Financial Sentiment Analysis - Understanding Financial Reports

Tao Jin
jin.hybr@gmail.com

Abstract

Natural Language Processing and Understanding is becoming a more important strategic tool for financial analysis. To understand unstructured financial reports, such as earning calls and Fed. Meeting Minutes,this project investigate the method and model to predict the related financial sentiment.

1 Problem Definition

Sentiment analysis is the key problem of Natural Language Understanding. Based on the amount of text, sentiment analysis models try to identify the opinions from those text such as various reviews or comments. Financial Sentiment Analysis differs with the general sentiment analysis such as movie or product reviews.The major purpose for financial sentiment analysis is to predict market trends based on the current financial reports such as companies' earning call report in Micro way - stock market reaction or Federal Reserve meeting minutes in Marco way - Monetary policy. The big challenge for financial sentiment analysis is due to lack of labeled data in financial domain, so general purpose models and methods are not effective used in financial related text.

2 Literature Review

The literature review will be divided into two parts: 1. general sentiment analysis frameworks and methods and 2. financial domain sentiment analysis.

2.1 Sentiment Analysis Framework and Method

In sentiment analysis field, the main research area is focused on the different levels of granularity of texts: from document level, sentence level and aspect level.

2.1.1 Document Level Method

Document level sentiment analysis is to predict the overall positive or negative opinion. This method is consider the whole document as the single entity.The bag-of-words (BoW) with N-grams norms is widely used in the to transfer the whole document to a scalar feature vector with fixed space, each element can be used as word occurrence, word frequency or TF-IDF value. The major problem for this method is the matrix is normally very sparse, dimension is very large which equals to the size of vocabulary from the whole document, further more, the matrix lost the semantics of words due to no order in the model. To tackle these problems , word embedding is proposed to encode some semantic and syntactic properties of words and generate dense vectors for word representation. Different models proposed by researchers in document level sentiment analysis, Le and Mikolov [1] proposed Paragraph Vector, an unsupervised learning algorithm that learns vector representations for variable-length texts such as sentences, paragraphs and documents. The vector representations are learned by predicting the surrounding words in contexts sampled from the paragraph.Tang et al [2] build a neural network model to learn the whole document representation with the sentence relationships.In the sentence learning, [2] employed the CNN or LSTM for sentence embedding, then a GRU to encode semantics of sentences and document to classify the sentiment.

2.1.2 Sentence Level Method

In Sentence Level is to classify the sentiment based on a single given sentence. Sentence level analysis can be inferred with subjectivity classification and polarity classification[3]. The sentence level classification problem is normally predict a sentence as positive, neural and negative. Adding more semantic information such as parse trees, opinion lexicons, and part-of-speech(POS) tag is key benefit in sentence level due its the word dimension is small compared to documents. The key area most researchers focused on is how to generate the best sentence representation to combine with the more semantics features from the words to sentences. [4] authors proposed a Character to Sentence CNN model to extract the features from words to sentences to predict sentiment. The other models include using the bi-directional LSTM to learn the sentiment strength based on context-sensitive lexicon-based method [5].

2.1.3 Aspect based Method

Aspect based sentiment analysis considers both sentiment and the entity aspect. It should model the surrounding context words towards to the target.An aspect is defined as an attribute or a feature an words or sentences possesses.There are two methods : Aspect-Term Sentiment Analysis and Aspect-Category Sentiment Analysis [6], the different between two methods is how to define the An Aspect, which is based on key words in sentences(Term) or from a pre-defined set of category. The important method to deal aspect-based problems is divided into three parts: Aspect Extraction, Text representations and Model to classify the sentiment.

- Aspect Extraction

  Aspect extraction is important for the sentiment analysis. The major methods are frequency-based, relation-based, supervised machine learning, unsupervised machine learning and hybrid approaches.Most recently, deep learning based methods achieved the better results due to the model can learn feature representation which properly integrated with context and semantics information. In [12], authors proposed bidirectional LSTM model for extracting of opinion entities and relations that connect the entities. [13] designed a joint model with RNN and Condition Random Fields(CRF) to extract aspects and opinion terms. In [14], authors first proposed an attention-based model for unsupervised aspect extraction. The major improvement is to let attention-mechanism is more focused on the aspect-related words while ignore the aspect irrelevant words during the embedding learning.[15] proposed an attention-based deep distance metric learning model to group aspect phrases. The attention-based model is to learn feature representation of contexts. Both aspect phrase embedding and context embedding are used to learn a deep feature subspace metric for K-means clustering.

