Concurrent Sentiment and Concept Extraction from Corporate Annual Reports for Financial Performance Forecasting

Petr HÁJEK, Vladimír OLEJ
Institute of System Engineering and Informatics
Faculty of Economics and Administration, University of Pardubice
Pardubice, 532 10, Czech Republic
petr.hajek@upce.cz, vladimir.olej@upce.cz

ABSTRACT

This paper aims to extract both sentiment and concept knowledge from a corpus. We extract linear combinations of features to examine the relations between sentiment categories and concepts obtained using singular value decomposition. We show that the features are related to current financial performance indicators. We use the extracted knowledge to compare decision trees algorithms (Best First Decision Tree, Alternating Decision Tree) on the one hand and Support Vector Machines and Multilayer Perceptron type Neural Networks on the other hand. The aim is to investigate the performance of document classification for a set of annual reports. We show that (1) Support Vector Machines provide the best accuracy in forecasting investment / non-investment grades, and (2) poorly performing companies frequently use commonality, human interest and negative terms and emphasize current business situation, without actively addressing problems associated with economic downturn.

Keywords: Annual Reports, Concept Extraction, Sentiment, Best First Decision Tree, Alternating Decision Tree, Support Vector Machines, Multilayer Perceptron Neural Networks.

1. INTRODUCTION

In traditional corporate financial forecasting models, the focus has always been on historical financial ratios of profitability, leverage, liquidity, etc. Over the past few years there has been a dramatic increase in interest in the role of textual information. Recent findings have shown that important information about both the past and future financial performance of companies can be extracted from textual documents such as news or reports. In particular, the narrative sections of the documents may serve as early warning signals for stakeholders. This is consistent with the growing share of qualitative business information compared with that of quantitative.

The research to date has tended to use the textual information to predict stock market return [11], volatility [21], bankruptcy [1] or credit ratings [10]. However, previous work has been focused on either word choice similarities [19] or sentiment analysis [10]. Most recently, attempts have been made to extract and visualize concepts from corporate documents [13]. However, far too little attention has been paid to examine the relations between sentiment categories and extracted concepts.

In this study, we used a set of corporate annual reports as a corpus. This is because they provide the best possible description of a company, its managerial priorities, business risks and policies. We hypothesize that annual reports differ in terms of the sentiment and concepts emphasized depending on future financial performance. In addition, we anticipate that concurrent extraction of quantitative and qualitative knowledge may reveal common factors of future financial performance.

We use the WordNet ontology to detect appropriate terms in annual reports. Thus, synonyms are detected for the corresponding domain. We further employ singular value decomposition (SVD) to extract the concepts from the set of candidate terms. We combine the concepts with two rule-based approaches that aim to extract sentiment information from the reports. Thus, both content and its sentiment can be combined with the information from financial statements in order to forecast financial performance in years to come. We perform the forecasting using several artificial intelligence classification algorithms. However, the choice of the most suitable algorithm becomes increasingly difficult. We employ three categories of artificial intelligence methods to examine the given aims, (1) Best First Decision Trees (BFDTs) and Alternating Decision Trees (ADTs), (2) Support Vector Machines (SVMs), and Multilayer Perceptron type Neural Networks (MLP NNs).

The remainder of this paper has been organized in the following way. The second section briefly reviews literature related to financial forecasting using textual analyses. The third section introduces the problem and data in more detail. The fourth section presents the results of the modeling and the final section concludes the paper and suggests future works.

2. FINANCIAL FORECASTING USING TEXTUAL ANALYSES

There are two general approaches to textual analysis reported in the literature, (1) word categorization (bag of words, rule-based approach) method and (2) statistical methods. The main problem of the former approach is that the available dictionary may be context (domain) sensitive, while the latter one requires the likelihood ratios to be estimated based on difficult to replicate and subjective classification of texts' tone [26]. A word classification scheme into positive and negative categories was used by Feldman et al. [6] to show that stock market reactions significantly respond to the tone change of the annual reports. Similar results have also been reported for statistical approaches such as Naïve Bayes classifier [4,19,20]. These studies suggest that the prediction of future stock return can be significantly improved using the information on sentiment categories used in corporate communication. This finding has been reported for both short- and long-term forecasting horizons [11].

