Towards Re-defining Relation Understanding in Financial Domain

Chenguang Wang
IBM Research-Almaden
chenguang.wang@ibm.com

Doug Burdick
IBM Research-Almaden
drburdic@us.ibm.com

Laura Chiticariu
IBM Research-Almaden
chiti@us.ibm.com

Rajasekar Krishnamurthy
IBM Research-Almaden
rajase@us.ibm.com

Yunyao Li
IBM Research-Almaden
yunyaoli@us.ibm.com

Huaiyu Zhu
IBM Research-Almaden
huaiyu@us.ibm.com

ABSTRACT

We describe our experiences in participating in the scored task for the 2017 FEIII Data Challenge. Our approach is to model the problem as a binary classification problem and train an ensemble model leveraging domain features that capture financial terminology. We share challenge results for our submission, which performed well achieving the highest score in four out of six evaluation criteria. We describe semantic complexities encountered with regards to the task definition and ambiguities in the labeled dataset. We present an alternative task formulation *Relationship Validation* that addresses some of these semantic complexities and demonstrate how our approach naturally extends to this simplified task definition.

KEYWORDS

Relation Understanding, Financial Domain, FEIII, Information Extraction, Text Classification

ACM Reference format:
Chenguang Wang, Doug Burdick, Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li, and Huaiyu Zhu. 2017. Towards Re-defining Relation Understanding in Financial Domain. In *Proceedings of DSMM'17, Chicago, IL, USA, May 14, 2017,* 6 pages.
DOI: http://dx.doi.org/10.1145/3077240.3077254

1 INTRODUCTION

The 2017 FEIII Data Challenge [3, 6] focuses on understanding the relationships among financial entities and the roles they play in financial contracts. The scored task is to evaluate whether textual passages (one or more sentences) in financial regulatory filings are relevant, interesting and validate a specific relationship between the filing financial entity and another financial entity mentioned in the text. In this paper, we describe our experiences from participating in this challenge.

We first describe how we modeled the problem as a classification task and developed an ensemble model that leverages domain features capturing financial terminology (Section 2). This approach provided good results on the challenge evaluation criteria, achieving the highest score on four out of six evaluation criteria (Section 3).

As we investigated our model performance over the challenge training dataset, we realized that the challenge task definition is complex and describe several complexities we observed in the labeled data (Section 4). These observations illustrate challenges that need to be addressed in general while defining semantic tasks that require deep domain knowledge.

One specific outcome of this analysis was our understanding of how the FEIII 2017 Challenge task attempts to achieve two objectives at the same time: (a) validate financial relationships and (b) identify interesting and relevant knowledge about financial entities. We define a simplified version of the task, *Relationship Validation* (*Rel_Val*), focused on (a), i.e., validating financial relationships and an associated labeling scheme (Section 5). We show how the approach outlined in Section 2 performs well under this refined task definition.

We conclude the paper with lessons learned and recommendations for the next iteration of the challenge (Section 6).

2 OUR APPROACH: ENSEMBLE MODEL LEVERAGING DOMAIN FEATURES

The challenge dataset consists of text from regulatory filings (*Context*) that potentially describe financial relationships. Each *Context* text is associated with a relationship triple (*Filing Entity*, *Role*, *Mentioned Entity*) and the objective is to evaluate whether the triple is validated by the *Context* text. The training dataset associates each *Context* text with one of four labels: *Highly Relevant*, *Relevant*, *Neutral* and *Irrelevant*. Table 1 shows examples from the challenge training dataset.

The task objective is to build a model that predicts the correct label given a *Context* text and the associated relationship triple. The participants were required to submit a ranked list of *Context* text from the evaluation dataset, assuming that the label set is ordered based on relevance (*Highly Relevant* > *Relevant* > *Neutral* > *Irrelevant*). The evaluation criteria was based on a Normalized Discounted Cumulative Gain (*NDCG*) scores over the ranked list.

