

CrowdIQ: A New Opinion Aggregation Model

Qianzhou Du
Virginia Tech
qiand12@vt.edu

Hong Hong
Xiamen University
honghongtju@126.com

G. Alan Wang
Virginia Tech
alanwang@vt.edu

Pingyuan Wang
Virginia Tech
fizer@vt.edu

Weiguo Fan
Virginia Tech
wfan@vt.edu

Abstract

In this study, we investigate the problem of aggregating crowd opinions for decision making. The Wisdom of Crowds (WoC) theory explains how crowd opinions should be aggregated in order to improve the performance of decision making. Crowd independence and a weighting mechanism are two important factors to crowd wisdom. However, most existing crowd opinion aggregation methods fail to build a differential weighting mechanism for identifying the expertise of individuals and appropriately accounting for crowd dependence when aggregating their judgments. We propose a new crowd opinion aggregation model, namely CrowdIQ, that has a differential weighting mechanism and accounts for individual dependence. We empirically evaluate CrowdIQ in comparison to four baseline methods using real data collected from StockTwits. The results show that, CrowdIQ significantly outperforms all baseline methods in terms of both a quadratic prediction scoring measure and simulated investment returns.

1. Introduction

Researchers have focused on making good decisions and improving the quality of predictive judgment in various domains, such as weather, finance, sports, culture, and economics [1, 6, 18]. However, individuals' judgments or predictions are always compromised with biases, since individuals' decisions are probably affected by overconfidence, emotionally available information, and ignorance among others [3, 19]. To reduce the above-mentioned biases, previous researchers propose a possible remedy, which mathematically aggregates multiple opinions or predictions from a group of individuals, e.g., knowledgeable experts and plain volunteers [5,

17]. Prior work often builds on the theory called "Wisdom of Crowds" (WoC) [18].

The WoC theory is applied to aggregating multiple opinions within a group in support of group decision making. Such an approach can outperform individuals, sometimes even domain experts, in various prediction and estimation tasks. Surowiecki claims that a mathematical or statistical aggregation over the judgments of a group of individuals can be more accurate than those of the average individuals because of the benefit of error cancellation [18]. Nowadays, the WoC theory has been applied to many domains including informed policy making and market prediction [12]. The aggregation of multiple sources is very persuasive and effective because it can maximize the information scope and reduce the potential impact of extreme or aberrant judgments. In addition, it increases the liability and validity of the aggregation methods.

However, simply aggregating individuals' judgments with a majority rule is not without shortcomings. One shortcoming has been pointed out by critics of prior crowd wisdom research. When the combined judgment is largely misrepresented by a systematic group bias or a large number of uninformed judges [16], the crowd wisdom model fails to deliver accurate predictions. To eliminate the flaws, some researchers propose a method to identify the expertise of each judge and aggregate only experts' judgments to make predictions [4]. Others propose weighted models where more experienced judges receive higher weights when aggregating individual judgments [2, 20]. Thus, in order to get the best prediction based on both the quality of experts and the quantity of the crowd, the most important issue of a crowd wisdom model is to precisely identify the expertise level of each individual in the crowd and assign appropriate weights to their judgments in the aggregation process.

Most of the existing crowd opinion aggregation methods assume independence in crowd opinions without considering the influence that often exist in

socially connected crowds, e.g., online communities, such as, Yahoo!Finance, SeekingAlpha, Amazon Reviews, CarComplaints, and StockTwits. These online communities have become popular venues for individuals to communicate their opinions with others [11, 4]. With the communities, individuals often read others' postings, reply to others, and follow experts, friends, and popular users. In this scenario, some individuals' judgments may be influenced by others when making predictions. Those original opinions should be given higher weights than those influenced by others. Hence, we need to take into account the influence between judges when we quantify the weighted contribution of each judgment.