Loughran and McDonald [21] developed a dictionary for financial domain as an alternative to traditional Harvard's General Inquirer. The dominance of the financial dictionary has been illustrated on several financial prediction problems including the forecasts of stock returns, trading volumes, return volatility, fraud, material weakness and unexpected earnings. This dictionary has recently been used to predict investment/non-investment rating grades [10,12] and distressed/non-distressed companies [14], respectively.

A word categorization approach was also used by Goel and Gangolly [9] to show that fraudulent financial reporting is reflected not only in a lower readability but also in a more positive, passive and uncertain tone of language. Chen et al. [2] finds that ambiguous and mild statements in the reporting year anticipate decreasing earnings, while assertive and positive statements show on the opposite.

The concepts were extracted from annual reports by Cecchini et al. [1] to discriminate between bankrupt/non-bankrupt companies. Similarly, Hajek and Olej [13] visualized the concepts using self-organizing maps to detect financially distressed companies. Other corporate documents have been examined recently (news stories [26], IPO prospectuses [16], earnings press releases [5]) to support the finding suggesting that qualitative verbal communication by managers contains important insider information on future financial performance.

3. SENTIMENT AND CONCEPT EXTRACTION

Fig. 1 depicts the block structure of sentiment and concept extraction.

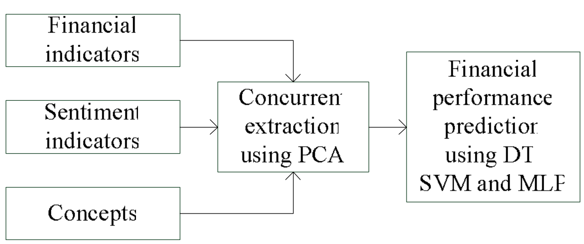

Fig. 1. Research methodology.

3.1 Financial Indicators

The prediction of corporate financial performance was realized as a two-class problem, where strong performance was represented by investment grade (IG) assignment and poor performance by non-investment grade (NG), respectively. This assignment was obtained from the Standard & Poor's rating agency in 2011. In addition to concurrent concept and sentiment information, we used previous year's financial indicators as predictors. These included size (enterprise value, revenues), profitability ratios (return on equity, enterprise value/earnings), liquidity (cash), leverage (market debt/total capital) and market value ratios (earnings per share, price to book value, price to earnings per share, high to low stock price, dividend yield, payout ratio, standard deviation of stock price). These indicators have shown promising predictive accuracy in previous studies [10,15]. The financial indicators were drawn from the Value Line database. After discarding rather specific mining and financial companies, the dataset covered 557 U.S. companies, 208 classified as IG and 349 as NG.

3.2 Set of Documents

The annual reports of corresponding companies were collected at the U.S. Securities and Exchange Commission EDGAR System.

For the corpus of documents, we set the following requirements: availability of financial indicators and the companies had to be listed on the NYSE or Nasdaq. In line with related studies, we downloaded all annual reports in10-Ks txt format (without amended documents) for the year 2010. Documents that only referred to other reports were withdrawn. Similarly, graphics, tables and SEC header were removed from the documents before text processing.

Concurrent sentiment and concept extraction required linguistic pre-processing (tokenization and lemmatization) and discarding both the stop-words and the least frequent words in the corpus. Next, the potential term candidates were compared with the WordNet ontology [22] to detect synonyms and the correct sense of the terms for the domain (those with the highest score for Economy, Commerce or Law domains were chosen), see [13] for details. The set of terms, comprising of 681 terms, was weighted using the *tf.idf* weighting scheme.

