CrowdPT: Summarizing Crowd Opinions as Professional Analyst

Chung-Chi Chen
Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
cjchen@nlg.csie.ntu.edu.tw

Hen-Hsen Huang
Department of Computer Science, National Chengchi University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
hhhuang@nccu.edu.tw

Chia-Wen Tsai
Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
cwtsai@nlg.csie.ntu.edu.tw

Hsin-Hsi Chen
Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
hhchen@ntu.edu.tw

ABSTRACT

This paper demonstrates a novel analytics service, CrowdPT, for capturing the key information, price target (PT), of individual investors on social media. PT, which is mentioned as a conclusion in most of analysts' reports, indicates not only the market sentiment (bullish/bearish) of investors, but also the analysis results. In order to provide the latest opinions of individual investors, we monitor Twitter in real time and update the information in price chart daily. For all component stocks in Dow Jones Industrial Average, textual information from numerous tweets is summarized into a single number, PT, in CrowdPT. Case studies confirm the effectiveness of our analytics service in the financial domain, and show that capturing the PT of individual investors is promising for stock price prediction. The Web API of CrowdPT is also provided for academic purpose.

CCS CONCEPTS

• **Information systems → Information extraction**.

KEYWORDS

price target, financial social media, crowd opinion

ACM Reference Format:
Chung-Chi Chen, Hen-Hsen Huang, Chia-Wen Tsai, and Hsin-Hsi Chen. 2019. CrowdPT: Summarizing Crowd Opinions as Professional Analyst. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3308558.3314122

1 INTRODUCTION

The application of artificial intelligence technologies to financial domain has attracted much attention in recent years. A variety of related tasks and researches are addressed in the shared tasks and workshops collocated with predominant conferences, including SemEval-2017 Task 5, FiQA-2018, FNP-2018, ECONLP-2018, and FinNum-2019.

Figure 1: Example of professional analyst's report.

According to the important information embedded in the numerals in financial narratives, it has caught many researchers' eyes [1, 2, 4]. Especially, forecasting the price target (PT), the price level that the price of a financial instrument will reach, is one of the major goals of analyzing financial instrument. As shown in Figure 1, most of professional analysts show their PTs in the top of the first page of analyst reports for the customers to capture their analysis results at a glance. In the left side of Figure 1, the analyst uses a piece of short text to summarize the report, and mentions the PT. In the right side of Figure 1, the analyst shows the analysis results in a table with the information such as the overweight comment and the PT. However, social media data is unstructured, and the PT information in tweets could only be extracted with natural language processing technique. Since PT is an extremely crucial indicator in the procedure of financial analysis, capturing the PT based on the crowd opinions is an important topic to investigate.

Sometimes the institutional investors may not release the report for a certain company. At this time, the price target forecasted by individual investors can be used to complement the missing information. For example, Bloomberg does not record the price targets of the analyst for $AXP in August, October, and November in 2017. With the PT sorted out by our system, the information can be filled up. Even though PT is quite meaningful for presenting the opinions and analysis results of investors, to the best of our

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3314122

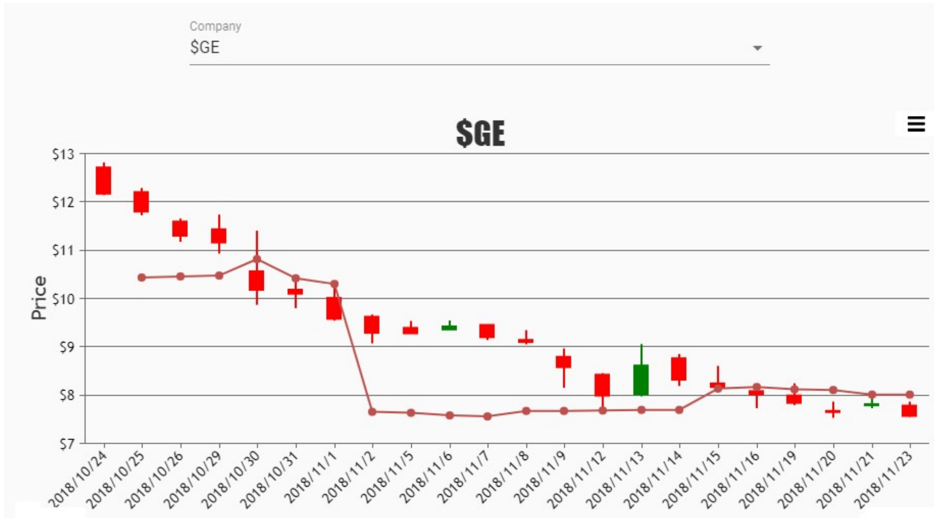

Figure 2: Screenshot of CrowdPT.

