Towards Re-defining Relation Understanding in Financial Domain

Chenguang Wang, Doug Burdick, Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li and Huaiyu Zhu
IBM Research-Almaden
SIGMOD DSMM’17, Chicago, USA
May 14, 2017
• Challenge Task Definition
• Our Approach
• Semantic Complexities in Challenge Task Definition
• Relation Validation based Task Definition
• Conclusion and Discussion
Outline

• Challenge Task Definition
• Our Approach
• Semantic Complexities in Challenge Task Definition
• Relation Validation based Task Definition
• Conclusion and Discussion
## FEIII Data Challenge Scored Task

### – Tuple

A **Tuple** is represented as a 4-tuple: \((\text{Context}, \text{Filing Entity}, \text{Mentioned Entity}, \text{Role})\). The task involves ranking these tuples such that those with the context that best supports the role assignment and contains financially relevant knowledge are at the top of the ranking.

### Labeling Scheme

- **Highly Relevant**
- **Relevant**
- **Neutral**
- **Irrelevant**

### Examples

<table>
<thead>
<tr>
<th>Example</th>
<th>Context</th>
<th>Filing Entity</th>
<th>Mentioned Entity</th>
<th>Role</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>For example, because Ally controls Ally Bank, Ally is an affiliate of Ally Bank for purposes of the Affiliate Transaction Restrictions.</td>
<td>Ally Financial Inc</td>
<td>Ally Bank</td>
<td>affiliate</td>
<td>Relevant</td>
</tr>
<tr>
<td>E2</td>
<td>Item 5. Market for Registrant’s Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities. M&amp;T’s common stock is traded under the symbol MTB on the New York Stock Exchange.</td>
<td>M&amp;T BANK CORP</td>
<td>Equity Securities</td>
<td>Issuer</td>
<td>Irrelevant</td>
</tr>
</tbody>
</table>
### FEIII Data Challenge Scored Task

**Tuple**

(\textit{Context}, \textit{Filing Entity}, \textit{ Mentioned Entity}, \textit{Role})

**Relationship Triple**

(\textit{Filing Entity}, \textit{ Mentioned Entity}, \textit{Role})

<table>
<thead>
<tr>
<th>Example</th>
<th>Context</th>
<th>Filing Entity</th>
<th>Mentioned Entity</th>
<th>Role</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>\textit{For example, because Ally controls Ally Bank, Ally is an affiliate of Ally Bank for purposes of the Affiliate Transaction Restrictions.}</td>
<td>Ally Financial Inc</td>
<td>Ally Bank</td>
<td>affiliate</td>
<td>Relevant</td>
</tr>
<tr>
<td>E2</td>
<td>\textit{Item 5. Market for Registrant’s Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities. M&amp;T’s common stock is traded under the symbol MTB on the New York Stock Exchange.}</td>
<td>M&amp;T BANK CORP</td>
<td>Equity Securities</td>
<td>Issuer</td>
<td>Irrelevant</td>
</tr>
</tbody>
</table>
FEIII Data Challenge Scored Task

- **Tuple**  
  \((\text{Context}, \text{Filing Entity}, \text{Mentioned Entity}, \text{Role})\)

- **Relationship Triple**  
  \((\text{Filing Entity}, \text{Mentioned Entity}, \text{Role})\)

<table>
<thead>
<tr>
<th>Example</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>For example, because Ally controls Ally Bank, Ally is an affiliate of Ally Bank for purposes of the Affiliate Transaction Restrictions.</td>
</tr>
<tr>
<td>E2</td>
<td>Item 5. Market for Registrant’s Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities. M&amp;T’s common stock is traded under the symbol MTB on the New York Stock Exchange.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Filing Entity</th>
<th>Mentioned Entity</th>
<th>Role</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ally Financial Inc</td>
<td>Ally Bank</td>
<td>affiliate</td>
<td>Relevant</td>
</tr>
<tr>
<td>M&amp;T BANK CORP</td>
<td>Equity Securities</td>
<td>Issuer</td>
<td>Irrelevant</td>
</tr>
</tbody>
</table>

- **Relationship Triple Ranking**: Rank the relationship triples such that the triples with the context that best supports that role assignment and contains (financially) relevant knowledge are at the top of the ranking.
## FEIII Data Challenge Scored Task

### – Tuple

\[(\text{Context}, \text{Filing Entity}, \text{Mentioned Entity}, \text{Role})\]

### – Relationship Triple

\[(\text{Filing Entity}, \text{Mentioned Entity}, \text{Role})\]

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

### – Relationship Triple Ranking:

Rank the relationship triples such that the triples with the context that best supports that role assignment and contains (financially) relevant knowledge are at the top of the ranking.