In this paper, we propose a new opinion aggregation method, namely CrowdIQ, to evaluate crowd wisdom by considering crowd dependence. To address the dependency problem in individual judgments, we propose a decay function to give different weights to judges based on their previous judgments along with the timestamp of each judgment. Those judges who make early/original judgments are given higher weights than those who simply follow others' judgments. We evaluate CrowdIQ in a stock prediction task using user-generated stock prediction tweets. The result shows that CrowdIQ outperforms all baseline methods.

The rest of this paper is structured as follows. Section 2 introduces the theoretical background. In Section 3, we present the proposed crowd opinion aggregation method, CrowdIQ. We test our model in comparison to four baseline methods and show the evaluation results in Section 4. We conclude the paper by discussing the findings, implications, and future work in Section 5.

2. Theoretical Background

Surowiecki's WoC theory provides a seminal framework for the study of crowd performance [18]. In his work he mentions two conditions for having a wise crowd: (1) independence of crowd participants' decision making: people's opinions are not determined by others; (2) a good aggregation mechanism: it appropriately merges individual judgments into a collective decision by assigning a reasonable weight to each judge [18].

Surowiecki argues that a crowd of individuals is likely to make better predictions when working independently [18]. If individuals do not consider independently before expressing their own opinions, their judgments can be biased by responses from the crowd [13]. Lorenz's work demonstrates that even

mild social influence can undermine the effect of crowd wisdom in simple estimation tasks [14].

Most existing opinion aggregation models use the simple average of judgments as the aggregated opinion [13]. However, this approach probably is not optimal because it ignores judges' diversity (e.g., expertise) and may reduce the capacity to get benefits from the crowd wisdom. For example, underperforming stock market analysts are more likely to make hasty predictions that may cause an aggregated prediction to a worse position [9]. Lee et al. examines the effect of crowd wisdom using the data of "The Price Is Right" game show. And they find that the aggregation models, especially those taking into account strategies and bidding history, outperform all individual estimations [13]. Those who use external information sources outperform the simple average method. That evidence suggests that considering judges' expertise could improve crowd wisdom in the opinion aggregation process.

French points out that the concept of "expertise" is ill-defined and has many interpretations [10]. One possible approach is to assign weights based on one's professional status, education level, seniority, expertise ratings provided by others, or a combination of them. Another possible method is to calculate the weights empirically based on the experts' past judgments. Budescu et al. [4] have compared the performance of Cooke's weights method [7] with equally weighted linear pools and their own empirical model (Contribution Weighted Model (CWM)) using only the best experts' judgments. The result shows that CWM generally outperforms Cooke's method [4]. However, different scoring rules can lead to different weight assignments. Armstrong puts forward that diversified crowds can have a better performance than the experts only [1]. CWM may be suboptimal because experts tend to have similar background and information bias. Diversified crowds are more likely to overcome the bias problem.

Existing opinion aggregation methods, including CWM, evaluate each single judgment using a binary variable. A judgment can be either a true prediction or a false one. The binary nature of the measure sometimes is not enough to reflect the degree of deviation between a prediction and the true outcome. For example, a stock price prediction closer to the true price change will result in higher investment returns than one that is in the same direction but not adjacent to the true price. Logically, the two judgments should receive different weights. The binary judgment measure discounts one's true level of expertise, and thus reduces crowd performance.

In summary, we find two major problems in existing opinion aggregation methods. First, they

assume independence in individuals' judgments, which is unrealistic especially in online communities. Second, they fail to capture the level of expertise when assigning weights to individual judgments in the aggregation process. In the next section, we propose a new crowd opinion aggregation method, CrowdIQ that collectively consider influence among individual judgments and the level of expertise.

3. The CrowdIQ Method

CrowdIQ includes two procedures as shown in Figure 1, namely weighting procedure and aggregation procedure. The Weighting procedure describes how we weigh each judge based on historical performance. The aggregation procedure defines how a set of judgments is aggregated for decision making. Finally, we briefly describe a crowd performance measure.