3.3 Concept Extraction

To capture the concepts, we performed SVD and kept those with singular values greater than 0. As a result, 27 basic concepts had to be labeled based on the term importance. The detected concepts covered the following broader topics: legal proceedings, financial and investment policy, cooperation and partnership, domestic and foreign markets, economic downturn, business goals and policy, organization and management, corporate restructuring and asset policy.

3.4 Sentiment Extraction

Sentiment extraction, on the other hand, was performed comparing the term candidates with two dictionaries: 1) financial dictionary developed by Loughran and McDonald [21]; and (2) Diction 5.0 [17]. We expected that combining the detected concepts with other linguistic attributes of the annual reports (optimism, uncertainty, etc.) improves the precision of the forecasting model. The financial dictionary covered the following categories of terms: negative, positive, uncertainty, litigious, modal strong and modal weak. Negation was detected to determine the frequency of positive words more accurately. The Diction 5.0 offers 31 additional categories such as ambivalence, self-reference, satisfaction, inspiration or concreteness. Again, the *tf.idf* term weighting scheme was used to obtain average importance of each category.

3.5 Principal Component Extraction

In the next step, we performed Principal Component Analysis (PCA) to obtain the linear combinations of both the quantitative financial indicators and qualitative linguistic attributes. Thus, the dimensionality was reduced from 77 (13 + 27 + 6 + 31) to 23. Only those components were drawn for which Eigenvalue was greater than 1. In total, the components explained 71.3 % of the variance in the data. Table 1 shows that significant differences were observed for x_1 – corporate restructuring, x_2 – financial market performance, x_4 – active investment policy, x_5 – profit, x_7 – liquidity, x_9 – growth and x_{19} – business situation and financial policy, suggesting that the use of financial indicators is critical for accurate financial performance prediction. However, the concepts related to corporate restructuring and, respectively, business situation and financial policy also provide relevant information about the future financial performance. More specifically, poorly performing companies emphasized the concept of restructuring together with centrality, accomplishment, passivity and cooperation terms. Further, active investment policy was accompanied with less negative terms. Similarly, companies with a high liquidity used more self-reference terms, and growth companies used human interest terms less frequently. We also performed Student's *t*-test for multiple variables, indicating that the classes were significantly different at p=0.000.

Table 1. Descriptive statistics of extracted components.

component	IG	NG	p-value
x_1	-0.363	0.216	0.100*
x_2	1.788	-1.066	0.000***
x_3	0.119	-0.071	0.218
x_4	0.223	-0.133	0.015**
x_5	0.197	-0.117	0.017**
x_6	0.038	-0.023	0.645
x_7	-0.219	0.130	0.004***
x_8	-0.004	0.002	0.959
x_9	-0.169	0.101	0.023**
x_{10}	-0.015	0.009	0.828
x_{11}	-0.046	0.028	0.487
x_{12}	0.070	-0.041	0.277
x_{13}	-0.100	0.059	0.115
x_{14}	-0.004	0.003	0.944
x_{15}	0.006	-0.003	0.924
x_{16}	-0.024	0.014	0.696
x_{17}	0.065	-0.039	0.282
x_{18}	-0.030	0.018	0.620
x_{19}	-0.228	0.136	0.000***
x_{20}	0.041	-0.025	0.477
x_{21}	-0.045	0.027	0.438
x_{22}	0.013	-0.007	0.823
x_{23}	-0.008	0.005	0.890

Legend: * significant at p=0.1, ** significant at p=0.05, *** significant at p=0.01.