We approach the problem as a binary classification problem, with the objective of predicting whether a *Context* text is *Highly Relevant*. We next describe two key aspects of our solution: (a) definition of *Domain-specific features* to capture domain knowledge about how financial entities and relationships are mentioned in text, and (b) training an ensemble model over three classifiers (Logistic Regression, Support Vector Machine and Learning to Rank)

DSMM'17, Chicago, IL, USA

DOI: http://dx.doi.org/10.1145/3077240.3077254

Example	*Context*	*Filing Entity*	*Mentioned Entity*	*Role*	*Label*
E1	*For example, because Ally controls Ally Bank, Ally is an affiliate of Ally Bank for purposes of the Affiliate Transaction Restrictions.*	Ally Financial Inc	Ally Bank	affiliate	*Relevant*
E2	*Item 5. Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities. M&T's common stock is traded under the symbol MTB on the New York Stock Exchange.*	M&T BANK CORP	Equity Securities	Issuer	*Irrelevant*

Table 1: Examples from Challenge Training Dataset

using the AdaBoost algorithm. We then used the probability scores produced by the ensemble classifier as the ranking criteria for our challenge submission.

2.1 Domain Features

We define two types of features: (i) *Generic features* such as token-level features (e.g., unigrams and bigrams), and (ii) *Domain features* to capture financial domain terminology about how financial entities and their relationships are mentioned in text. In this section, we describe the two categories of domain features we used for the challenge.

Financial Vocabulary Features. These features identify candidate entity mentions and candidate role mentions in the text. Example patterns include:

- `<Candidate Entity>` identifying whether a candidate entity is mentioned and whether the mention matches *Filing Entity* or *Mentioned Entity*. The semantic role of the entity is also represented as additional features.
- `<Candidate Role>` identifying whether a candidate role is mentioned and whether the mention matches *Role*.

Financial Relationship Pattern Features. These features represent proximity based patterns across the candidate entities and roles. Example patterns include:

- `<Candidate Entity> <upto n tokens> <Candidate Role>`
- `<Candidate Role> <upto n tokens> <Candidate Entity>`
- `<Candidate Entity> <upto n tokens> <Candidate Entity>`
- `<Candidate Entity> <upto n_1 tokens> <Candidate Role> <upto n_2 tokens> <Candidate Entity>`

For each relationship pattern, the part-of-speech information of the text matching the proximity pattern is represented as additional features.

We use the SystemT [2] text analytics engine to extract these features, which are then transformed into a boolean feature vector.

2.2 AdaBoost Ensemble Model

Given the feature vectors described above, and the label *Highly Relevant* as the class to predict, the training data is represented by the feature matrix $\mathbf{X}$ and the label vector $\mathbf{y}$. Let matrix $\mathbf{X}$ be the matrix where $\mathbf{X}_{\cdot i} = \boldsymbol{x}_i^T$, matrix $\mathrm{Y} = \mathrm{diag}(\boldsymbol{y})$. We choose the following base classifiers and combine their results using an ensemble method to build a more robust classifier as described below.

2.2.1 Base classifiers.

Logistic Regression (LR) is a linear model for classification. The probabilities describing the possible outcomes of a single trial are modeled using a logistic function. We use a binary class L2 penalized logistic regression, which minimizes the following cost function:

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \log(\exp(-y_i(X_i^T w + c)) + 1). \tag{1}$$

The solver uses a coordinate descent (CD) algorithm [5].

Support Vector Machine (SVM) Let vector $\mathbf{1}$ be an n-dimensional vector of all ones and C be a positive trade-off parameter. Then, the dual formulation of 1-norm soft margin SVM [7] is given by

$$\max_{\alpha} \mathbf{1}^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{YKY} \boldsymbol{\alpha} \tag{2}$$
$$s.t. \quad \boldsymbol{y}^T \boldsymbol{\alpha} = 0, 0 \le \boldsymbol{\alpha} \le C\mathbf{1}.$$

We adopt RBF kernel $\mathbf{K}$. To learn the SVM classifier, we use a convex quadratic programming to solve the dual problem in Eq. (2).

Learning to Rank (LTR) The following ranking model is formulated as a weighted combination of matching features:

$$\mathcal{R}(q, X_k) = \sum_{k=1}^{K} w_k X_k \tag{3}$$

Learning the weight vector $\{w_k\}_{k=1}^K$ is a standard learning-to-rank task. The goal of learning is to find an optimal weight vector $\{\bar{w}_k\}_{k=1}^K$, such that for any two instances X_i and X_j, the following condition holds:

$$\mathcal{R}(q, X_i) > \mathcal{R}(q, X_j) \Leftrightarrow r_{X_i} > r_{X_j} \tag{4}$$

where r_X denotes a numerical relevance rating labeled by human annotators denoting the relevance between q and X.