- Feature Representation

  The difference between Document and Sentence Level representation is how to generate the Aspect/Target representation which need combined with the context words and how to learn and identify the important context sentiment words.Word embedding plays an important role in sentiment analysis. Mass et al in [16] first proposed the model to learn both semantics and sentiment information. In [17], authors proved that an n-gram model combined with latent representation would be more suitable for sentiment classification.[18] and [19] presented models to learn Sentiment-specific Word Embeddings(SSWE), in which not only the semantic but also sentiment information is embedded in the learned word vectors. The other methods for enriching the feature representation including using multi-sense word embedding[20] and multi-prototype word embedding [21]. Zhang et al [22] proposed word embedding with matrix factorization for personalized review-based rating prediction and refine existing semantics-oriented word vectors (e.g., word2vec and GloVe) using sentiment lexicons.

- Model

  For the modeling to classify aspect based sentiment, most of recently models are based on LSTM. Dong et al [7] propsoed AdaRNN for target-dependent twitter sentiment classification, which learns to propagate the sentiments of words towards the target depending on the context and syntactic structure. It uses the representation of the root node as the features, and feeds them into the softmax classifier to predict the distribution over classes. Ruder et al [8] proposed the hierarchical bidirectional LSTM model which is integrated intra-sentences and inter-sentence relations. The network input is the word embedding for a sentence-level bidirectional LSTM, then in the final layer concatenated with the target embedding as input into the review level LSTM. The model output is the a probability distribution over the sentiments.To improve the integration the context words and semantics information, researchers achieved a lot of progresses by using the attention mechanism into LSTM model. [9] proposed an attention-based LSTM model with target embedding, in this way, the model has effectively integrated with related part of a sentence. Furthermore, [10] used two attention-based bidirectional LSTM model to improve the performances. [11] designed a attention-based model to combine the left-side context and the right-side context of a given Aspect.

2.2 Financial Sentiment Analysis

Based on the traditional sentiment analysis method and model, researchers have achieved a lot progresses in Financial area. In [23] authors proposed a comprehensive survey of financial text analysis by applying Bag-of-Words and Lexicon-based method. Pagolu et al.[24] investigate n-grams model for Financial related Twitter are fed into supervised machine learning algorithms to detect the sentiment. In [25], authors first applied the LSTM model to analysis the company announcements to predict stock market movements. Sohangir in [26] investigate the generic deep learning models to classify the StockTwits data set to classify the market sentiment. In [27] authors use a combination of text simplification and LSTM network to classify market sentiment based on the Financial Phrase Bank.

- Word Embedding v.s. Language Model

  In the traditional sentiment analysis, the models use the word embedding or LSTM model to learn the feature representation. The key difference between embedding method and language model is the language model returns more contextualized embedding instead of a single vector presentation. Recently progresses in language model can be success successfully fine-tuned for most downstream NLP/NLU tasks with small modifications. These models are usually trained on very large corpora, and then with addition of suitable task-specific layers fine-tuned on the target data set.By using these language models, it can achieve a better performance in sentiment analysis, such as ELMo (Embedding from Language Models) model[29], ULMFit (Universal Language Model Finetuning) [31] and Bidirectional Encoder Representations from Transformers (BERT) [30]. In [32], author proposed FinBert which is first build BERT model into financial domain. Author investigated three baseline models: LSTM classifier with GLoVe embedding, LSTM classifier with ELMo embedding and ULMFit Classifier. The results show BERT based model outperform other baseline models in every measured metric on current state-of-the-art results for two financial sentiment analysis data sets - Reuter TRC2 dataset [1], FiQA Sentiment Dataset[38] and Financial PhraseBank[37].

- BERT model based Aspect-based Sentiment Analysis

  With the recent success of BERT-based models, various research works have used BERT to generate contextualized embedding for input sentences, which are then used to classify sentiment for target-aspect pairs [33]. In [34] authors constructed auxiliary sentences with different pairs of targets and aspects or modifying the top-most classification layer to also take in targets and aspects [34].

- Financial Domain Sentiment Index

  Financial Domain related sentiment index is the domain related knowledge to measure financial market sentiment. In [35], the paper proposed a 10 different dictionary-based sentiment indexes for forecasting economics movement.These indexes give the throughout reference and explanation how financial reports will impact on the market movement and an useful reference in the financial sentiment analysis.

3 Experiment Protocol

3.1 Hypotheses

The hypotheses of this project are based on the language model and word embedding feature presentation in Financial domain. Based on the domain financial dataset, the model will predict the sentiment of sentence from financial reports.

