Next Cashtag Prediction on Social Trading Platforms with Auxiliary Tasks

Chung-Chi Chen*, Hen-Hsen Huang†‡, Hsin-Hsi Chen*‡
*Department of Computer Science and Information Engineering National Taiwan University, Taipei, Taiwan
†Department of Computer Science National Chengchi University, Taipei, Taiwan
‡MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
cjchen@nlg.csie.ntu.edu.tw, hhhuang@nccu.edu.tw, hhchen@ntu.edu.tw

***Abstract*—Social trading platforms provide a forum for investors to share their analysis and opinions. Posts on these platforms are characterized by narrative styles which are much different from posts on general social platforms, for instance tweets. As a result, recommendation systems for social trading platforms should leverage tailor-made latent features. This paper presents a representation for these latent features in both textual data and market information. A real-world dataset is adopted to conduct experiments involving a novel task called next cashtag prediction. We propose a joint learning model with an attentive capsule network. Experimental results show positive results with the proposed methods and the corresponding auxiliary tasks.**

***Index Terms*—interest prediction, social trading, joint learning**

I. Introduction

Providing attractive information for potential audiences is an important topic; application scenarios appear in e-commerce and online advertising. In this work, we aim to leverage textual data in social trading platforms to understand user interests, and propose a novel joint learning model to infer audience preferences with auxiliary tasks.

A social trading platform provides a forum for investors to share their analysis results of financial instruments such as stocks and foreign exchange. Investors use cashtags, tags which start with the $ sign, to mark the mentioned financial instruments. For example, $AAPL stands for the stock of Apple Inc., and $USDJPY stands for the exchange rate of US Dollar and Japanese Yen. Given the current trend of financial technology (FinTech), more and more natural language processing researchers are focusing their attention on financial social media [1]–[3]. However, most previous work focuses on mining opinions from social trading platforms. In this paper, we propose personalized next cashtag prediction, a novel task for social trading platform users. We formulate this task in detail in Section III.

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08
http://dx.doi.org/10.1145/3341161.3342945

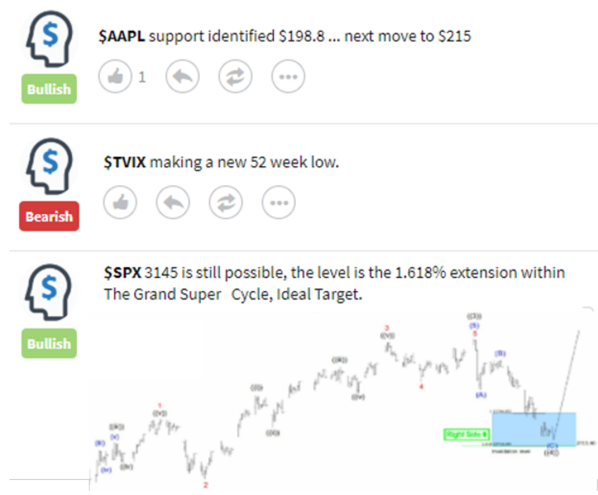

Fig. 1. Financial tweets in financial social media

Fig. 1 shows examples of financial tweets in financial social media. It is intuitive to recommend information on $AAPL to an investor who mentions $AAPL. However, only recommending the mentioned cashtag to users is not sufficient. According to our dataset statistics, within five days of a given day, 80.22% of users mention at least one cashtag that they did not mention in the previous week. This indicates the importance of exploration in the proposed task. For the investor who posted the second financial tweet in Fig. 1, instead of information on $TVIX, it would be better to recommend stocks whose prices have fallen to a 52-week low. This example shows the investor providing the focus—a 52-week low—of their financial tweets. For the investor who posted the third financial tweet in Fig. 1, the price chart of the mentioned cashtag could be a hint that reveals the interest of the investor. Along this line, this paper explores both textual information and price information to represent the latent features of target cashtags for the task of predicting the cashtag that could next interest investors.

II. Related Work

Recommendation systems have been widely discussed in the last decade. Several overview papers [4]–[7] analyze the development of recommendation systems from different aspects. In particular, Camacho et al. [7] review papers related

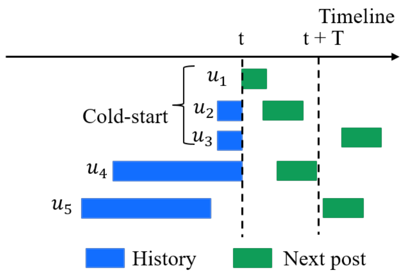

Fig. 2. Task design

to the cold-start problem from 2011 to 2017. Li et al. [8] propose a behavior-intensive neural network for next-item recommendation, and show that considering both preference and session behavior of users is useful for cold-start cases. Halder et al. [9] regard recommendation as a multi-label classification problem, and use bi-directional gated recurrent units (GRU) to take into account recommending never-seen-before items. Inspired by previous work, we use textual data to represent the preferences of the target investor and the dynamic representation of the target cashtags. The proposed neural network architecture also considers the recommendation task as a multi-label classification task.