Table 1: Examples of subcategories.

Subcategory	Target Numeral	Tweet
price target	14.35	$CIEN, CIEN seems to have broken out of a major horizontal resistance. Targets $**14.35**.
support or resistance	46	$CTRP, $**46** Breakout Should be Confirmed with Wm%R Stochastic Up
buying	137.89	$SPY Long 1/2 position **137.89**
selling	36.50	$KOG - hopefully a better outcome than getting kneecapped by $BEXP selling itself dirt cheap at **36.50**
other	110.2	$FB (**110.20**) is starting to show some relative strength and signs of potential B/O on the daily.

knowledge, none of the previous systems provide the function for extracting the PTs of individual investors. Once the PT is extracted, the user gets not only the market sentiment, but also the degree of the market sentiment, calculating by the difference between the PT and the newest close price. Moreover, the expected value of a certain financial instrument is estimated. Many in-depth empirical experiments can be extended with the PT extracted from social media text.

This work presents CrowdPT, a service providing the PTs of individual investors based on social media text. For each stock, our system provides the PT extracted from social media text. Case studies show that PT from social media data is useful for investors when making decisions, and provides more insights for the behavior of investors.

Our contributions are three-fold: (1) We propose CrowdPT, a novel Fintech application based on natural language processing, and demonstrate its potential values. (2) In addition to the demonstration website, we also provide our model as a cloud service for downstream applications and extended research. (3) CrowdPT provides the real-time market information for all component stocks in Dow Jones Industrial Average. The case studies show the PTs of individual investors based on financial social media data help investors make investment decisions.

2 SYSTEM OVERVIEW

Figure 2 shows a screenshot of CrowdPT. The red bar and green bar, called candlestick chart, are used to show the daily open, high, low, and close price of a stock. The red line shows the average of the PTs from financial tweets, which can be easily compared with the market price shown in the candlestick chart. For those days on which no numerals are classified as PT, approximate PTs are shown with dotted lines. In Figure 2, we use $GE, General Electric Company, as an example. On Oct 25, the price of $GE was about 12, and the PT that we sorted out from the financial tweets was about 10.5. It shows that individual investors forecast the price will fall down to 10.5 in the future. On Oct 30, $GE's stock price came to the PT on Oct 25. Moreover, individual investors modified their PT to about 7 on Nov 2. Then, the price matched this PT on Nov 12, and descended to crowd's PT on Nov 15. These cases show the usefulness of crowd price target and indicate that CrowdPT provides the leading indicator for stock market prediction.

CrowdPT summarizes the textual information from financial tweets into single numeral (PT), and provides the real-time information for all component stocks in Dow Jones Industrial Average. The market sentiment, bullish or bearish, can be captured at a glance with the price chart. When the PT is higher than the close price, it means that individual investors take bullish market sentiment toward the stock. On the other hand, investors have bearish market sentiment to the stock when the PT is lower than the close price. Beyond market sentiment, CrowdPT provides more fine-grained information for individual investors' opinions. With the PTs sorted out, not only the degree of the market sentiment but also the expected price level would be captured.

Additionally, we provide the Web API for disambiguating the meaning of the numerals in financial tweet. We follow the taxonomy tailor-made for the numerals in financial social media data [1]. Categories and subcategories are shown in Figure 3. Table 1 shows the example tweets for each subcategory.

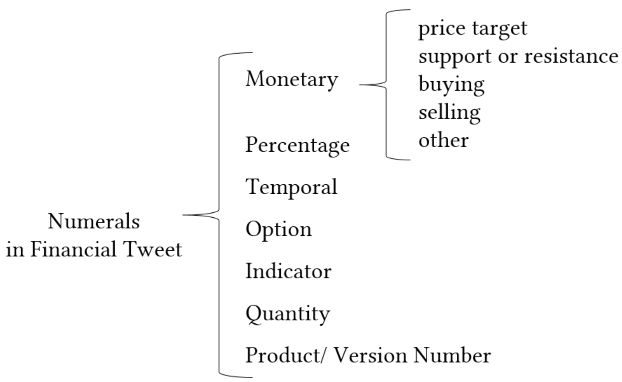

Figure 3: Numeral taxonomy in financial tweets.

3 TECHNICAL IMPLEMENTATION

In CrowdPT, we adopt the word-based CNN model for financial numeral understanding [1]. Here, we merge the subcategories that do not imply any subjective opinions of investors, including "money", "quote", "change", and the rare subcategory, "stop loss" in Monetary category into one class, called "other".