The list of extracted components (+ 5 attributes with the highest weight):

x_1 = 0.206centrality + 0.206accomplishment + 0.195corporate restructuring + 0.194passivity + 0.193cooperation, (1)
label: *corporate restructuring*;

x_2 = - 0.352high low stock price - 0.343std. dev. of stock price + 0.333enterprise value + 0.31revenue - 0.284market debt to capital, (2)
label: *financial market performance*;

x_3 = 0.31weak modal + 0.307litigious + 0.279uncertainty - 0.258human interest + 0.246financial restructuring, (3)
label: *financial restructuring*;

x_4 = 0.362dividend yield - 0.331negative + 0.302payout ratio + 0.275investment policy - 0.258relation to environment, (4)
label: *active investment policy*;

x_5 = 0.343return on equity + 0.317price to book value + 0.233 legal proceedings implications - 0.225 price to earnings - 0.221revenue, (5)
label: *profit*;

x_6 = - 0.374market debt to capital + 0.23organization and management - 0.225litigious - 0.202legal proceedings implications + 0.185self-reference, (6)
label: *equity*;

x_7 = 0.367legal proceedings - 0.311relation to environment + 0.255self-reference + 0.225cash - 0.225financial restructuring, label: *liquidity*; (7)

x_8 = 0.311positive + 0.284financial cooperation and partnership + 0.254reclassification of tangible assets - 0.22legal proceedings implications + 0.213payout ratio, (8)
label: *payout*;

x_9 = 0.547price to book value + 0.52return on equity - 0.17human interest + 0.161foreign markets + 0.16price to earnings, (9)
label: *growth*;

x_{10} = 0.34new market opportunities - 0.288concreteness - 0.255economic downturn impacts - 0.217investment policy - 0.217human interest, (10)
label: *new market*;

x_{11} = 0.415restructuring strategy - 0.295dealing economic downturn - 0.279e-commerce - 0.278motion + 0.274 reclassification of financial assets, (11)
label: *passive restructuring strategy*;

x_{12} = 0.444domestic market difficulties + 0.325enterprise value to earnings - 0.32foreign markets - 0.275technological change + 0.255business policy, (12)
label: *business policy for domestic market*;

x_{13} = 0.284e-commerce - 0.255technological change - 0.237 reclassification of tangible assets + 0.237reclassification of financial assets + 0.223foreign markets, (13)
label: *foreign markets*;

x_{14} = - 0.439business policy - 0.347technological change - 0.336 foreign markets + 0.285financial cooperation and partnership - 0.262financial risk, (14)
label: *financial cooperation and partnership*;

x_{15} = 0.345financial accounting principles - 0.333production policy - 0.285business policy - 0.264enterprise value to earnings - 0.232economic downturn, (15)
label: *financial accounting*;

x_{16} = 0.491business goals - 0.338financial accounting principles - 0.317production policy - 0.225tangible assets+ 0.223 restructuring strategy, (16)
label: *business goals and restructuring strategy*;

x_{17} = - 0.496financial policy + 0.368enterprise value to earnings + 0.364financial cooperation and partnership + 0.272business situation + 0.218foreign markets, (17)
label: *current business situation*;

x_{18} = 0.482debt policy + 0.358production policy + 0.316 tangible assets + 0.243business goals - 0.227technological change, (18)
label: *debt and production policy*;

x_{19} = 0.477business situation - 0.268financial assets - 0.262 dealing economic downturn + 0.26financial policy + 0.257 enterprise value to earnings, (19)
label: *business situation and financial policy*;

x_{20} = 0.384restructuring strategy + 0.37economic downturn - 0.31financial cooperation partnership + 0.265new markets + 0.244dealing economic downturn, (20)
label: *economic downturn*;

x_{21} = 0.586financial risk - 0.327business goals - 0.312financial assets + 0.239organization management - 0.233technological change, (21)
label: *financial risk*;

x_{22} = 0.459financial accounting principles + 0.373debt policy - 0.327tangible assets + 0.291dealing economic downturn + 0.252tenacity, (22)
label: *financial accounting connoting confidence and totality*;

x_{23} = - 0.367debt policy - 0.356financial policy - 0.346financial assets + 0.302production policy + 0.279business policy, (23)
label: *production and business policy*.

