tolerance to risk. To supplement the data on a company, the user has access to a *Breaking News agent*, which gathered financial news from on-line sources such as Reuters, CNN Financial Network, Business Wire, Forbes.com and others. In this paper, we describes our endeavor to create an agent that analyzes news articles that were retrieved by the Breaking News agent for their content about a company's financial well-being, for presentation to the user in a meaningful way.

To accomplish this task, we proposed a new text classification algorithm that classifies financial news into the predefined five classes: "good", "good, uncertain", "neutral", "bad, uncertain", and "bad" and a sampling algorithm that predicts the label of unlabeled data information-theoretically. Our hypotheses for how this goal can be achieved are as follows: (1) frequently

[1] The system is named after Warren Buffet, a famous American investor and author about investment strategies.

co-located phrases can permit the classifier to estimate the class of a news article with high probability; (2) the *domain experts* algorithm based on the voting process among frequently co-located phrases can perform well for this problem; (3) the performance of the method can be improved by using a sampling method which made use of *vote entropy* as the sampling criteria. Briefly, the proposed method predicts the label of a news articles through voting process among the frequently co-located phrases. In here, a frequently co-located phrase (FCP) is very often appeared in a certain class (e.g. good) and thus is considered to discriminate its class boundary well from others.

Our task is similar to the established approaches on the text classification domain in that it is a task to assign each given document to the appropriate class based on the semantic content of the document. Numerous statistical and machine learning methods have been applied to this domain in recent years including nearest neighbor classification [17], naive Bayes with EM (Expectation Maximization) [12] [14], Winnow with active learning [10], Support Vector Machines (SVMs) [6], [8], Maximum-Entropy model [13]. It is, however, slightly different with others in that our task deal with more objective and confined classes, such as "good" or "bad" for a company's financial outlook than classification of news articles into "politics" or "economics."

The text classification task has several characteristics that make it a difficult domain for the use of machine learning, including a very large number of input features, high levels of attribute and class noise, and a large percentage of features that are irrelevant. In particular, the exploitation of supervised learning requires a relatively large number of labeled examples. When it is given a small set of labeled training data, classification accuracy will suffer because the variance in the probability distribution of data will be high. However it is expensive to obtain labeled training data, while unlabeled data are cheaply available. Several methods have been used for coping with the problem which comes from insufficient labeled data, such as Expectation Maximization (EM) [12], [14], selective-sampling [2], sub-sampling and uncertainty sampling [9]. The proposed sampling method, self-confident sampling, picks out least uncertain data from unlabeled data sets in terms of entropy concept. It is similar to uncertainty-sampling in that it predicts the label of unlabeled data on the basis of the learner's confidence which is acquired during training phase. The examples that are predicted with the least uncertainty will be added to the training set in the next iteration. The overall procedure of self-confident sampling is described in Section 4.2.

The paper is organized as follows. Section 2 will give the overview of our task in terms of the text classification context. Section 3 details the method of selecting co-located phrases. In Section 4, we describes the procedure of classification with consideration of the company's financial well-being. Section 5 provides the experimental results and compares them with those of existing methods. Section 6 discuss the results and the future work respectively.

2 Classification of Financial News Articles

Concisely, our task is to build an algorithm that classifies each given news article into the predefined classes in terms of the referred company's financial well-being.

Most financial news articles that we gathered for experiments were manually labeled into 5 classes by considering how explicitly they mentioned the company's financial status which we want to know. The following five classes are considered to be pertinent by considering the nature of financial news articles:

```
[id] 000x-xx [\id]
[title] Goldman Profits Fall 13 Percent [\title]
[date] Mar 20 6:35 PM ET [\date]
[source] Reuters [\source]
[company] Goldman Sachs (GS) [\company]
[body]
Goldman Sachs Group Inc.(NYSE:GS - news), one of Wall Street's top firms,
on Tuesday said first-quarter profits fell 13 percent but were above reduced estimates
as fees for advising companies on stock sales declined in a slumping market.
...
The value of Goldman's principal investments fell $140 million, compared with a gain of
$214 million in last year's first quarter. Principal investments were down across the board,
Viniar said.
[\body]
[label] bad [\label]
```

Figure 1: A example of financial news article.