### – Labeling Scheme

- Highly Relevant, Relevant, Neutral, Irrelevant
Key Challenges in the Task – Financial Domain

- Complex financial relationships
  - Context describes multiple relationships through an entity chain
  - Triple validated by inferring across the entity chain

  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

- Labeling guideline challenges
  - Definition of financial entity, e.g., “USA”
  - Ambiguous entity mentions, e.g. does “PNC” refer to “PNC Financial Services Group Inc.” or “PNC Bank”

- Ambiguities in labeled data
  - The two experts who labeled the training data did not agree on labels in 15% of the case

Domain Features!
Key Challenges in the Task – Learning Method

- Ranking models
  - Millions of training samples vs. 900+ samples
  - The label scheme naturally provides category information
- Classification models
  - No model is perfect across domains
  - Different kinds of model have different properties/advantages
Outline

• Challenge Task Definition
• Our Approach
• Semantic Complexities in Challenge Task Definition
• Relation Validation based Task Definition
• Conclusion and Discussion
Domain Features + Ensemble Classification Model
## Financial Domain Feature Exploration

### Feature types

<table>
<thead>
<tr>
<th>Feature types</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic features</td>
<td>Token unigrams and bigrams</td>
</tr>
</tbody>
</table>
| Domain features          | **Financial Vocabulary Features:** Identify candidate entity mentions and candidate role mentions in the text
                            - Semantic roles of the entities                                            |
|                          | **Financial Relationship Pattern Features:** Represent proximity based patterns across the candidate entities and roles
                            - Parts of speech for the tokens in the gaps
                            - Whether entities and roles match elements of the triple                  |

### Features are extracted using IBM Research SystemT

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.
Adaboost based Ensemble Model

1. Initialize the observation weights $w_i = 1/n, i = 1, 2, \ldots, n$.

2. For $m = 1$ to $M$:
   
   (a) Fit a classifier $T^{(m)}(x)$ to the training data using weights $w_i$.
   
   (b) Compute the error of a classifier
   
   \[
   \text{Error} = \sum_{i=1}^{n} w_i \mathbb{I}(c_i \neq T^{(m)}(x_i)) / \sum_{i=1}^{n} w_i.
   \]

   (c) Compute the weight of the classifier
   
   \[
   \alpha^{(m)} = \log \frac{1 - \text{Error}^{(m)}}{\text{Error}^{(m)}}.
   \]

   (d) Set
   
   \[
   w_i \leftarrow w_i \cdot \exp\left(\alpha^{(m)} \cdot \mathbb{I}(c_i \neq T^{(m)}(x_i))\right),
   \]
   
   for $i = 1, 2, \ldots, n$.

   (e) Re-normalize $w_i$.

3. Output

   \[
   C(x) = \arg \max_k \sum_{m=1}^{M} \alpha^{(m)} \cdot \mathbb{I}(T^{(m)}(x) = k).
   \]

Base classifiers:

1. Logistic Regression
2. SVM
3. Learning to Rank
Evaluation Results – Challenge Results

- **NDCG**
  - Standard ranking metric yields score between 0~1, the higher the better
  - Competition provides different weighting method