4. MODELING AND ANALYSIS OF THE RESULTS

To predict financial performance using the extracted components, we employed commonly used BFDTs and ADTs classification algorithms. Furthermore, these classification algorithms were compared with SVMs and MLP NNs. In order to avoid over-fitting, we carried out experiments with different values of BFDTs, ADTs, SVMs, and MLP NNs parameters using 10-fold cross-validation. The measures of classification performance are represented by the averages of standard statistics applied in classification tasks [24]: true positives (TP rate), false positives (FP rate), precision (Pre) and recall (Re), F-measure (F-m), and the area under the receiver operating characteristic (ROC) curve. F-m is the weighted harmonic mean of Pre and Re, or the Matthews correlation coefficient, which is a geometric mean of the chance-corrected variants. A ROC is a graphical plot which illustrates the performance of a binary classifier system, which represents a standard technique for summarization classifier performance over a range of tradeoffs between TP and FP error rates.

4.1 Decision Trees

Standard decision tree learners such as expand nodes in depth first order, while in BFDT [25] learners the best node is expanded first. The best node is the node whose split leads to maximum reduction of impurity among all nodes available for splitting. The resulting tree will be the same when fully grown, just the order in which it is built is different. In practice, some branches of a fully expanded tree do not truly reflect the underlying information in the domain. This problem is known as over fitting and is mainly caused by noisy data. Pruning is necessary to avoid over fitting the training data, and discards those parts that are not predictive of future data. Best first node expansion enables us to investigate new pruning techniques by determining the number of expansions performed based on cross-validation. The BFDTs was tested for minimal number of instances at the terminal nodes = {1,2, … ,50}, the number of folds in internal cross-validation = {1,2, … ,5}, pruned strategy, random number seed {1,2, … ,5}, and percentage of the training set size {1,2, … ,5}. The following parameters of the BFDTs were examined to obtain the best classification performance:

– the number of instances at the terminal nodes = 30,
– the number of folds in internal cross-validation = 2,
– pruning strategy = pruned,
– random number seed = 1,
– percentage of the training set size=1.

ADTs [7,8], a natural extension of both voted-stumps and decision trees, consist of alternating layers of prediction and decision nodes. The structure of an ADT represents decision paths; when a path reaches a decision node, it continues with the specific offspring node that corresponds to the decision outcome as in standard decision tree. On the other hand, when a path reaches a prediction node, the path continues with all of the offspring nodes. Thus the classification rule that it represents is basically a weighted majority vote over base prediction rules. ADT can resolve the problem of over-fitting which occurs when limited amount of data is available. ADTs are a generalization of weighted sums of decision trees. The ADTs was tested for the number of boosting iterations = {1,2,5,10, … ,200}, and random number seed = {1,2, … ,5}. The following parameters of the ADTs were examined to obtain the best classification performance:

– the number of boosting iterations = 100,
– random number seed = 3.

The best classification results for the BFDTs and ADTs simulations (average of 10-fold cross validation) are shown in Table 2.

Table 2. Best results for BFDTs and ADTs.

	BFDT		ADT	
Accuracy [%]	88.51		88.15	
Class	IG	NG	IG	NG
TP rate	0.851	0.905	0.827	0.914
FP rate	0.095	0.149	0.086	0.173
Precision	0.843	0.911	0.851	0.889
Recall	0.851	0.905	0.827	0.914
F-m	0.847	0.908	0.839	0.906
ROC	0.896	0.896	0.916	0.916