GOOD News scripts which show good evidences of the company's financial status explicitly.
e.g.) ... Shares of ABC Company rose 1/2 or 2 percent on the Nasdaq to $24-15/16. ...

GOOD, UNCERTAIN News scripts which refer to predictions of future profitability, and forecasts.
e.g.) ... ABC Company predicts fourth-quarter earnings will be high. ...

NEUTRAL News scripts which did not mention anything about the financial well-being of the company explicitly.
e.g.) ... ABC and XYZ Inc. announced plans to develop an industry initiative. ...

BAD, UNCERTAIN News scripts which refer to predictions of future losses, or no profitability.
e.g.) ... ABC (Nasdaq: ABC) warned on Tuesday that Fourth-quarter results could fall short of expectations. ...

BAD News scripts which show bad evidences of the company's financial status explicitly.
e.g.) ... Shares of ABC (ABC: down $0.54 to $49.37) fell in early New York trading. ...

Any news articles that do not mention financial clues of a company explicitly were classified into "neutral" class because we could not determine its current financial status. In case of "uncertain" classes ("good, uncertain" and "bad, uncertain", one may be allowed to decide it as a good (or bad) news for the company, but we could not be sure of it (i.e. uncertain.) The prediction of future earning is the very example of these classes: "good, uncertain" and "bad, uncertain." Figure 1 shows an example of news article used for our experiments.

3 Co-located Phrase as A Domain Expert

The proposed algorithm predicts the label of a financial news articles by means of voting process among selected frequently co-located phrases (FCP). A FCP is frequently occurred in a class and sequence of nearby, but not necessarily consecutive words. It is often desirable to consider such a contextual information (i.e. word-collocation) with respect to the characteristics of English text [11]. Thus, we believed that a FCP can discriminate the class of financial news articles well because it is strongly correlated with its class. For example, *Shares* and *rose* can be selected as a FCP for "good" class from the sentence, which is "Shares of Company A rose 1/2 or 2 percent on the Nasdaq to $24-15/16..."

It, however, is not easy to select such a phrase which express a class well due to the inherent properties of text classification, such as large features and much noise. To overcome this problem, we made use of a heuristic that considers the characteristic of financial news report. Since most of financial news articles report several company's stories in a news article, they mentioned a company's name (or a company's ticker[2]) explicitly. From this observation, we built an abridged version of a news article. That is, if a sentence contains a company's name or ticker, it is added to the abridged version from the original news. Indeed, an abridged version of article still has noise, but it also has sufficient information that make a classifier learn the correlation between phrases and a class. Most phrases in abridge version are the candidate of co-located phrases to a class after removing stop-words. To select the most informative feature, we taken advantage of information gain as the phrase goodness criterion [16]. Let $\{c_j\}_{j=1}^{m}$ denote the set of classes in the target space. The information gain of kth co-located phrase in jth class, $fcp_{k,j}$ is defined to be:

$$\begin{aligned} Gain(\mathrm{fcp}_{k,j}) = -\sum_{j=1}^{m} P(c_j)\log P(c_j) \\ +P(\mathrm{fcp}_{k,j})\sum_{j=1}^{m} P(c_i|\mathrm{fcp}_{k,j})\log P(c_j|\mathrm{fcp}_{k,j}) \\ +P(\bar{\mathrm{fcp}}_{k,j})\sum_{j=1}^{m} P(c_j|\bar{\mathrm{fcp}}_{k,j})\log P(c_j|\bar{\mathrm{fcp}}_{k,j}) \end{aligned} \tag{1}$$

Equation 1 was applied to estimate the importance of each FCP candidate, which is made by combining each word in condensed version and five consecutive words toward the end of a sentence. One of the five FCP candidates which has the highest value of information gain is selected for a domain experts for a class. Table 1 shows the example of selected FCPs for each class.