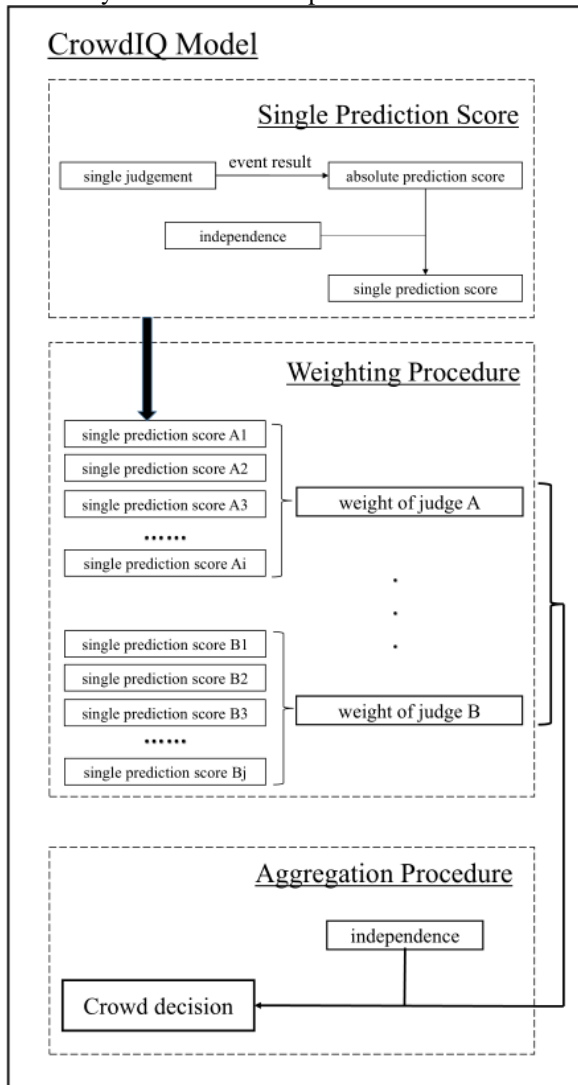


Figure 1. The CrowdIQ Method

3.1. Weighting Procedure

We define a judge's weight to be the judge's overall performance in all his/her past predictions. The weight reflects the judge's level of domain expertise. For each event prediction, an individual can get a prediction score based on the degree of correctness of a past prediction. The prediction score is not a binary value, but a continuous value normalized into a range between -1 and 1. An individual's past event predictions and scores can be summarized into an individual weight,

$$Individual_weight_j = f(\mu_j, \sigma_j),$$

where μ_j and σ_j are the numerical average and standard deviation of individual j 's all prediction scores, respectively. The individual weight represents not only an individual's average historical performance, but also the degree of consistency over time. To get a judge's weight, single prediction scores must be calculated for all his/her past predictions based on true outcomes.

3.1.1. Raw Prediction Score. Past research considers a prediction as being either correct or incorrect, i.e., a binary variable. We propose a raw prediction scoring method that evaluates the degree of correctness in a prediction using a real value. We assume that a prediction is made in the format of a text message such as an online posting. Given a prediction message, we determine the direction of the prediction using message sentiment and the degree of prediction correctness using a measure derived from the prediction consequence. For example, in the context of stock prediction, the prediction consequence will be the extent of the stock price change. We calculate the raw prediction score for each judgment as follows:

$$Raw_Prediction_Score_i = Opinion_i * \Delta Result_i,$$

where i denotes the judgment i , $\Delta Result_i$ (a continuous value normalized between [-1,1].) is the degree of correctness for judgment i . $Opinion_i$ is defined as follows:

$$Opinion_i = 1 \text{ if sentiment is positive, or} \\ -1 \text{ if sentiment is negative,}$$

The calculation of $\Delta Result_i$ is domain dependent. In the Evaluation section, we illustrate the calculation method for stock prediction using an online stock discussion community. Calculation in other domains must define a method appropriate for the target domain.