- **Results**
  - *Highest scores in 4 out of 6 evaluation metrics*

<table>
<thead>
<tr>
<th></th>
<th>gt1</th>
<th>gt1_500</th>
<th>gt2</th>
<th>gt3</th>
<th>gt4</th>
<th>gt5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our score P17</strong></td>
<td><strong>0.9223</strong></td>
<td><strong>0.8209</strong></td>
<td><strong>0.9593</strong></td>
<td><strong>0.7270</strong></td>
<td><strong>0.9431</strong></td>
<td><strong>0.8029</strong></td>
</tr>
<tr>
<td><strong>Min score</strong></td>
<td><strong>0.7107</strong></td>
<td><strong>0.3347</strong></td>
<td><strong>0.8852</strong></td>
<td><strong>0.6228</strong></td>
<td><strong>0.8575</strong></td>
<td><strong>0.7682</strong></td>
</tr>
<tr>
<td><strong>Max score</strong></td>
<td><strong>0.9223</strong></td>
<td><strong>0.8209</strong></td>
<td><strong>0.9593</strong></td>
<td><strong>0.7917</strong></td>
<td><strong>0.9431</strong></td>
<td><strong>0.8581</strong></td>
</tr>
<tr>
<td><strong>Mean score</strong></td>
<td><strong>0.8281</strong></td>
<td><strong>0.618</strong></td>
<td><strong>0.9252</strong></td>
<td><strong>0.6991</strong></td>
<td><strong>0.8973</strong></td>
<td><strong>0.8127</strong></td>
</tr>
</tbody>
</table>
Evaluation Results – Domain Feature Effectiveness

- Top important features in ensemble model

<table>
<thead>
<tr>
<th>Feature types</th>
<th>Number of features among top 1,000</th>
<th>Percentage among all features in the type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>464</td>
<td>9.23%</td>
</tr>
<tr>
<td>Domain-specific</td>
<td>536</td>
<td>89.33%</td>
</tr>
</tbody>
</table>

Percentage of domain features is substantially larger than the generic features among most important ones.

- Domain features are effective
  - Complex financial relationships, Labeling guideline challenges, Ambiguities in labeled data
• Precision, recall and F1

<table>
<thead>
<tr>
<th>Method</th>
<th>Train</th>
<th>Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Base</td>
<td>1.00</td>
<td>0.75</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Learning to Rank</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ensemble</td>
<td>1.00</td>
<td>0.75</td>
</tr>
</tbody>
</table>

– AdaBoost ensemble model is robust
  • No single model is perfect
  • We use ensemble outputs in our submission

Base classifiers overfitting
  • Perfect on training data
  • Poor on test data
Outline

- Challenge Task Definition
- Our Approach
- Semantic Complexities in Challenge Task Definition
- Relation Validation based Task Definition
- Conclusion and Discussion
### Semantic Complexities in Challenge Task Definition

#### Tuple \((\text{Context}, \text{Filing Entity}, \text{Mentioned Entity}, \text{Role})\)

#### Definition Dimensions

<table>
<thead>
<tr>
<th>Challenge Definition</th>
<th>Definition Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Validate financial relationships</td>
<td>1. Validate financial relationships</td>
</tr>
<tr>
<td>✓ Identify interesting and relevant knowledge about financial entities</td>
<td>2. Identify interesting and relevant knowledge about financial entities</td>
</tr>
</tbody>
</table>

#### Example

<table>
<thead>
<tr>
<th>Example</th>
<th>Context</th>
<th>Filing Entity</th>
<th>Mentioned Entity</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>E3</td>
<td>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.</td>
<td>PNC Financial Services Group Inc.</td>
<td>PNC Bank, N.A.</td>
<td>Seller</td>
</tr>
<tr>
<td>E4</td>
<td>4.1 Indenture, dated as of October 21, 2010, between JPMorgan Chase &amp; Co. and Deutsche Bank Trust Company Americas, as Trustee (incorporated by reference to ...</td>
<td>JPMorgan Chase &amp; Co.</td>
<td>Deutsche Bank Trust Company Americas</td>
<td>Trustee</td>
</tr>
</tbody>
</table>
• Challenge Task Definition
• Our Approach
• Semantic Complexities in Challenge Task Definition
• Relation Validation based Task Definition
• Conclusion and Discussion
### Definition Dimensions

<table>
<thead>
<tr>
<th>Challenge Definition</th>
<th>Relation Validation Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Validate financial relationships</td>
<td>✓</td>
</tr>
<tr>
<td>2. Identify interesting and relevant knowledge about financial entities</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Labeling Scheme