4 Classifying News by Considering Financial Status

To begin with, let us describe text representation. As mentioned earlier, a news article is divided into two group of sentences: abridged text and others. The abridged news article of each news article is represented as a weight vector:

$$\vec{d_i} =< w_1, w_2, ..., w_k, ..., w_{|T|} >, \tag{2}$$

[2] A ticker is a symbol that usually is used for representing a company's name briefly in stock trade market.

class	selected FCP
$+$	"revenue rose", "exceeds expectations", "share rose", "rose profit"
$+/?$	"expect earnings", "forecasts earnings", "anticipate earnings"
$+/-$	"alliance company", "alliance corp", "introduce ", "announces product"
$-/?$	"warning profits", "short expectation", "warning earnings"
$-$	"share off", "share down", "profit decrease", "fall percent", "sales decrease"

Table 1: Examples of selected frequently co-located phrases (FCP) for each class. Each symbol at leftmost column stands for five classes: "good", "good, uncertain", "neutral", "bad, uncertain", and "bad" as +, +?, +/-, -?, and -, respectively.

where w_k is the weight of k th word (or phrase) in i th document vector, d_i, which is made up of $|T|$ number of weights.

4.1 Naive Bayes Classifier with EM

One of the popular classifiers is the naive Bayes classifier. It learns from training data the conditional probability of each attribute w_k given the class c_j. Classification is done by applying Bayes rule to compute the probability of c_j given the particular instance of $w_1, ..., w_T$, and then predicting the class with the highest posterior probability. In equation 3, $freq_{k,j}$ is the number of times word t_k is occurred, $|J|$ is the total number of unique words in class j, and $|V|$ is the total number of unique words in data set. This computation is possible by making a strong independence assumption: all the attributes w_k are conditionally independent given the value of class c_j. The naive Bayes classifier estimates the probability that a new article is a member of a certain class using the probabilities of terms occurring in the class.

$$\begin{aligned} Pr(c_j|d_i) &= \arg\max_{c_j \in C} Pr(c_j) \prod_k Pr(w_k|c_j) \\ &\approx \frac{Pr(c_j) \prod_{k=1} Pr(w_k|c_j)}{\sum_{j=1}^{|C|} Pr(c_j) \prod_{k=1} Pr(w_k|c_j)} \\ & where, \\ w_k &= \frac{1 + freq_{k,j}}{|J| + |V|} \end{aligned} \tag{3}$$

The problem of applying naive Bayes classification to the real world problem is that the its performance could be decreased by variance from training data. In other words, when it is given a small set of labeled training data, the accuracy of classification will suffer because variance in the probability distribution of data would be high. However, it is expensive to acquire a sufficient number of labeled data for training. In [12], they tried to decrease a variance in classifying unseen data by a combination of a variant of Active Learning and Expectation Maximization (EM). Active learning is used to actively select documents for labeling, then EM with a naive Bayes model further improves classification accuracy. EM is a class of iterative algorithms for maximum likelihood estimation in problems with incomplete data. Given a model of data generation, and data with some missing values, EM will converge to a set of generative parameters that locally maximizes the likelihood of both the labeled and unlabeled data. In

Input:

Training data $D \leftarrow D_l$, where, D_l: labeled news articles

Feature Selection

Choose the co-located phrases among the current labeled news articles.

Do while performance of classifier improve

1. Generate unlabeled data D_u, where, D_u: unlabeled news articles
   Pick news articles randomly from unlabeled data set
2. Use the current classifier and feature sets to estimate the class of each unlabeled article D_u.

   - $\begin{cases} D_u \leftarrow D_u + \lambda d_i & \text{if } d_i > \theta_{ve} \\ \text{skip } d_i \text{ to the next iteration} & \text{otherwise} \end{cases}$

   - $D \leftarrow D_l + D_u$
3. Re-select feature sets by new data set D.

Output:

A classifier that takes an unlabeled articles and predicts a class label.

Figure 2: The self-confident sampling for domain experts algorithm.

this paper, EM is applied for an iterative two-step process: E-step calculates probabilistic class label of a document, $Pr(c_j|d_i)$, for given a set of unlabeled document using a current estimate of a class $Pr(c_j)$ and M-step calculates a new maximum likelihood estimates for $Pr(c_j)$ using all the labeled data, both original and probabilistically labeled.