2.2.2 Ensemble Model. The goal of ensemble methods is to combine the predictions of several individual classifiers in order to improve the generalizability and robustness over the individual classifiers. We use the popular *AdaBoost* algorithm [4, 9]. It iteratively train a sequence of weak learners (e.g., the base classifiers described above) on repeatedly modified versions of the data. The boosted model produces its predication as a weighted combination of the weak learner predictions. In each iteration the sample weights are adjusted to emphasize those data points predicated wrongly by the boosted model, forcing the next weak learner to focus more on these examples.

Parameter Tuning We tune the parameters of the three base classifiers and the ensemble by performing five-fold cross validation on the challenge training set.

Score output The overall output of our model is the probability produced by the ensemble method. These are used as the scores that define the rank of the test data used for evaluation.

	gt1	gt1_500	gt2	gt3	gt4	gt5
score	**0.9223**	**0.8209**	**0.9593**	0.7270	**0.9431**	0.8029

Table 2: NDCG scores of our approach over the Challenge Evaluation Dataset

Features	Number@5	Number@500	Number@1,000
Generic	4 (0.07%)	255 (5.07%)	464 (9.23%)
Domain-specific	1 (0.16%)	245 (40.83%)	536 (89.33%)

Table 3: Feature Importance Ranking Distributions

Method	Train			Dev		
	P	R	F1	P	R	F1
LR	1.00	0.75	0.86	0.25	0.14	0.18
SVM	1.00	1.00	1.00	0.31	0.71	0.44
LTR	1.00	1.00	1.00	0.50	0.43	0.46
AdaBoost	1.00	0.75	0.86	1.00	0.30	0.46

Table 4: Comparing the Ensemble Model with Base Classifiers

3 EVALUATION RESULTS

In this section, we first describe the results of our submission for the challenge. We then analyze how domain features and an ensemble model helped our submission achieve good results in the challenge.

3.1 Challenge Results

The evaluation metric for the challenge is a weighted Normalized Discounted Cumulative Gain (NDCG) score over the ranked results [6]. Weights are assigned to the different labels to see how individual techniques performed under varying scenarios. Six configurations were used in the challenge evaluation: (gt1, gt1_500, gt2, gt3, gt4 and gt5) as described in [6].

Table 2 shows the NDCG scores of our challenge submission (P17 in [6]).

- Our submission achieved the highest score in four out of six evaluation metrics (gt1, gt1_500, gt2 and gt4). In three of these scenarios, the NDCG scores are over 0.9
- For the other two scenarios (gt3 and gt5), the NDCG scores were much lower for our submission. The maximum score achieved by any submission was also comparatively lower in these two scenarios (0.79 for gt3 and 0.86 for gt5).

3.2 Analysis of Our Approach

Two key aspects of our approach are the use of domain features and training an ensemble model. To better understand the contributions of these two components, we performed the following analysis. We use the challenge training dataset for this analysis, where we train our classifiers on 80% of the training set (Train) and use the remaining 20% of the training set for evaluation (Dev).

3.2.1 Do Domain Features contribute to model performance ? To see the contributions of the two groups of features (generic and domain features) designed in Sec. 2.1, we examine how they are utilized by our classifiers. We conduct a feature importance test [1] with the relative feature importance scores automatically generated by AdaBoost [8]. Table 3 shows for each group the number of features as well as the corresponding percentage among all features in the group, in top-N ranked positions (N = 5, 500 and 1,000) based on the relative feature importance scores from AdaBoost.

As can be seen, the percentage of domain features utilized by the ensemble model are substantially larger than the percentage of generic features. This indicates that the model relies more heavily on domain features, which are effective in capturing how relationships between entities and roles are mentioned in financial text.

3.2.2 Does an Ensemble Model help? Table 4 shows the precision (P), recall (R) and the F1 measure for the three base classifiers, and the ensemble method combining them. We see that the three base classifiers overfit the training data. In particular, both SVM and LTR learn the characteristic of the training data perfectly (precision and recall are equal to 1), but when applied on the Dev set, SVM exhibits very low precision, while LTR has both low precision and low recall. In contrast, the results with AdaBoost are more balanced across the Train and Dev sets, with precision remaining high on the Dev set. This indicates that for our problem setting, AdaBoost is less susceptible to overfitting, contributing to the good performance on the test dataset in the competition.