[1]https://trec.nist.gov/data/reuters/reuters.html

3.2 Datasets

- Financial PhraseBank Dataset
  Maloet. [37] built the sentence-based dataset from financial domain and labeled by the 16 experts. The PhrasBank includes 4845 sentences from LexisNexis database. The labels are 'Negative', 'Neutral' and 'Positive'. The aspect is focused on the potential impact on the mentioned company's stock price. The objective of the phrase level annotation task was to classify each example sentence into a positive, negative or neutral category by considering only the information explicitly available in the given sentence. Since the study is focused only on financial and economic domains, the annotators were asked to consider the sentences from the view point of an investor only; i.e. whether the news may have positive, negative or neutral influence on the stock price. As a result, sentences which have a sentiment that is not relevant from an economic or financial perspective are considered neutral. the dataset is also include the agreeable level of labels by the annotators. There are three levels from 70% ,66% to 50 %.

- Customized FiQA Sentiment Dataset
  FiQA is Aspect-based financial sentiment analysis dataset. Given a text instance in the financial domain (microblog message, news statement or headline) in English, detect the target aspects which are mentioned in the text (from a pre-defined list of aspect classes) and predict the sentiment score for each of the mentioned target aspect. Sentiment scores will be defined using continuous numeric values ranged from -1(negative) to 1(positive). 686 annotated financial statements will be made available. In this project, will customize the FiQA dataset to an extra Financial PhraseBank dataset. By using mapping the sentiment scores values to positive, negative or neutral labels, this method will give model more data, also the FiQA dataset will include the Aspect information.

- Earnings Call and Federal Federal Open Market Committee Transcripts
  Earning call transcripts[2] and FOMC meeting minutes are available online[3].This project will chose some suitable examples from both resource to build a new sentiment dataset.

3.2.1 Evaluation Metrics

For evaluate the NLP/NLU classification model,this project will use macro F1 average. Also the model will use the Accuracy and cross entropy loss to evaluate the model.Macro F1 average calculates F1 scores for each of the classes and then takes the average of them. The average F1 will be a better metric for some imbalanced dataset.

3.3 Baseline Model

The baseline model will use the traditional Scikit-Learn Models and Bidirectional LSTM model.

- Baseline Model - Logistic Regression and Naive Bayes With TF-IDF
  In this traditional baseline model, this project will use the TF-IDF to vectorize the financial text data and use Logistic Regression, multinominal Naive Bayes classifier and Gradient Boosting Classifier as the basic models to analysis the sentiment.

- Baseline Model - BiLSTM Classifier with GloVe embedding
  By using bi-directional LSTM model as the baseline to compare with the language models, such as Bert and RoBerta Model. The experiments will apply the GLoVe embedding(300D) as the word embedding.The final layer is soft-max prediction for likelihood of three labels.

3.4 General Reasoning

Based on the fine tuning BERT language model to predict sentiment and with domain dataset, this project is trying to build a domain related BERT model to predict sentiment.The tasks are divided into following parts:

1. The vanilla BERT model
   The vallina BERT model will give a good reference on how language model to classify the sentiment.Through the BERT tokenizer, build the DataLoader, train the vallia BERT model.

[2]https://seekingalpha.com/earnings/earnings-call-transcripts

[3]https://www.federalreserve.gov/monetarypolicy/fomchistorical2010.htm

2. The RoBERTa model
   RoBERTa model is built upon BERT with longer with more data, bigger batch sizes while only pre-training on masked language modeling as opposed to pre-training on next sentence prediction as well. This project will also train the related financial dataset and predict the sentiment based on financial reports.

4 Experiments and Results

This project has setup experiments to investigate and compare the methodology and results of baseline models vs Language model based sentiment analysis. The baseline model will apply the machine learning models such as Logistic Regression, Naive Bayes and Grading Boosting Model to classify the sentiment. Word embedding such as GloVe model will apply into BiLSTM model. The language model method will use pre-trained BERT, RoBerta and XLM models.
The experiment environment is based on Google Co-Lab and laptop.

4.1 Dataset

This project uses Financial PhraseBank dataset. The whole datasets include the the different agree-grade sets. For simply experiment setup purpose,the experiment dataset combines all data into one csv file with three sentiment labels - Positive, Neural and Negative and 4846 sentences.

4.2 Baseline Models - ML models

The pipeline for the ML models such as Logistic Regression, Naive Bayes and Grading Boosting Model is to prepossess the text data,then use the TF-IDF to vectorize the sentence, the final step is to use the specific models to fit and predict the results.
The key methods for text data cleaning include the remove the punctuation, stop words, frequent words, stemmize and lemmatize the words. The baseline model use the simple numerical statistic method such as TD-IDF to vectize the text data.