4.2 Domain Experts with Self-confident Sampling

Figure 3 describes the predictions procedure of domain experts in detail. A group of domain experts consists of $|K|$, which is determined in experiment empirically, FCPs and represents a class. When it make prediction of a news article, it make use of voting among the group of domain experts and then learns by altering the weight associated with each domain expert. One attractive property of the proposed algorithm is that it is able to accommodate inconsistent hypothesis as well as consistent hypothesis. In other words, it does not eliminate a hypothesis that is found to be inconsistent with some training documents, but rather reduces its weight with the degree of β. Since we made use of our own text data set, we can tell that there are little phrase which appears only a class. Domain experts algorithm is similar with Sleeping experts algorithm [1], in that they consider each of selected "phrases" as a consistent expert (or hypothesis) to the class. But Sleeping expert did not allow a classifier to have the inconsistent

$C = \{$"good", "good, uncertain", "neutral",
"bad, uncertain", "bad"$\}$
$fcp_{k,j}$ denotes k th expert in j th domain (or class).
$w_{k,j}$ denotes the weight associated with $fcp_{k,j}$.
E_i denotes the most probable class of i th document.

- For each training example $< d_i, c(d_i) >$
  - Initialize e_j to 0
  - For each domain expert $fcp_{k,j}$
    $e_j \leftarrow e_j + w_{k,j} \quad \text{if } fcp_{k,j} \in d_i$
  - Predict
    $E_i(d_i) = \arg\max_j \frac{e_j}{\sum_j \sum_k w_{k,j}}$
  - Update weight
    $w_{k,j} \leftarrow \beta w_{k,j} \quad \text{if } \ e_j \neq c(d_i) \text{ and } fcp_{k,j} \in e_j$

Figure 3: Domain-Experts algorithm.

hypotheses.

The self-confident sampling method which we have proposed in figure 2 shares a property of the uncertainty-sampling [9], in that it predicts the label of an unlabeled data on the basis of the learner's confidence which is obtained through the training phase. The examples that are labeled with the least uncertain will be added to the training set in the next iteration. Unlike uncertainty-sampling, our method rely only on the vote by each of member of domain experts group, which has knowledge induced from the labeled data. We, however, could not rely on its knowledge completely. In this regards, λ is introduced for regulating the degree of reliance on learner's experience. Empirically, the proposed sampling method shows the best performance at 70 % confidence.

The class uncertainty of an unlabeled news article is determined by the value of vote entropy. Vote entropy is the entropy of the class label distribution resulting from having each group member deterministically "vote" for its winning class [3]. Let $V(j)$ be the number of domain experts which are involved in 'voting' for d_i for the class j:

$$VE(d_i) = -\sum_{j}^{|C|} \frac{V(j)}{|K|} \log \frac{V(j)}{|K|} \quad \text{if } fcp_k \in d_i, \tag{4}$$

where $|K|$ is the number of domain experts which took parts in voting of ith data, d_i which is ith data from the unlabeled data set.

While the vote entropy is 0 if a number of domain experts participating in the vote belong to the same class, the vote entropy is 1 when the vote committee is consist of an equal number of each class. We found empirically that the vote entropy for a class assigned correctly was less than 0.25, whereas the average entropy for incorrectly classified data was greater than 0.7. This approach is similar with active learning in terms of voting among the constituent of a classifier.

Data	+	+?	+/–	–?	–	Total
Labeled	239	70	526	60	344	1239
Unlabeled	–	–	–	–	–	5000
Total	239	70	526	60	344	6239

Table 2: The number of news articles for each class. Relatively smaller data in "uncertain (+/? and −/?)" classes could be explained by the objective contents of financial news article in terms of "good" or "bad."

Our sampling method specifies a process that determines the label of a new article, whereas active learning guides the direction that focuses what kind of instance is useful for learning.

5 Experimental Results

In this section, we describe the experimental results of the proposed methods, as compared with conventional methods.

