ABSTRACT
Relation extraction models trained on a source domain cannot be applied on a different target domain due to the mismatch between relation sets. In the current literature, there is no extensive open-source relation extraction dataset specific to the finance domain. In this paper, we release FinRED, a relation extraction dataset curated from financial news and earning call transcripts containing relations from the finance domain. FinRED has been created by mapping Wikidata triplets using distance supervision method. We manually annotate the test data to ensure proper evaluation. We also experiment with various state-of-the-art relation extraction models on this dataset to create the benchmark. We see a significant drop in their performance on FinRED compared to the general relation extraction datasets which tells that we need better models for financial relation extraction.

KEYWORDS
- Deep learning → Span extraction; • Natural language processing → Information extraction, Relation Extraction; • Domain → Financial.

INTRODUCTION
The task of relation extraction (RE) is defined as identifying triplets from text. Recently joint entity and relation extraction models have been proposed to eliminate the dependency on an external NER module. These models have achieved remarkable performance in this task. Freebase-New York Times dataset (FB-NYT) [9] is obtained by mapping relational triplets from Freebase [3], WebNLG [7] is a natural language generation dataset created semi-automatically from DBPedia [2] triplets. Although KBs such as Freebase, Wikidata, DBpedia contain a lot of relational triplets from different domains such as finance, material science, legal, but that is not reflected in the existing RE datasets as they used very generic sources of text corpus. There have been some efforts to create datasets for particular domain-specific relation extraction such as legal [1], biomedical [6, 8, 12, 17], scientific articles [11, 13]. There have been some efforts to create datasets in the financial domain [10, 18, 21], however to the best of our knowledge, there is currently no standard dataset in English to understand financial relations. There has been some work on financial numeral understanding [4, 5], however, our paper focuses on textual relations and not numerical. To bridge this gap, we create a finance domain-specific distance supervised relation extraction dataset. We obtain the financial relations from Wikidata KB and use text from the finance domain to create the FinRED dataset.

DATASET SOURCES AND CREATION
The FinRED dataset has been created using two sets of documents: 1) Webhose Financial News, and 2) Earning Call Transcript (ECT).

Webhose Financial News: For the financial news articles, we use the freely available dataset created by Webhose. It contains 47,851 English financial news articles crawled from July 2015 - October 2015.

Earning Call Transcripts: We collect 4,713 ECTs dated from June 2019 to September 2019 from seekingalpha.com. We use the presentation as well as the questionnaire portion of the transcript and remove monologues with less than 200 characters. Here monologue refers to a monologue given by a company participant during the presentation or the response given by a company participant during the questionnaire. In the ECT corpus, we have about 200K monologues, 1.8M sentences with an average of 7.19 sentences per monologue. In total, we have 152K tokens in the corpus.

Knowledge Base (KB): We use a subset of the Wikidata KB as a source of relational triplets for financial domain containing manually filtered 29 financial relations. Using the idea of distance supervision [14], we align the relational triplets to the text corpus.

Dataset: In total we obtain about 21,000 sentences using the distance supervision method, however a lot of these sentences are
Table 1: Comparison between different RE datasets.

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Train</th>
<th>Test</th>
<th>#Relations</th>
<th>#Financial Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB-NYT</td>
<td>56,196</td>
<td>5,000</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>WebNLG</td>
<td>5,519</td>
<td>703</td>
<td>216</td>
<td>12</td>
</tr>
<tr>
<td>FinRED (Ours)</td>
<td>5,699</td>
<td>1,068</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 2: Examples in the dataset

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Head</th>
<th>Relation</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antony Jenkins has been sacked as chief executive officer of Barclays Plc.</td>
<td>Anthony Jenkins</td>
<td>chief executive officer</td>
<td>Barclays</td>
</tr>
<tr>
<td>MEXICO CITY — State-owned Mexican oil company Pemex is reporting second quarter losses of $US 2 billion ($A7.16 billion) due mainly to lower petroleum prices</td>
<td>Pemex</td>
<td>product or material produced</td>
<td>petroleum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 EXPERIMENTS

To avoid the use of external NER module, we choose 3 joint entity and relation extraction models, namely, SPN [16], TPLinker [19], and CasRel [20] for our experiments. These models achieved state-of-the-art performance on standard relation extraction datasets Freebase-New York Times [9] and WebNLG [7]. We report precision, recall and F1 score for triplet extraction based on exact entity matching criterion and report the results in Table 3.