3.1.2. Independent Prediction Score. According to the WoC theory, if individuals do not make their judgments independently, their judgments can create biases and hurt the crowd performance. Unfortunately, in the online community context, individuals often read others' postings before making their own judgments. It is unlikely for all community participants to make judgments independently. Hence, removing the influence effect among judges is crucial for calculating an accurate prediction score closer to the reality. A judge, who makes judgments without being influenced by others, should receive a higher prediction score than one who is influenced by others' judgments. In this study, we propose a decay function to quantify the degree of influence as shown in Figure 2.

$$\text{Degree of Independence} = e^{[1-\lambda*(p-1)]},$$

where p is the order of judgment determined by its timestamp for the same type of judgments. For example, if an individual is the first one who predicts a bear market for a stock that day, his/her judgment order is 1 among all the bear predictions for that stock on the same day. λ is a decaying factor between 0 and 1, determining the decaying speed, i.e., the effect of the influence on subsequent judgments.

We can calculate an independent prediction score that accounts for the influence of earlier predictions as follows.

$$\text{Independent_Prediction_Score}_i = \text{Raw_Prediction_Score}_i * e^{[1-\lambda*(p-1)]},$$

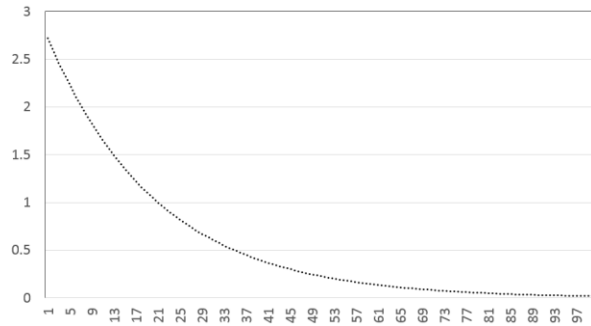


Figure 2. Decay Function ($\lambda = 0.05$)

3.1.3 Individual Weight. In the previous section, we showed how to calculate an independent prediction score for each prediction that a judge made in the past. We define a judge's overall individual weight as the following,

$$\text{individual_weight} = (\mu_j/\sigma_j),$$

which considers both the average prediction score (μ_j) and consistency of historical performance (σ_j) for a judge. According to this equation, a judge, who has a high average prediction score and high consistency (i.e., small σ_j), deserves a high weight.

3.2. Aggregation Procedure

The aggregation procedure is used to aggregate crowds' opinions toward a new event, in which the opinions are weighted based on the individual weights of the authors. We adopt a quadratic scoring method [8] to aggregate and quantify the aggregated crowd performance. Let N be the number of events forecasted, and symbol C_n be the number of prediction categories for event n (where $n = 1, \dots, N$). Additionally, we define W_{nc} as the aggregated crowd opinion score for outcome c (where $c = 1, \dots, C_n$). O_{nc} is a binary indicator that represents two possible event outcomes: true (i.e., the event occurred) and false (i.e., the event did not occur). Now the crowd's score of event n can be represented as follows.

$$S_n = a + b \sum_{c=1}^{C_n} (O_{nc} - W_{nc})^2$$

where $O_{nc}=0$ (event n occurs) or 1 (event n does not occur). W_{nc} is the aggregated probability of the crowd for each outcome and defined as follows.

$$W_{nc} = \frac{\sum_{j \in J_{nc}} \text{individual_weight}_j * e^{[1-\lambda*(p_{jnc}-1)]}}{\sum_{c=1}^{C_n} \sum_{j \in J_{nc}} \text{individual_weight}_j * e^{[1-\lambda*(p_{jnc}-1)]}}$$

where p_{jnc} denotes the temporal order of judge j 's judgment among all judges who predict outcome c for event n . We also consider judgment dependency when calculating the aggregated probability. Those judgments that are likely influenced by earlier judgments are discounted in the measure.

We use constants $a = 100$ and $b = -50$ as scaling parameters recommended in [4]. The score S_n ranges from 0 to 100. In this case, 0 indicates the worst performance (all judges' predictions are wrong) and 100 indicates the best performance (all judges' predictions are correct).

4. Evaluation

To evaluate the CrowdIQ method, we collected user-generated stock predictions extracted from a financial social networking community, StockTwits.