As mentioned earlier, experiments were performed using the text data which we had made by ourselves. The labeled financial news articles data set amounts to 1,239 financial news articles gathered from various electronic news sources: CNN Financial Network, Forbes, Reuters/Reuters Securities, NewsFactors, Motley Fool, CNet, ZDNet, Morningstar.com, Business Week, AP Financial, Business Wire, PR News Wire, and Associated Press. Table 2 describes the distributions of news articles for each class. The phenomenon that "neutral" class has more data than others could be explained by the fact that larger part of them did not mentioned anything about the company's financial well-being explicitly, but deal with general information about the company.

Experiments aimed to verify the proposed methods in terms of two performance criterion: how well it make use of unlabeled data for improving classification accuracy and how accurately it classifys the latest news articles into predefined classes. Firstly, we evaluated whether the proposed sampling method would improve classification performance rates better than those trained by conventional methods. The experiment was performed to show the performance of domain experts with self-confident sampling, naive Bayes with EM, domain experts with EM and naive Bayes with self-confident sampling. Through the experiments, about 25% of the labeled data was used for testing and the rest of labeled set were used to produce classifiers. A domain experts group is empirically made up of 200 FCP. After training phase, each methods was tested in terms of classification accuracy: the proportion of the number of news articles classified correctly to the number of total news articles used.

Figure 4 and 5 show results of testing the accuracy performance of each sampling method with different number of labeled data. Total 50 iterations were carried out for each method. At each iteration, 50 unlabeled news article were given to each methods. When 690 out of 1239 labeled data feeds on training, the performance of the proposed method, the combination of Domain Experts and Self-Confident Sampling, is going up until sampling of 1,750 unlabeled news articles, and shows the best performance on accuracy measure at the point. From this, we assumed that around 2,000 news articles allow us to make a classifier with 75 % accuracy because it seems to largely depend on the fact that most of news providers delivered financial news with a restricted vocabulary set. With self-confident sampling, 16% accuracy is improved

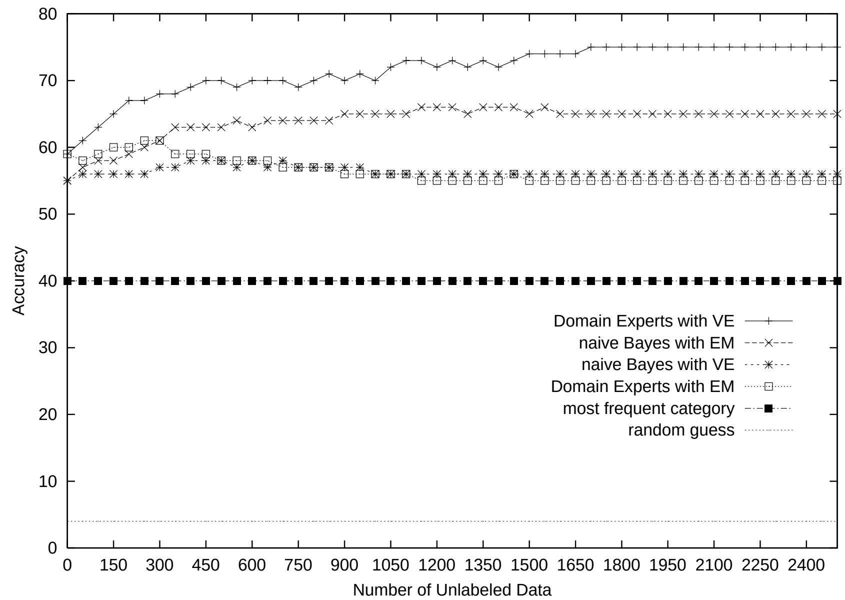

Figure 4: A result of sampling experiment was represented after training each of methods with 690 labeled data. "Most frequent class" and "random guess" are base line of performance. Since about 40% (526/1239) news articles out of labeled data are "neutral" class, we can assume that a method is able to gain 40% accuracy when it answers consistently the label of each news article in test set as "neutral."

with 56 % of labeled data (690/1239) and 35% of unlabeled data (1,750/5000) from the result in figure 4. As another goal of our task is to classify the label of the on-line financial news articles, the second experiment was performed to show the accuracy of the latest financial news data. A online data set is made up of the articles that gathered from the same news sources as the labeled data set and reports the latest financial news at the experimental time. At each trial, 30 news articles for a company was gathered from various news sources. However news article about a minor company could not meet the number of test documents at a trial (i.e. 30). The second row of Table 3 tells us the distribution of online test set. As a result, the proposed method has 79% averaged accuracy, which means 433 out of 549 total financial news articles were classified correctly. Table 3 shows the accuracy of tested methods per each class.