4.2 Support Vector Machines and Multilayer Perceptron Type Neural Networks

The output $f(\mathbf{x}_t)$ of SVMs is [3,23] defined this way

$$f(\mathbf{x}_t) = \sum_{i=1}^{N} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}_t) + b, \tag{24}$$

where $\mathbf{x}_t$ is the evaluated pattern, N is the number of support vectors, $\mathbf{x}_i$ are support vectors, α_i are Lagrange multipliers determined in the optimization process, k is the actual kernel function $k(\mathbf{x},\mathbf{x}_i)$. Given some training data D, a set of n points of the form $D=\{(\mathbf{x}_i,y_i)| \mathbf{x}_i \in R^p, y_i \in \{-1,1\}\}_{i=1}^{n}$ where the y_i is either 1 or −1, indicating the class to which the point $\mathbf{x}_i$ belongs. For standard SVM problem, the smallest possible optimization involves two Lagrange multipliers because the Lagrange multipliers must obey a linear equality constraint. At every step, the SVM trained by the Sequential Minimal Optimization (SMO) [23] chooses two Lagrange multipliers to jointly optimize, finds the optimal values for these multipliers, and updates the SVM to reflect new optimal values. The SVM trained by the SMO was tested for linear, polynomial (exponent = {1,2,3,4}) and RBF kernel functions γ = {0.1, 0.001, 0.0001} with the complexity parameter C = {1,2, … ,100}. The following parameters of the SVMs were examined to obtain the best classification performance:

- polynomial kernel functions, exponent = 2,
- complexity parameter = 20,
- round-off error ε = 1.0E-12,
- tolerance parameter = 0.001.

The j-th output $f_j(\mathbf{x},d,\mathbf{w})$ MLP NNs [18] can be expressed for example as follows

$$f_j(\mathbf{x},d,\mathbf{w}) = \sum_{k=1}^{K} \mathbf{v}_k (d(\sum_{j=1}^{J} \mathbf{w}_{j,k}\mathbf{x}_{j,k})), \quad (25)$$

where $\mathbf{v}_k$ is vector of synapses' weights among neurons in hidden layer and output neuron, $\mathbf{w}_{j,k}$ is the vector of synapses' weights among input neurons and neurons in hidden layer, k is the index of neuron in hidden layer, K is the number of neurons in hidden layer, d is the activation function, j is the index of the input neuron, J is the number of the input neurons per one neuron in hidden layer, and $\mathbf{x}_{j,k}$ is the input vector of MLP NN. The MLP NN was trained using the back propagation algorithm with momentum. The following parameters of the MLP NN were set and examined: the number of neurons in the hidden layer = {5,10, … ,50}, learning rate = {0.05, 0.1, 0.2, 0.3}, momentum = {0.05, 0.1, 0.2, 0.3}, and the number of epochs = {5,10, ... ,200}. The following parameters of the MLP NNs were examined to obtain the best classification performance:

- the number of neurons in the hidden layer = a,
- learning rate = 0.4,
- momentum = 0.05,
- the number of epochs = 30.

The best classification results for the SVMs and MLP NNs simulations (average of 10-fold cross validation) are shown in Table 3.

Table 3. Best results for SVMs and MLP NNs.

	SVM		MLP NN	
Accuracy [%]	90.13		89.95	
Class	IG	NG	IG	NG
TP rate	0.851	0.931	0.875	0.914
FP rate	0.069	0.149	0.086	0.125
Precision	0.881	0.913	0.858	0.925
Recall	0.851	0.931	0.875	0.914
F-m	0.866	0.922	0.867	0.919
ROC	0.891	0.891	0.940	0.940

To compare the performance of the classifiers, we used paired t-test (Table 4). SVM and MLP NN performed significantly better than decision trees in terms of classification accuracy and F-measure, respectively. On the other hand, ADT provided statistically similar performance with that of MLP NN regarding the ROC measure. Nevertheless, the FP rate on IG class becomes critical to investors due to a significantly higher credit risk assigned to NG class. In that case, SVM performed significantly better than the remainder of the classifiers.

Table 4. Performance comparison across classifiers.