StockTwits has become the largest and most representative peer-based investment discussion community in recent years. It provides a social platform for investors to share their own stock analysis on financial securities. There are more than 10 million messages posted in StockTwits each year. Compared to SeekingAlpha, which is a crowd-sourced content service for financial markets, StockTwits has more active discussions for short to medium term investing strategies because there are more active users on StockTwits. Figure 3 shows a screenshot of user postings on StockTwits.

Users can post a prediction message for a particular stock ticker using a Hashtag \$, e.g., \$AAPL denotes that this prediction is related to the Apple stock. Moreover, unlike other financial social media, StockTwits allows users to post a message with an opinion label, either “Bullish” or “Bearish”. This unique feature provides a good opportunity for researchers to aggregate crowd opinions.

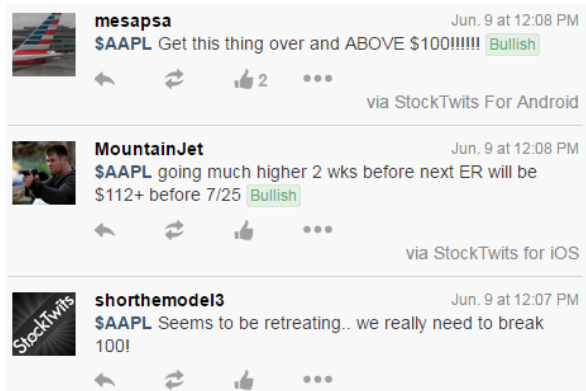


Figure 3. StockTwits Screenshot of AAPL

4.1. Data Collection and Processing

We collected approximately 11 million messages about 9,303 stocks, marketing indices, and exchange traded funds, for the year of 2014. Each message with an opinion label is regarded as a prediction. However, the statistics shows that only 16% of all messages include the opinion labels. To expand the judgments set, a reliable labeling tool is needed to extract opinions from the posted messages.

To classify the opinion label for those messages without a user specified label, we followed a previous work [15] and build a classifier using the unigram words in the messages as classification features. We removed the stop words, ticker symbols, company names, infrequent words from the message text to reduce the noise because they do not indicate an opinion. To train the classifier, we randomly sampled 10K messages labeled as “Bullish” and 10k messages

labeled as “Bearish” as our training data set. We used several commonly used supervised classification methods, such as Naive Bayes, Logistic Regression, and Support Vector Machine (SVM) and ran a 10-fold cross validation. Our experimental results showed that SVM had the highest accuracy (about 76%). Therefore, we chose SVM to build our final classifier. The opinion scores of StockTwits are continuous and range from 0 to 1 with 1 being absolute bullish and 0 absolute bearish. We arbitrarily chose 0.9 and 0.1 as our “Bullish” and “Bearish” thresholds, respectively. Therefore, any opinion scores equal to or above 0.9 (equal to or below 0.1) is considered as a bullish prediction (a bearish prediction). The messages with the opinion score between 0.9 and 0.1 are considered as neutral opinions and not considered in our evaluation.

For our evaluation, we only collected data for the top 11 tickers (namely AAPL, FB, GILD, KNDI, MNKD, NQ, PLUG, QQQ, SPY, TSLA, and VRNG) at StockTwits that had the most active discussions in 2014. In total there are 349,439 messages created by 13,933 individual users. Each event is defined as a prediction of the closing price for a stock on a trading day. We assume that crowd predictions made for a stock on a given day are the predictions of the closing price for the same stock three days later ($t+3$). For each day in 2014, we merged user prediction messages from the same judge for the same ticker into an average opinion score. We only considered those active judges who made predictions for 50 or more events (an event is a prediction made for a stock on a specific day). Finally, our sample contains 78,947 judgments created by 754 judges for 2,772 events. These 754 judges made predictions for 119 prediction events on average.