An astute reader might observe how our solution achieved high NDCG scores over the test dataset (Table 2), but the F-measure under a more typical classification evaluation setting over the Dev set is only 0.46 (Table 4). It turns out that NDCG scores used in this challenge with rank weight $1/\log(r+1)$, where r is the rank of an item, is less sensitive to variations of rank beyond the first few entries in the ranked list, as $\log(x)$ is a very-slowly increasing function for large x. An alternative of using $1/r$ as rank weight is likely to produce a more sensitive score. This alternative has the additional benefit of being robust against changes in the test data size: if test size is changed by a factor a, the rank of an item is approximately changed from r to ar. The NDCG scores calculated using weight $1/r$ remain unchanged, while those with the original rank weight $1/\log(r+1)$ suffer changes that make them incomparable. More generally, interesting questions arise about alternative evaluation metrics for this problem and developing a better understanding of the relationship across these different metrics.

4 CHALLENGES IN DEFINING LABELING GUIDELINES

As we investigated our model performance over the challenge training dataset, we discovered several semantic complexities in the labeled dataset providing for training. In this section, we describe some of these semantic complexities, which illustrate the challenges in precisely defining labeling guidelines for deep domain semantic tasks.

Consider the example training dataset entries shown in Table 5.

- Example E3 is labeled as *Highly Relevant* even though the financial relationship (PNC Financial Services Group Inc., PNC Bank, Seller) is not validated by the *Context* text. This entry has possibly been labeled as *Highly Relevant* due to the *Context*

Example	*Context*	*Filing Entity*	*Mentioned Entity*	*Role*	*Label*
E3	*During the fourth quarter of 2013, PNC finalized the wind down of Market Street Funding LLC (Market Street), a multi-seller asset-backed commercial paper conduit administered by PNC Bank, N.A.*	PNC Financial Services Group Inc.	PNC Bank, N.A.	Seller	*Highly Relevant*
E4	*4.1 Indenture, dated as of October 21, 2010, between JPMorgan Chase & Co. and Deutsche Bank Trust Company Americas, as Trustee (incorporated by reference to ...*	JPMorgan Chase & Co.	Deutsche Bank Trust Company Americas	Trustee	*Neutral*

Table 5: Examples illustrating Labeling Complexities

text being relevant to the *Filing Entity* and not as a validation of the relationship.

- Example E4 is labeled as *Neutral* even though the financial relationship (JPMorgan Chase & Co., Deutsche Bank Trust Company Americas, Trustee) is validated by the *Context* text.

The above examples illustrate complexities in the labeling due to the FEIII Challenge attempting to achieve two objectives at the same time: (i) validating financial relationships and (ii) identifying interesting and relevant knowledge about financial entities. In Section 5 we explore a simplified version of the task focused just on validating financial relationships and an associated labeling scheme.

The complexities of real-world relationships between multiple financial entities and how they are represented in regulatory filings raise additional challenges. We next describe some of these challenges.

Complex Financial Relationships. Many context passages describe multiple relationships involving several financial entities and roles. A further complication is that relationships often involve one or more other entities, forming an "entity chain," with each entity pair in the chain having a financial relationship. A triple could be validated by inferring a relationship through reasoning across the "entity chain". For instance, consider the *Context* text: "U.S. Bank, in its role as trustee of CHL Mortgage Loan Trust 2006-4SL, filed a lawsuit against Bank of America and affiliates, including Countrywide and Merrill Lynch". Several triples can be validated by this text involving five entities and two roles.

Labeling Guideline Challenges. The definition of financial entities and relationships has many nuances. For instance, should financial terms such as "Equity Management" be considered as financial entities? How does one handle scenarios where the entity mentions are ambiguous? E.g., does PNC refer to 'PNC Financial Services Group Inc.' or 'PNC Bank'? If the financial relationship is partially validated, what should be the label assigned?

Ambiguities in Labeled Data. Experts did not agree on the labels for a non-trivial fraction of data points in the training dataset (about 15%). For our submission, we chose the label provided by the first expert. The organizers addressed this issue in the test dataset by implementing a resolution step in the labeling process.

5 RELATIONSHIP VALIDATION

In this section we propose *Relationship Validation* (*Rel_Val*), a simplified task definition that focuses on evaluating whether the *Context* text validates the provided financial relationship triple.