classes	$+$	$+/?$	$+/-$	$-/?$	$-$	total
articles	85	1	243	0	220	549
DE	.76	1	.8	–	.78	.79
naive	.61	0	.68	–	.62	.65

Table 3: Accuracy measure of each class to the online news data. Each column at the third and fourth row represents the accuracy of each class in terms of the proportion of the number of news article classified correctly to the total number of news article for the class. That the column about "good, uncertain $(+/?)$" class has the value of 1 means that only one news article which is labeled by human to that class was classified correctly. There is no news articles about "bad, uncertain $(-/?)$" class.

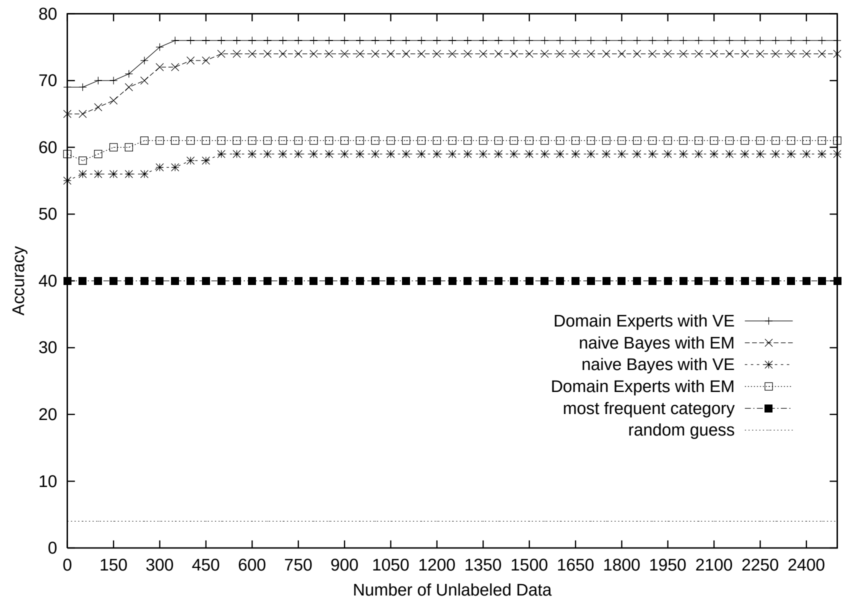

Figure 5: A result of sampling experiment was represented after training each of methods with 1239 labeled data.

6 Conclusions

We introduced an application of text classification that classifies financial news articles by considering referred company's financial well-being from their contents. The proposed algorithm which observed co-located phrase of a certain class from news contents and predicted the label with Weighted-Majority voting outperformed naive Bayes classifier about 14 %. For further improvement of accuracy, we proposed a sampling technique of which determine the class of an unlabeled news article with its entropy value. With the proposed sampling method, self-confident sampling, 16% accuracy is improved with 56 % of labeled data (690/1239) and 35% of unlabeled data (1,750/5000). The successful results from sampling test and online test supports that the proposed algorithms effectively works in this task, even though the promising results are partially come from the task characteristics of which its decision boundaries are relatively objective and are confined with a specific company's name.

But the proposed method has several weak points that prevent it from reaching the performance above 75 % accuracy. One is the difficulty in determining the label of news article of which made up of commensurate number of co-located phrases of each class. To illustrate, "Shares of company B rose 5 % in contrast with company A of which shares fall 7 %." In this example, domain experts may fail to predict "good" for company B. Because both phrases, which are "shares" with "rose" and "shares" with "fall", are very strong indicators of company's financial well-being at the moment, even though they did not indicate the same company and are not assigned with the same weight value during the training phase. As mentioned earlier, another weak point is that the proposed method does not consider the co-referred sentence. In other words, that it does not consider sentences, which did not mention company's name or ticker explicitly, as the financial evidence. For example, "Company C expects to boost revenue next quarter, Chief Operating Officer xxx said Wednesday. Despite of these anticipation, the company's shares fall again." In here, the prediction by the proposed method could be "good, uncertain", even though the true label of this example might be "bad" because "Company C" and "the company" are co-referred as the company's current financial well-being is not good.