4.3 Baseline Model - GloVe + BiLSTM

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.There are several vector size in GloVe embedding, in this project use vector size is 300D. Bi-directional LSTM model has achieved a lot of progresses in NLP field. In this baseline model, will use the GloVe embedding(300D) with a small BiLSTM network. The maximum len for text is set to 46(max. length of sentences.
The baseline model result is i the following table.

Model	Accuracy	F1 Score(Macro avg)
Logistic Regression	0.76	0.70
Multinomia NB	0.71	0.53
Grading Boosting	0.78	0.74
GloVe + BiLSTM	0.75	0.70

4.4 Language Model - BERT, RoBERTa and XLNet

To compare with the baseline models, this project use the pre-trained language based transformer models such as BERT, RoBERTa and XLNet.For simplifying the training processes, using the Simple Transformers packages which is based on the Hugging Face library to train the model. The results for transformer model is based on one epoch training. The following models are used.

1. BERT Model
   Pre-trained Model : bert-base-uncased
   Parameters: 2-layer, 768-hidden, 12-heads, 110M parameters

2. RoBERTa Model
   Pre-trained Model : roberta-base, Parameters: 125M parameters RoBERTa using the BERT-base architecture

3. XLNet Model Pre-trained Model : xlnet-base-cased, Parameters: 110M parameters2-layer, 768-hidden, 12-heads

The one-epoch training results are in the following table.

Model(1 epoch)	Accuracy	F1 Score(Macro avg)
BERT	0.84	0.82
RoBERTa	0.72	0.63
XLNet	0.77	0.74

4.5 Model Prediction

The whole training processes use the Financial PhraseBank dataset, for prediction, this projects created a headline sentence set from major business medias such as Bloomberg, CNBC, Benzinga etc. The example predictions are as following:

1. Headline :
'Apple supplier Foxconn warns that component shortages will last until 2022',
Result: positive
2. Headline :
'Gartner Stock Gives Every Indication Of Being Modestly Overvalued'
Result: negative
3. Headline :
'Japan stocks jump more than 1.5 percent as other major markets close for Good Friday'
Result: neutral

4.6 Error Analysis

For all the results, only BERT model achieved F1 score 0.82.
For example - the error prediction :
Headline: 'Dollar heads for third weekly gain as payrolls data looms'
Prediction: neutral. True: positive
The major problem for this error and performance problem lies below :

1. Lack of Domain Corpus
   The training data is still limited. Due to data limitation, the performance is hard to improve.

2. Data Precision
   The Financial Phrasebank data is reviewed by 16 reviewers, so there are a lot of disagreement on the labels. This will cause the training data is not accuracy based on different reviewers' opinions.

3. Fine tuning Model
   Because all transformer models used are pretrained models. So there is no fine tuning training processes to decide which layer to perform best classification results

5 Conclusion and Future Works

The sentiment analysis of financial report is very important for the market. In the current market situation, a simple headline may cause the major market volatility. This project examined the domain dataset and methodology, built ML based model, GloVe Embedding BiLSTM and Transformer based model and further more to compare with different model's performance. Further work should more focus on building more Aspect-based dataset, creating more accuracy labels such as somehow positive or somehow negative with aspect related result and the Fine-tuning training process to achieve the better result.

References

[1] Le Q, Mikolov T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning (ICML 2014), 2014.

[2] Tang D, Qin B, Liu T. Document modelling with gated recurrent neural network for sentiment classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), 2015.

[3] Wiebe J, Bruce R, and O'Hara T. Development and use of a gold standard data set for subjectivity classifications. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 1999), 1999.

[4] Dos Santos, C. N., Gatti M. Deep convolutional neural networks for sentiment analysis for short texts. In Proceedings of the International Conference on Computational Linguistics (COLING 2014), 2014.

[5] Teng Z, Vo D-T, and Zhang Y. Context-sensitive lexicon features for neural sentiment analysis. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), 2016.

[6] Wei Xue, Wubai Zhou, Tao Li, and Qing Wang. 2017. Mtna: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), volume 2, pages 151–156

[7] Dong L, Wei F, Tan C, Tang D, Zhou M, and Xu K. Adaptive recursive neural network for target-dependent Twitter sentiment classification. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2014), 2014.

[8] Ruder S, Ghaffari P, Breslin J.G. A hierarchical model of reviews for aspect-based sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2016), 2016.

[9] Wang Y, Huang M, Zhu X, and Zhao L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), 2016.