5.1 Relationship Validation Labeling Scheme

Consider the tuple (*Context*, *Filing Entity*, *Mentioned Entity*, *Role*), where *Context* appears in a regulatory filing filed by *Filing Entity*. *Context* validates (*Filing Entity*, *Mentioned Entity*, *Role*) if *Context* confirms both (a) *Filing Entity* and *Mentioned Entity* are financial entities engaged in a relationship, and (b) Either *Filing Entity* or *Mentioned Entity* plays the role *Role* in the relationship.

Labels are assigned per the following scheme:

Validated: *Context* validates the relationship (*Filing Entity*, *Mentioned Entity*, *Role*)

Partial Validated: *Context* partially confirms the relationship (*Filing Entity*, *Mentioned Entity*, *Role*). One of the entities is confirmed to be in the relationship, and it is unclear whether the other entity is involved in the relationship.

Not Validated: *Context* invalidates the relationship (*Filing Entity*, *Mentioned Entity*, *Role*). One of the entities is confirmed to be in the relationship, and the other entity is confirmed not be part of the same relationship.

Irrelevant: Either *Filing Entity* or *Mentioned Entity* is not a financial entity, or neither of them participate in a relationship with role *Role*.

Revisiting the examples in Table. 5, using the *Rel_Val* labeling scheme, Example E3 is labeled as *Not Validated* and Example E4 is labeled as *Validated*.

5.2 Evaluation Results

We relabeled part of the challenge training dataset (345 out of 975 instances) according to the *Relationship Validation* labeling scheme and trained an ensemble model leveraging domain features as described in Section 2. We then labeled part of the challenge test dataset (350 out of 900 instances) according to the *Relationship Validation* labeling scheme and conducted experiments based on these labels. We used the configurations shown in Table 6 to compute the weighted *NDCG* scores, keeping them similar to the weight assignments used in the challenge evaluation.

	gt1	gt2	gt3	gt4	gt5
Validated	4	4	4	4	4
Partial Validated	0	3	0	3.5	3.5
Not Validated	3	2	0	3	0
Irrelevant	0	0	0	0	0

Table 6: Configurations for Evaluation under Relationship Validation Scheme.

	gt1	gt2	gt3	gt4	gt5
score	**0.9105**	**0.9411**	0.7259	**0.9681**	0.8379

Table 7: NDCG scores of our approach using Relationship Validation Scheme

Features	Number@5	Number@500	Number@1,000
Generic	4 (0.07%)	426 (8.48%)	883 (17.57%)
Domain	1 (0.16%)	74 (12.33%)	117 (19.50%)

Table 8: Feature Importance Ranking Distributions under the Relationship Validation Scheme

Table 7 shows the NDCG scores of our approach under the *Relationship Validation* labeling scheme. The scores are over 0.9 in three of the five configurations. Overall, these results closely track our challenge results in Table 2, coming within a range of 0.01 to 0.03 in all configurations. This provides promising evidence that the good performance of our approach combining domain features with an ensemble model carries over under alternative problem formulations.

Similar to Sec 3.2.1, we examined feature importance under the *Relationship Validation* scheme, as shown in Table 8. The percentage of domain-specific features utilized by the ensemble model is larger than the generic features. This indicates that domain features are effective in capturing the relationships between entities and roles, even under this task formulation.

Method	Train			Dev		
	P	R	F1	P	R	F1
LR	0.90	0.83	0.86	0.75	0.46	0.57
SVM	0.90	0.87	0.88	0.77	0.77	0.77
LTR	0.98	1.00	0.99	0.46	0.92	0.61
AdaBoost	0.92	0.82	0.86	0.77	0.77	0.77

Table 9: Comparing the Ensemble Model with Base Classifiers under the Relationship Validation Scheme

Similar to Sec 3.2.2, we examined how the ensemble model compares to the individual classifiers under the *Relationship Validation* scheme. The results are shown in Tables 9. We again see that the AdaBoost method performs the best on the test dataset, while two of the base classifiers (Logistic Regression and LTR) overfit the training data. SVM performs comparable to the AdaBoost method on this dataset.

5.3 Comparing Relationship Validation and Challenge Labeling Schemes

In order to understand how the two labeling schemes compared with each other, we analyzed how the 345 entries in the training dataset were labeled according to the two schemes. Table 10 summarizes how many entries were labeled with each pair of labels across the two schemes.