	BFDT	ADT	SVM	MLP NN
Acc.	88.51±4.35	88.15±4.45	90.13±4.02*	89.95±4.87*
FP_{IG}	0.10±0.03	0.09±0.03	0.07±0.02*	0.09±0.03
F-m	0.90±0.04	0.90±0.04	0.92±0.03*	0.91±0.04*
ROC	0.90±0.05	0.92±0.04*	0.89±0.05	0.94±0.03*

* significantly higher at p<0.05

To determine the effect of the extracted components on classification performance, we calculated the average relative importance of each attribute by adding up the improvement in classification accuracy gained by using the corresponding predictor. The importance ranged from 0 (no improvement) to 1 (the most important component). The importance between 0 and 1 is relative to the most important predictor. Table 5 shows only those predictors that increased classification accuracy (importance > 0) of the SVM models.

The results corroborate the tests performed in Table 1 and suggest that financial market performance (x_2) is by far the most important predictor in the SVM model. However, companies using excessively terms related to current business situation and financial policy (x_{19}) also indicate financial difficulties.

Table 5 Average importance (AI) of extracted components (EC) in SVM.

EC	x_2	x_{19}	x_7	x_5	x_1
AI	1.00	0.114	0.045	0.034	0.028
EC	x_9	x_4	x_{16}	x_{15}	x_6
AI	0.017	0.011	0.011	0.011	0.006

5. CONCLUSIONS

The present study was designed to determine the effect of concurrent sentiment and concept extraction on the accuracy of financial performance forecasting. Taken together, the evidence from this study suggests that, although financial indicators incorporate the most important information, concurrent sentiment and concept extraction from annual reports may be used as an important predictor of corporate financial performance. Specifically, the results indicate that poorly performing companies emphasized current business situation together with financial policy tools, without actively addressing problems associated with economic downturn. Additionally, financial distress may be indicated by a frequent use of commonality, human interest and negative terms. The best classification accuracy obtained using SVM was 90.13 %, performing significantly better for IG class in particular.

The current findings suggest a key role for financial market determinants. Sentiment and concept analysis seems to be related to financial ratios such as profitability and liquidity in particular. However, the current study was unable to analyze the frequency of the sentiment collocates near the key concepts in the reports. We recommend that for further research. It would be also interesting to assess the effect of concurrent sentiment and concept extraction on related future financial distress indicators such as bankruptcy, profitability and stock price movements.

The experiments in this study were carried out in Statistica 10 (linguistic preprocessing) and Weka 3.7.5 in MS Windows 7 operation system.

Acknowledgment: This work was supported by the scientific research project of the Czech Sciences Foundation Grant No: 13-10331S.

6. REFERENCES

[1] M. Cecchini, H. Aytug, G.J. Koehler, P. Pathak, "Making Words Work: Using Financial Text as a Predictor of Financial Events", **Decision Support Systems**, Vol. 50, No. 1, 2010, pp. 164-175.

[2] C.L. Chen, C.L. Liu, Y.C. Chang, H.P. Tsai, "Opinion Mining for Relating Subjective Expressions and Annual Earnings in US Financial Statements", **Journal of Information Science and Engineering**, Vol. 29, No. 3, 2013, pp. 1-24.

[3] N. Cristianini, J. Shawe-Taylor, **An Introduction to Support Vector Machines and other Kernel-based Learning Methods**, Cambridge: Cambridge University Press, 2000.

[4] S. Das, M. Chen, **Yahoo! for Amazon: Opinion Extraction from Small Talk on the Web**, Santa Clara University, 2001.

[5] E.A. Demers, C. Vega, **Soft Information in Earnings Announcements: News or Noise?**, Working paper, INSEAD, 2010.

[6] R. Feldman, S. Govindaraj, J. Livnar, B. Segal, "Management's Tone Change, Post Earnings Announcement Drift and Accruals", **Review of Accounting Studies**, Vol. 15, 2010, pp. 915-953.

[7] J. Friedman, T. Hastie, R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting", **Annals of Statistics**, Vol. 28, No. 2, 2000, pp. 337-407.