To cope with these problems, we consider to employ several natural language processing techniques, such as the consideration of more wide range of a sentence and resolution of co-reference. For the purpose of verifying the applicability of proposed method, we also are about to try to apply the proposed method to the domains of which has similar characteristics to our task.

7 Acknowledgements

We would like to thank Massimo Paolucci and Sean Owens for their maintenance of the Warren financial portfolio agent system. This research has been sponsored in part by DARPA Grant F-30602-98-2-0138 and the Office of Naval Research Grant N-00014-96-16-1-1222.

References

[1] Cohen, W.W., and Singer, Y., Context-sensitive learning methods for text categorization, In *Proc. of Int'l ACM Conf. on Research and Development in Information Retrieval* (SIGIR-96), pp. 307-315, 1996

[2] Cohn, D.A., Atlas, L., and Ladner, R., Improving generalization with active learning, *Machine Learning*, Vol. 15, No. 2, pp. 201-221, 1994.

[3] Dagan, I., and Engelson, P., Committee-based sampling for training probabilistic classifiers, In *Proc. of Int'l Conf. on Machine Learning* (ICML-95), 1995.

[4] Decker, K., Sycara, K., Pannu, A., and Williamson, M., Designing behaviors for information agents, In *Proc. of Int'l Conf. on Autonomous Agents* (AA-97), pp. 404-413, 1997.

[5] Domingos, P., and Pazzani, M., Beyond independence: Conditions for the optimality of the simple Bayesian classifier, *Machine Learning*, Vol. 29, pp. 103-130, 1997.

[6] Dumais, S., Using SVMs for text categorization, *IEEE Intelligent Systems*, Vol. 13, No. 4, 1998.

[7] Freund, Y., Seung, S.H., Shamir, E., and Tishby, N.,Information, prediction, and Query-by-Committee algorithm, In *Proc. of Neural Information Processing Systems* (NIPS-92), pp. 483-490, 1992.

[8] Joachims, T., Text categorization with support vector machines: Learning with many relevant features, In *Proc. of European Conf. on Machine Learning*, pp. 137-142, 1998.

[9] Lewis, D.D., and Gale, W.,Training text classifiers by uncertainty sampling, In *Proc. of Int'l ACM Conf. on Research and Development in Information Retrieval* (SIGIR-92), pp. 3-12, 1994.

[10] Liere, R., and Tadepalli, P., Active learning with committees for text categorization, In *AAAI-98 Workshop on Learning for Text Categorization*, 1998.

[11] Manning, C.D., and Schutte, H., *Foundations of Statistical Natural Language Processing*, MIT Press, 1999.

[12] McCallum, A., and Nigam, K., Employing EM and pool-based active learning for text classification, In *Proc. of Int'l Conf. on Machine Learning* (ICML-98), pp. 359-367, 1998.

[13] Nigam, K., Lafferty, J., and McCallum, A., Using maximum entropy for text categorization, In *IJCAI-99 Workshop on Machine Learning for Information Filtering*, pp. 61-67, 1999.

[14] Nigam, K., McCallum, A., Thrun, S., and Mitchell, T., Text classification from labeled and unlabeled documents using EM, *Machine Learning*, Vol. 39, No. 2/3, pp. 103-134, 2000.

[15] Sycara, K., Decker, K., and Pannu, A., Distributed intelligent agents, *IEEE Expert*, 1996.

[16] Yang, Y., and Pedersen, J.O., A comparative study on feature selection in text categorization, In *Proc. of Int'l Conf. on Machine Learning* (ICML-97), pp. 412-420, 1997.

[17] Yang, Y., and Liu, X., A re-examination of text categorization methods, In *Proc. of Int'l ACM Conference on Research and Development in Information Retrieval* (SIGIR-99), pp. 42-49, 1999.