From these numbers, we observe that

	Validated	*Partial Validated*	*Not Validated*	*Irrelevant*
Highly Relevant	0	14	18	36
Relevant	40	6	26	57
Neutral	33	6	40	55
Irrelevant	1	3	2	18

Table 10: Comparing the Relationship Validation and Challenge Labeling Schemes

	gt1	gt1_500	gt2	gt3	gt4	gt5
score	0.7107	0.3347	0.8852	0.6228	0.8656	0.8332

Table 11: NDCG scores of model trained using Relationship Validation Scheme evaluated under Challenge Evaluation

- The challenge scheme attempts to capture multiple dimensions: interesting, relevant or validating *Context*. So, most of the entries are marked as *Highly Relevant*, *Relevant* or *Neutral*, as they fall under one of these requirements. Consequently very few entries are labeled as *Irrelevant*.
- In contrast, *Relationship Validation* scheme focuses only on validation, so less than one-third of the entries are labeled as *Validated* or *Partial Validated*.
- Entries (partially) validated in the *Relationship Validation* scheme (*Validated*, *Partial Validated*) are distributed across the *Highly Relevant*, *Relevant* and *Neutral* labels in the challenge scheme.

We also submitted the model trained under *Relationship Validation* scheme to the challenge. Table 11 shows the NDCG scores of this submission (P11 in [6]). Notice how these NDCG scores are much lower than our other submission (Table 2). These results indicate that the two labeling schemes are substantially different.

6 CONCLUSION AND DISCUSSION

In this paper, we described our approach to the scored task for the 2017 FEIII Data Challenge: an ensemble classification model leveraging domain features that capture financial domain terminology. Our submission performed well in the challenge evaluation obtaining the highest score in four out of six evaluation criteria. We presented multiple semantic complexities we encountered with regards to the task definition, labeled dataset and evaluation metrics. We presented a simplified version of the task focused on *Relationship Validation* and demonstrated how our approach performs well even in this alternative problem setting.

Potential directions to consider in subsequent versions of the challenge include:

- Defining precise labeling guidelines for the challenge task and alternative task definitions such as *Relationship Validation* introduced in this paper.
- Exploring the relative impact of domain features under various problem formulations and developing a systematic framework for leveraging domain features.
- Exploring alternative evaluation metrics, both under the ranked classification task definition (extensions to *NDCG*) and in a multi-class classification setting.

- Extending the formulation of the learning tasks so that additional information such as document structure and metadata can be provided as contextual information.
- Expanding the task definition to analyze, reason about and validate financial relationships from a corpus, going beyond document-at-a-time analysis and evaluation.
- Developing a comprehensive labeling, training and evaluation framework for deep-domain semantic tasks such as the financial relationship validation task.

REFERENCES

[1] Yi-Wei Chen and Chih-Jen Lin. 2006. Combining SVMs with various feature selection strategies. In *Feature extraction*. Springer Berlin Heidelberg, Berlin, Heidelberg, 315–324.

[2] Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li, Sriram Raghavan, Frederick R. Reiss, and Shivakumar Vaithyanathan. 2010. SystemT: An Algebraic Approach to Declarative Information Extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 128–137.

[3] DSFIN. 2016. FEIII: Financial Entity Identification and Information Integration. (2016). https://ir.nist.gov/dsfin/

[4] Yoav Freund and Robert E. Schapire. 1995. A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. In *Proceedings of the Second European Conference on Computational Learning Theory (EuroCOLT '95)*. Springer-Verlag, London, UK, UK, 23–37.

[5] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* 33, 1 (2010), 1.

[6] Louiqa Raschid, Doug Burdick, Mark Flood, John Grant, Joe Langsam, Ian Soboroff, and Elena Zotkina. 2017. Financial Entity Identification and Information Integration (FEIII) Challenge 2017: The Report of the Organizing Committee. In *Proceedings of the Workshop on Data Science for Macro-Modeling (DSMM@SIGMOD)*. ACM, New York, NY, USA.

[7] Chenguang Wang, Yangqiu Song, Haoran Li, Ming Zhang, and Jiawei Han. 2016. Text Classification with Heterogeneous Information Network Kernels.. In *AAAI*. AAAI Press, 2130–2136.

[8] Ruihu Wang. 2012. AdaBoost for feature selection, classification and its relation with SVM, a review. *Physics Procedia* 25 (2012), 800–807.

[9] Ji Zhu, Hui Zou, Saharon Rosset, and Trevor Hastie. 2009. Multi-class adaboost. *Statistics and its Interface* 2, 3 (2009), 349–360.