3.1 Dataset

We adopt FinNum dataset for experiments [1]. Three experts in the financial domain are involved to annotate the numerals in the SemEval-2017 Task 5 dataset, in which financial tweets from Twitter and StockTwits are collected and labeled with market sentiment. Total 1,341 numerals are annotated. The average agreement of the annotation on price target achieves kappa value of 65.59%, which is considered as substantial agreement [3].

3.2 System Flow

Figure 4 illustrates the flowchart of price target extraction. Firstly, we use tweepy [1], a Python library, to monitor the latest tweets published in Twitter. Only the tweets mentioning the cashtags of the constituent stocks of Dow Jones Index will be crawled. Secondly, the price target filter is performed at the end of the day (GMT -8) to sort out the price target for each cashtag. Then, a classifier based on convolutional neural network (CNN) is adopted for extracting the price target. Finally, we collect the close price from Yahoo Finance

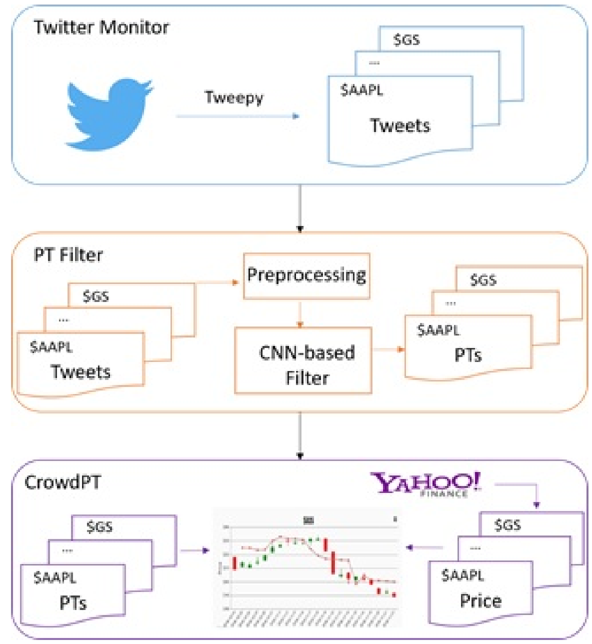

Figure 4: System flowchart of price target extraction.

Table 2: Experimental results of subcategory classification.

	target	sup/res	buying	selling	other
P	42.50	41.34	36.80	35.36	50.66
R	50.90	30.68	27.29	54.05	80.53
F1	45.32	34.27	28.55	40.17	61.51

with the fix_yahoo_finance [2] and render the price chart. Our system also provides the analysis results in json format via Web APIs.

3.3 Results of Numeral Disambiguation

In the 7-way category classification task, our word-based CNN model achieves an accuracy of 67.61%, which significantly outperforms other models with RNN structures under 95% confidence level. Please refer to our previous work [1] for more details of model setting and comparison. In the subcategory classification task, the precision, recall and F1-score for each subcategory are shown in Table 2.

4 DETAILS OF WEB API

Our Web API returns the classification results of the numerals in input financial tweets. A financial tweet should contain at least one cashtag, and our target tweets are those financial tweets that have at least one numeral. In this paper, we demonstrate the usefulness of the extracted PT, and more empirical studies could be extended based on other (sub)categories.

The participants of the financial market can be separated into two groups, institutional investors and individual investors. Yet, only the

[1]http://www.tweepy.org/

[2]https://github.com/ranaroussi/fix-yahoo-finance

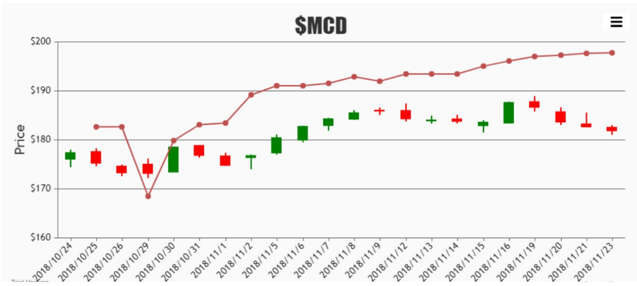

Figure 5: Correlation of stock price and crowd's PT.