Joint learning models have been successful in several machine learning tasks [10]. One recent highlight is the BERT model [11], which learns sentence embeddings via two auxiliary tasks and achieves state-of-the-art performance on the GLUE benchmark [12]. In this paper, we adopt BERT pre-trained embeddings and propose a joint learning model with two tailor-made auxiliary tasks. Section IV presents the details of the proposed ACN model.

III. Problem Formulation

We propose a next cashtag prediction task for social trading platforms. In this task, we attempt to predict the next cashtag given personal preference information. Fig. 2 illustrates the possible three cases in the proposed task. Cases u_1, u_2, and u_3 are considered cold-start cases because few historical data of the user are available. In this paper, we focus on cases such as u_2, who posts at time t and also posts in the time window between $t+1$ and $t+T$. That is, only one day (time t) of historical data of the user is used in our setting. We use the post(s) of $user_i$ at time t to represent his or her preferences. With the help of this information, we predict the cashtags that $user_i$ will mention within next T days. T is set to 5.

IV. Methods

A. Information Representation

In our models, we use both textual information and price information to represent user interests, the analysis results of the users in Stocktwits, and the market information of the candidate cashtags.

1) Textual Information: To represent word information, we use the pre-trained uncased BERT-large model [11] to extract the embeddings for the masked word. We mask the target word and extract the last 1024-dimension layer output as the word embedding. To represent a financial tweet, we average each dimension of the word embeddings of the financial tweets.

As shown in Fig. 1, users of Stocktwits describe interest patterns or signals in their posts. Along this line, we convert the interests of $user_i$ into a vector by averaging each dimension of the posts from the user at time t. This embedding is used to represent his or her interests and preferences.

To represent the candidate cashtag information, we sort the financial tweets at time t related to $cashtag_j$, and also average each dimension of the financial tweets. This embedding is used to represent the analysis results and the view of users toward $cashtag_j$. Finally, we use a matrix to represent the latent features of candidate cashtags from a textual aspect.

2) Price Information: As shown in Fig. 1, price information provides the clues for cashtags, and can be linked to the textual information. Previous work [13], [14] shows that market data such as price and volume provide useful information for tasks concerning financial data. Along this line, we concatenate the normalized open, high, low, close, and volume data from $t-w$ to t to represent the market information of the cashtag. A tensor with the market information vectors of the candidate cashtags is used to represent the market information of the past w days. In this paper, w is set to 30.

B. The Joint Learning Model

We propose two auxiliary tasks for co-training with next cashtag prediction, the major task. These two auxiliary tasks are used to imitate the intention of investors:

- **Hottest cashtag prediction** (A_{hot}): Predict which cashtag has the highest growth rate of mentions within the following T days.
- **Most profitable cashtag prediction** (A_{profit}): Predict which cashtag has the highest return within the following T days.

1) Attentive Capsule Network: To complete the proposed task, we develop the ACN model to learn the language model of analysis vectors and the pattern of the chart vectors. The ACN architecture is composed of four major layers: an attention layer, a convolutional neural network (CNN), CaspNet [15], and a BiGRU (bi-directional gated recurrent unit) layer. The attention layer is used for weighting important latent features in input embeddings, the subsequent CNN layer and CapsNet are used to shrink the feature matrix, and the BiGRU layer captures feature information from both directions.

2) Joint Learning Processing: We concatenate the two outputs from ACN with the interest vector of $user_i$ to share the cashtag information and the target user preferences. We use a fully connected layer as the last layer of the model, and output the prediction results for the three tasks. We use the same loss function as in previous work [16]. To avoid overfitting, early stop is triggered with a patience setting of 10 epochs.

TABLE I
EXPERIMENTAL RESULTS AND ABLATION ANALYSIS

Model	*hit*@2	*hit*@3	*hit*hit@5	*Precision*@2	*Precision*@3	*Precision*@5	*Recall*@2	*Recall*@3	*Recall*@5
Joint ACN	**69.03%**	**74.01%**	**80.33%**	**45.50%**	**38.23%**	30.91%	35.70%	**41.44%**	**50.13%**
-Attention	67.30%	71.95%	78.04%	43.44%	36.40%	29.35%	33.71%	39.03%	47.19%
-CapsNet	66.90%	72.05%	78.87%	43.62%	36.70%	30.03%	33.59%	39.09%	48.07%
-A_{hot}	68.96%	73.78%	80.24%	45.42%	38.12%	**30.96%**	**35.71%**	41.17%	50.02%
-A_{profit}	68.74%	73.60%	79.89%	45.19%	38.01%	30.84%	35.49%	41.06%	49.74%

V. EXPERIMENTS

A. Dataset Construction

We use data collected from Stocktwits, a widely used social trading platform. We collected the data from 288,400 users over 244 days, from May 27, 2018 to Feb 23, 2019. A total of 1,036,365 financial tweets were collected. In this paper, we focus on cases such as as u_2 in Fig. 2, and use all component stocks in the Dow Jones Industrial Average Index (DJI) as the candidate cashtags for prediction. We separate the data into the training and test sets by Jan 3, 2019.