[8] Y. Freund, L. Mason, "The Alternating Decision Tree Learning Algorithm", **16th International Conference on Machine Learning**, Bled, Slovenia, 1999, pp. 124-133.

[9] S. Goel, J. Gangolly, "Beyond the Numbers: Mining the Annual Reports for Hidden Cues Indicative of Financial Statement Fraud", **Intelligent Systems in Accounting, Finance and Management**, Vol. 19, No. 2, 2012, pp. 75-89.

[10] P. Hajek, V. Olej, "Evaluating Sentiment in Annual Reports for Financial Distress Prediction Using Neural Networks and Support Vector Machines", **Communications in Computer and Information Science**, Vol. 384, 2013, pp. 1-10.

[11] P. Hajek, V. Olej, R. Myskova, "Forecasting Stock Prices using Sentiment Information in Annual Reports - A Neural Network and Support Vector Regression Approach", **WSEAS Transactions on Systems**, Vol. 10, No. 4, 2013, pp. 293-305.

[12] P. Hajek, V. Olej, "Predicting Firms' Credit Ratings Using Ensembles of Artificial Immune Systems and Machine Learning - An Over-Sampling Approach", **IFIP Advances in Information and Communication Technology**, Vol. 436, 2014, pp. 29-38.

[13] P. Hajek, V. Olej, "Comparing Corporate Financial Performance and Qualitative Information from Annual Reports using Self-organizing Maps", **10th International Conference on Natural Computation** (ICNC'14), Xiamen, China, 2014, pp. 93-98.

[14] P. Hajek, V. Olej, R. Myskova, "Forecasting Corporate Financial Performance using Sentiment in Annual Reports for Stakeholders' Decision-Making", **Technological and Economic Development of Economy**, 2014, DOI: 10.3846/20294913.2014.979456, in press.

[15] P. Hajek, K. Michalak, "Feature Selection in Corporate Credit Rating Prediction", **Knowledge-Based Systems**, Vol. 51, 2013, pp. 72-84.

[16] K.W. Hanley, G. Hoberg, "The Information Content of IPO Prospectuses", **Review of Financial Studies**, Vol. 23, No. 7, 2010, pp. 2821-2864.

[17] R.P. Hart, "Redeveloping DICTION: Theoretical Considerations", M.D. West (Ed.), **Theory, Method, and Practice in Computer Content Analysis**, 2001, pp. 43-60.

[18] S. Haykin, **Neural Networks: A Comprehensive Foundation**, 2nd edition, New Jersey: Prentice-Hall Inc., 1999.

[19] A. Huang, A. Zang, R. Zheng, **Informativeness of Text in Analyst Reports: A Naïve Bayesian Machine Learning Approach**, Working paper, The Hong Kong University of Science and Technology, 2010.

[20] F. Li, "The Information Content of Forward-Looking Statements in Corporate Filings - A Naïve Bayesian Machine Learning Approach", **Journal of Accounting Research**, Vol. 48, No. 5, 2010, pp. 1049-1102.

[21] T. Loughran, B. McDonald, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks", **The Journal of Finance**, Vol. 66, No. 1, 2011, pp. 35-65.

[22] G.A. Miller, "WordNet: A Lexical Database for English", **Communications of the ACM**, Vol. 38, No. 11, 1995, pp. 39-41.

[23] J.C. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization", B. Schoelkopf, C. Burges, A. Smola (Eds.), **Advances in Kernel Methods - Support Vector Learning**, MIT Press, 1998.

[24] D.M.W. Powers, "Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation", **Journal of Machine Learning Technologies**, Vol. 1, No. 2, 2011, pp. 37-63.

[25] H. Shi, **Best-First Decision Tree Learning**, Department of Computer Sciences, The University of Waikato, Hamilton, New Zeland, 2007.

[26] P.C. Tetlock, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", **Journal of Finance**, Vol. 62, 2007, pp. 1139-1168.