From Table 3, we can observe a drop of ∼4% for SPN and ∼25-30% for TPLinker and Casrel models in F1 as compared to FB-NYT and WebNLG datasets. Drop in the performance showcases that more research is required to better capture performance. For the SPN model, we observe that the relation F1 value is 88.37% and the entity F1 value is 96.36% showcasing that while the model can predict the entity properly, the lower model performance can be attributed to incorrect relation classification. We see similar trends in the performance of TPLinker and CasRel.

Table 3: Performance of the three baseline models on FinRED.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPN</td>
<td>89.22%</td>
<td>86.42%</td>
<td>87.80%</td>
</tr>
<tr>
<td>TPLinker</td>
<td>76.59%</td>
<td>63.45%</td>
<td>69.40%</td>
</tr>
<tr>
<td>CasRel</td>
<td>69.71%</td>
<td>55.84%</td>
<td>62.01%</td>
</tr>
</tbody>
</table>

Table 4: Performance of SPN model on the sentence classes with different number triplets and with type of overlapping triplets in them.

<table>
<thead>
<tr>
<th>Metric</th>
<th>1 triplet</th>
<th>2 triplets</th>
<th>3 triplets</th>
<th>4 triplets</th>
<th>5 triplets+</th>
<th>NEO</th>
<th>EPO</th>
<th>SEO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>76.15%</td>
<td>86.46%</td>
<td>88.88%</td>
<td>87.60%</td>
<td>80.98%</td>
<td>78.40%</td>
<td>89.61%</td>
<td>87.36%</td>
</tr>
<tr>
<td>Recall</td>
<td>83.36%</td>
<td>88.88%</td>
<td>87.60%</td>
<td>87.36%</td>
<td>80.98%</td>
<td>78.40%</td>
<td>89.61%</td>
<td>87.36%</td>
</tr>
<tr>
<td>F1</td>
<td>83.36%</td>
<td>87.60%</td>
<td>87.36%</td>
<td>87.36%</td>
<td>80.98%</td>
<td>78.40%</td>
<td>89.61%</td>
<td>87.36%</td>
</tr>
</tbody>
</table>

Performance of SPN model on the sentence classes with different number of triplets is reported in Table 4. The performance improves with the number of triplets, but degrades marginally with more than 5 triplets. We hypothesise that a few more triplets (typically EPO and SEO) assist the machine make better predictions, but too many confuse it.

In Table 4 we also report the performance of SPN model on the sentence classes with the type of overlapping triplets in them. Based on the overlap of entities, the sentence can be divided into three types: 1) NEO (No entity overlap) 2) EPO (Entity Pair Overlap) and 3) SEO (Single Entity Overlap). Here, we can observe that the model shows a slightly higher performance for SEO (F1: 87.98%) and EPO (F1: 87.98%) as compared to NEO (F1: 83.36%). We surmise that this is because the entity prediction part of the model performs well so for multi-label and overlapping triplets, the performance is higher since it needs to identify less number of entities.

4 CONCLUSION

In this paper, we propose FinRED, a relation extraction dataset for the finance domain curated from earning call transcript corpus and financial news articles. It contains more finance domain-specific relations than the existing RE datasets. We experimented with three state-of-the-art joint entity and relation extraction models on this dataset and saw a significant drop in F1 score compare to general domain RE datasets showcasing that more research is required on the models for FinRED.

ACKNOWLEDGMENTS

This research was partially supported by Goldman Sachs sponsored research grant FTHS. (FinTalk: Research towards creating a platform for highlight generation and summarization of financial documents while taking into account user feedback).
REFERENCES