B. Experimental Results

For evaluation, we used hit@k, Precision@k, and Recall@k as the evaluation metrics. Table I shows the experimental results. Joint ACN is the full proposed model and yields the best overall performance. We conducted an ablation analysis to show the contribution of the attention layer and the CapsNet in the proposed task. The full model outperforms the other models under hit@k evaluation metrics. The results show the effectiveness of the proposed architecture. We conducted another ablation analysis to verify the contribution of each auxiliary task. Although the performance of the Joint ACN model and the model without A_{hot} are similar, the proportions of errors with hit@k evaluation is significantly different on our test set: 21,538 instances under McNemar's test with $p < 0.05$ when $k \in \{2, 3\}$. This shows that the performance can be improved via the proposed auxiliary tasks.

VI. CONCLUSION AND FUTURE WORK

In this paper, we present preliminary results for the next cashtag prediction task in the cold-start case. Experimental results show the effectiveness of the proposed ACN model and the usefulness of the corresponding auxiliary tasks. In addition to recommending interesting information for users, the proposed task can be extended to user grouping, dynamic interest prediction, and market information prediction such as price movement prediction and market volatility prediction.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Science and Technology, Taiwan, under grants MOST-106-2923-E-002-012-MY3, 108-2218-E-009-051-, MOST-108-2634-F-002-008-, and MOST 107-2218-E-009-050-, and by Academia Sinica, Taiwan, under grant AS-TP-107-M05.

REFERENCES

[1] K. Cortis, A. Freitas, T. Daudert, M. Huerlimann, M. Zarrouk, S. Handschuh, and B. Davis, "Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, pp. 519–535.

[2] C.-C. Chen, H.-H. Huang, Y.-T. Shiue, and H.-H. Chen, "Numeral understanding in financial tweets for fine-grained crowd-based forecasting," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2018, pp. 136–143.

[3] S. E., L. Yang, M. Zhang, and Y. Xiang, "Aspect-based financial sentiment analysis with deep neural networks," in *Companion Proceedings of the The Web Conference 2018*, 2018.

[4] P. Rastogi and V. Singh, "Systematic evaluation of social recommendation systems: Challenges and future," *dimensions*, vol. 7, no. 4, 2016.

[5] F. Aznoli and N. J. Navimipour, "Cloud services recommendation: Reviewing the recent advances and suggesting the future research directions," *Journal of Network and Computer Applications*, vol. 77, pp. 73–86, 2017.

[6] P. Rahayu, D. Sensuse, B. Purwandari, I. Budi, F. Khalid, and N. Zulkarnaim, "A systematic review of recommender system for e-portfolio domain," in *Proceedings of the 5th International Conference on Information and Education Technology*. ACM, 2017, pp. 21–26.

[7] L. A. G. Camacho and S. N. Alves-Souza, "Social network data to alleviate cold-start in recommender system: A systematic review," *Information Processing & Management*, vol. 54, no. 4, pp. 529–544, 2018.

[8] Z. Li, H. Zhao, Q. Liu, Z. Huang, T. Mei, and E. Chen, "Learning from history and present: Next-item recommendation via discriminatively exploiting user behaviors," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1734–1743.

[9] K. Halder, L. Poddar, and M.-Y. Kan, "Cold start thread recommendation as extreme multi-label classification," in *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018, pp. 1911–1918.

[10] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the conference of the North American chapter of the association for computational linguistics*, 2019.

[12] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Glue: A multi-task benchmark and analysis platform for natural language understanding," *arXiv preprint arXiv:1804.07461*, 2018.

[13] S. Murakami, A. Watanabe, A. Miyazawa, K. Goshima, T. Yanase, H. Takamura, and Y. Miyao, "Learning to generate market comments from stock prices," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2017, pp. 1374–1384.

[14] T. Aoki, A. Miyazawa, T. Ishigaki, K. Goshima, K. Aoki, I. Kobayashi, H. Takamura, and Y. Miyao, "Generating market comments referring to external resources," in *Proceedings of the 11th International Conference on Natural Language Generation*, 2018, pp. 135–139.

[15] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in Neural Information Processing Systems*, 2017, pp. 3856–3866.

[16] C.-C. Chen, H.-H. Huang, and H.-H. Chen, "Numeral attachment with auxiliary tasks," in *The 42nd International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 2019.