[10] Yang M, Tu W, Wang J, Xu F, and Chen X. Attention-based LSTM for target-dependent sentiment classification. In Proceedings of AAAI Conference on Artificial Intelligence (AAAI 2017), 2017.

[11] Liu J, Zhang Y. Attention modeling for targeted sentiment. In Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017), 2017.

[12]Katiyar A, Cardie C. Investigating LSTMs for joint extraction of opinion entities and relations. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2016), 2016.

[13] Wang W, Pan SJ, Dahlmeier D, and Xiao X. Recursive neural conditional random fields for aspect-based sentiment analysis. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), 2016.

[14] He R, Lee WS, Ng HT, and Dahlmeier D. An unsupervised neural attention model for aspect extraction. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2017), 2017.

[15] Xiong S, Zhang Y, Ji D, and Lou Y. Distance metric learning for aspect phrase grouping. In Proceedings of the International Conference on Computational Linguistics (COLING 2016), 2016.

[16] Mass A. L, Daly R. E, Pham P. T, Huang D, Ng A. Y. and Potts C. Learning word vectors for sentiment analysis. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2011), 2011.

[17] Bespalov D, Bai B, Qi Y, and Shokoufandeh A. Sentiment classification based on supervised latent n-gram analysis. In Proceedings of the International Conference on Information and Knowledge Management (CIKM 2011), 2011.

[18]Tang D, Wei F, Yang N, Zhou M, Liu T, and Qin B. Learning sentiment-specific word embedding for twitter sentiment classification. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2014), 2014.

[19] Tang D, Wei F, Qin B, Yang N, Liu T, and Zhoug M. Sentiment embeddings with applications to sentiment analysis. IEEE Transactions on Knowledge and Data Engineering, 2016.

[20] Li J, Jurafsky D. Do multi-sense embeddings improve natural language understanding? In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), 2015.

[21] Ren Y, Zhang Y, Zhang, M and Ji D. Improving Twitter sentiment classification using topic-enriched multiprototype word embeddings. In Proceeding of AAAI Conference on Artificial Intelligence (AAAI 2016), 2016.

[22] Zhang W, Yuan Q, Han J, and Wang J. Collaborative multi-Level embedding learning from reviews for rating prediction. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2016), 2016.

[23] Tim Loughran and Bill Mcdonald. 2016. Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research 54, 4 (2016), 1187–1230.

[24] Venkata Sasank Pagolu, et al. Sentiment Analysis of Twitter Data for Predicting Stock Market Movements, 2016, arXiv:1610.09225

[25] Mathias Kraus and Stefan Feuerriegel. 2017. Decision support from financial disclosures with deep neural networks and transfer learning. Decision Support Systems,104 (2017), 38–48.

[26]Sahar Sohangir, Dingding Wang, Anna Pomeranets, and Taghi M Khoshgoftaar.2018. Big Data: Deep Macedo

[27]Maia, Andrï£¡ Freitas, and Siegfried Handschuh. 2018. FinSSLx: A Sentiment Analysis Model for the Financial Domain Using Text Simplification. In 2018 IEEE 12th International Conference on Semantic Computing (ICSC). IEEE, 318–319.

[28] Neel Kant, Raul Puri, Nikolai Yakovenko, and Bryan Catanzaro. 2018. Practical Text Classification With Large Pre-Trained Language Models. (2018).

[29]Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. (2018).

[30] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018).

[31] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine tuning for Text Classification.

[32] Dogu Araci, FinBERT: Financial Sentiment Analysis with Pre-trained Language Models, arXiv 1908.10063.

[33] Huang, B.; and Carley, K. M. 2019. Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 5472–5480.

[34] Rietzler, A.; Stabinger, S.; Opitz, P.; and Engl, S. 2020. Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification. In Proceedings of The 12th Language Resources and Evaluation Conference, 4933–4941.

[34] Sun, C.; Huang, L.; and Qiu, X. 2019. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, 380–385.

[35] Eleni Kalamara, Arthur Turrell, Chris Redl, George Kapetanios and Sujit Kapadia, Making text count: economic forecasting using newspaper text, Bank of England, Staff Working paper No. 865.

[36]

[37]PekkaMalo, Ankur Sinha, Pekka Korhonen, Jyrki-Wal lenius, and Pyry Takala. 2014. Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 65(4):782–796.

[38] Macedo Maia, Siegfried Handschuh, André Freitas, Brian Davis, Ross Mcdermott,Manel Zarrouk, Alexandra Balahur, and Ross Mc-Dermott. 2018. Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon, France, April 23-27, 2018. ACM. https://doi.org/10.1145/3184558