<table>
<thead>
<tr>
<th>Challenge Definition</th>
<th>Relation Validation Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Relevant</td>
<td><strong>Validated</strong>: Context validates the relationship</td>
</tr>
<tr>
<td>Relevant</td>
<td><strong>Partial Validated</strong>: One of the entities is confirmed to be in the relationship</td>
</tr>
<tr>
<td>Neutral</td>
<td><strong>Invalidated</strong>: Context invalidates the relationship</td>
</tr>
<tr>
<td>Irrelevant</td>
<td><strong>Irrelevant</strong>: One of the entities is not financial entity, or neither of them confirms the role</td>
</tr>
</tbody>
</table>

### Example

<table>
<thead>
<tr>
<th>Example</th>
<th>Context</th>
<th>Filing Entity</th>
<th>Mentioned Entity</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>E3</td>
<td>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.</td>
<td>PNC Financial Services Group Inc.</td>
<td>PNC Bank, N.A.</td>
<td>Seller</td>
</tr>
<tr>
<td>E4</td>
<td>4.1 indenture, dated as of October 21, 2010, between JPMorgan Chase &amp; Co. and Deutsche Bank Trust Company Americas, as Trustee (incorporated by reference to ...</td>
<td>JPMorgan Chase &amp; Co.</td>
<td>Deutsche Bank Trust Company Americas</td>
<td>Trustee</td>
</tr>
</tbody>
</table>
Evaluation Results – NDCG Results based on Relation Validation Definition

• NDCG Results

<table>
<thead>
<tr>
<th></th>
<th>gt1</th>
<th>gt2</th>
<th>gt3</th>
<th>gt4</th>
<th>gt5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>0.9105</td>
<td>0.9411</td>
<td>0.7259</td>
<td>0.9681</td>
<td>0.8379</td>
</tr>
</tbody>
</table>

Scores closely track our challenge results

– Our approach (domain features + ensemble classification model) is robust
  • Works for very different labeling schemes

– Size of RV labels: 345 out of 975 instances
Evaluation Results – Features and Models for Relation Validation Definition

- Top important features in ensemble model

<table>
<thead>
<tr>
<th>Feature types</th>
<th>Number of features among top 1,000</th>
<th>Percentage among all features in the group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>883</td>
<td>17.57%</td>
</tr>
<tr>
<td>Domain-specific</td>
<td>117</td>
<td>19.50%</td>
</tr>
</tbody>
</table>

- Domain features are effective

- Precision, recall and F1 measure

<table>
<thead>
<tr>
<th>Method</th>
<th>Train</th>
<th>Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Base</td>
<td>Logistic Regression</td>
<td>0.90</td>
</tr>
<tr>
<td>Base</td>
<td>Support Vector Machine</td>
<td>0.90</td>
</tr>
<tr>
<td>Base</td>
<td>Learning to Rank</td>
<td>0.98</td>
</tr>
<tr>
<td>Ensemble</td>
<td>AdaBoost</td>
<td>0.92</td>
</tr>
</tbody>
</table>

- AdaBoost ensemble model is robust

LR and LTR overfit the training data  
SVM and AdaBoost are similar, both robust  
Percentage of domain features is larger than the generic features
Outline

• Challenge Task Definition
• Our Approach
• Semantic Complexities in Challenge Task Definition
• Relation Validation based Task Definition
• Conclusion and Discussion
Future Explorations for the Challenge

- Precise labeling guidelines
  - Labeling could be done across multiple dimensions, e.g., relationship validation and interestingness from financial perspective

- Expand task definition
  - Document-at-a-time analysis: Allow additional information such as document structure and metadata to be provided as context
  - Collection-level analysis: analyze, reason and validate financial relationships across a corpus, going beyond document-at-a-time analysis

- Unified framework for deep-domain semantic tasks such as financial domain relation understanding
  - Labeling: precise guideline
  - Model: domain features + ranking or classification model pool
  - Evaluation: alternative evaluation metrics
  - Overall: system and platform includes the above components
• **Our Approach**
  – Domain features + Ensemble classification model

• **Our Result**
  – Our approach achieves the highest score in four out of six evaluation criteria

• **Our Exploration**
  – *Concentration is the secret of strength* - Ralph Waldo Emerson
  – A more simplified and precise challenge problem definition:
    • E.g., our relation validation definition: concentrating on relation validation