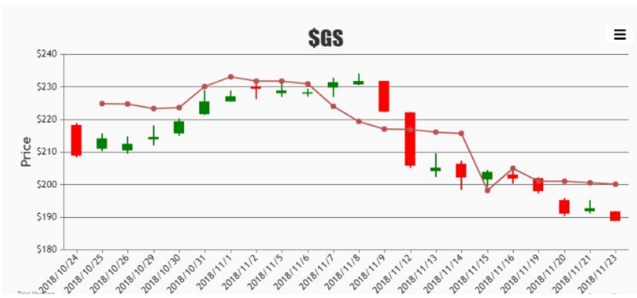

Figure 6: Evidence for crowd's PT as leading indicator.

analysis results of institutional investors are recorded in structural format by market information providers like Bloomberg. None of previous systems attempts to summarize the textual information on the financial social media data into price target. In CrowdPT, the latest PT from individual investors can be obtained and be compared with the latest close price in the provided price chart.

5 INVESTIGATION OF CROWD PRICE TARGET

We perform backtesting of the PT of both institutional and individual investors in 2017, and discuss our findings as follows. The performances of both institutional investors and individual investors can be compared within CrowdPT. For example, the price of $DWDP reached the September price target of institutional investors, but did not reach their November price target until now. In contrast, the price did not reach the September price target of individual investors, but reached their November price target. That shows the information of both institutional investors and individual investors may be comparable. With CrowdPT, investors can obtain more insights than before and develop new trading strategies.

Figure 5 shows the correlation between stock price and crowd's PT. The price did not match the PT during November, but the close prices follow the trend of PT. In other words, the crowds upgrade the PT first and the close price of $MCD rises toward the PT. Figure 6 shows that CrowdPT not only provides the bullish information but also the bearish information. It also shows that the crowd's PT can be the leading indicator for stock market movement prediction. The cases in Figure 2, 5, 6 show the evidence that CrowdPT provide

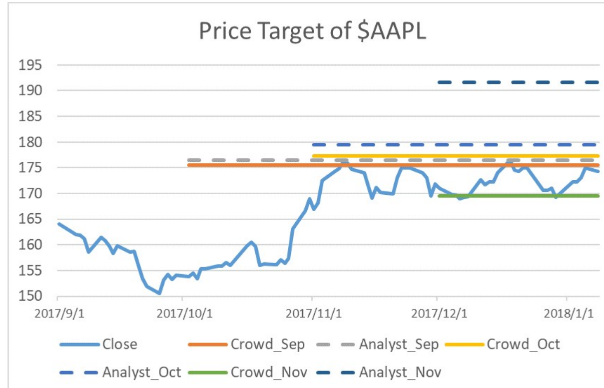

Figure 7: Price target of $AAPL.

the longer-term market movement forecast, which is more sensible for investment, and different from previous work [5] focus on speculation (predict the short-term price movement).

More empirical studies can be conducted with the analysis results of CrowdPT. For instance, the difference of the tendency of setting price target between institutional investors and individual investors can be investigated. Figure 7 shows the comparison of the price target of $AAPL in September, October, and November. We find that the price targets of the analysts are always higher than that of the crowd. That shows the analysts may be more optimistic than individual investors, or individual investors may tend to focus on the short-term prediction. According to this finding, investors could consider to follow the opinions of the crowd for short-term trading and follow the opinions of analysts for long-term investment.

6 CONCLUSION

We present CrowdPT [3], a service aiming at providing price-target oriented analysis results based on financial social media text. Novel applications based on our model are demonstrated. Investors can benefit from the information provided by our system. In particular, trading strategies combine the price targets of institutional investors and individual investors have great potential for further studies. With our Web API [4], users can get the fine-grained classification results based on the tailor-made taxonomy for numerals in financial social media data.

ACKNOWLEDGMENTS

This research was partially supported by Ministry of Science and Technology, Taiwan, under grants MOST 107-2218-E-009-050-, MOST-106-2923-E-002-012-MY3, MOST-107-2634-F-002-011-, 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] Chung-Chi Chen, Hen-Hsen Huang, Yow-Ting Shiue, and Hsin-Hsi Chen. 2018. Numeral Understanding in Financial Tweets for Fine-grained Crowd-based Forecasting. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 136–143.

[3] http://nlg6.csie.ntu.edu.tw/CrowdPT

[4] http://nlg6.csie.ntu.edu.tw/FinNum/?demo=CrowdPT&tweet= <input tweet>

[2] Matthew Lamm, Arun Tejasvi Chaganty, Christopher D Manning, Dan Jurafsky, and Percy Liang. 2018. QSRL: A Semantic Role-Labeling Schema for Quantitative Facts. In *Proceedings of the 1st Financial Narrative Processing Workshop*.
[3] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
[4] Soichiro Murakami, Akihiko Watanabe, Akira Miyazawa, Keiichi Goshima, Toshihiko Yanase, Hiroya Takamura, and Yusuke Miyao. 2017. 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. 1374–1384.
[5] Yumo Xu and Shay B Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1970–1979.