We also arbitrarily set the parameter of the independence factor $\lambda=0.05$, and the $\Delta Result_{t,s} = (Price_{t+3,s} - Price_{t,s})/Price_{t,s}$, where $Price_{t,s}$ is the closing price of stock ticker s on day t , and $Price_{t+3,s}$ is the closing price of stock s on day $t+3$. For example, when a judge predicts \$AAPL to be bullish, we then check the closing price of \$AAPL three days later. If the stock price of \$AAPL increases by 10%, this judge is given a score of 10 points for this raw prediction score; if the price of \$AAPL decrease by 5%, then -5 is the raw prediction score.

4.2. Performance Evaluation

To evaluate the performance of CrowdIQ, we chose four baseline opinion aggregation models summarized in Table 1. The first model, the unweighted mean model (UWM) [12] assumes that all judges have an equal weight. BWM and XBWM are weighted models where weights are determined by

binary prediction scores. The difference between BWM and XBWM is that XBWM uses only the top judges who perform better than average judge performance, while BWM uses all judges. CWM is built based on the quadratic scoring rule. Similar to XBWM, it only uses the top judges.

Table 1. Baseline Opinion Aggregation Models Compared to CrowdIQ

Model	Weighting	Aggregation
UWM	Equal weights for all judges.	Aggregates all judges' opinions
BWM	Weights depend on the judges' past binary prediction scores.	Aggregates all judges' opinions
XBWM	Weights depend on the judges' past binary prediction scores.	Aggregates opinions from the top judges.
CWM	Weights are calculated using the quadratic scoring rule.	Aggregates opinions from the top judges.
CrowdIQ	Weights are calculated using a fine-grained scoring rule.	Aggregates all judges' opinions

Table 2. Performance Comparison of All Models

Model	Mean	Median	SD
UWM	59.59	74.62	39.10
BWM	59.88	73.85	39.16
XBWM	65.71	75.0	37.47
CWM	67.98	75.04	33.29
CrowdIQ	72.69	81.85	27.69

To avoid overfitting, we evaluated all opinion aggregation models using the 10-fold cross validation method. Our sample data set was randomly split into 10 folds, among which 9 folds were training data and the other fold is for testing. The training data are used to compute individual judges' weights. A summary of the performance of all opinion aggregation models is shown in Table 2, in which the models are listed with their mean prediction score, median score, and standard deviation. We used the performance improvement metric defined in [3] as $(\text{difference between the two compared models' mean scores}) / (100 - \text{the baseline model mean score})$.

Table 3. Student T Test for Performance Differences

Diff. (p-value)	UWM	BWM	XBWM	CWM	CrowdIQ
UWM	-				
BWM	0.29	-			
XBWM	6.13***	5.83***	-		
CWM	8.39***	8.09***	2.26***	-	
CrowdIQ	13.10***	12.80***	6.97***	4.71***	-

Note: ***, **, and * denote p-value < 0.01, < 0.05, and < 0.1 respectively

As Table 2 shows, CrowdIQ outperformed all baseline methods. Specifically, it outperformed UWM and BWM by approximately 32% and XBWM by 17.2%. CrowdIQ also outperformed the state-of-the-art model, CWM, by as much as 12%. To show the statistical significance of performance differences, we ran the Student *T* test for each pair of model. Table 3 shows that the performance of CrowdIQ is significantly better than the four baseline methods. Therefore, we can conclude that our model statistically outperforms the four baseline models.

4.3. CrowdIQ vs. CrowdIQ-beta

Two unique features of our proposed opinion aggregation methods are a differential weighting mechanism and accounting for the dependence among judges. To show the usefulness of both features, we also consider a variant model, CrowdIQ-beta, which does not account for dependence among judges. We ran the same cross-validation and compared its performance to CWM and CrowdIQ. The performance measures and the Student *T* test result are shown in Table 4.

Table 4. Comparison of CWM, CrowdIQ and CrowdIQ-beta

Model	Mean	Median	SD
CrowdIQ	72.69	81.85	27.69
CrowdIQ-beta	71.40	80.59	29.06
CWM	67.98	75.04	33.29

(a)

Diff. (p-value)	CWM	CrowdIQ_beta	CrowdIQ
CWM	-		
CrowdIQ-beta	3.42***	-	
CrowdIQ	4.71***	1.29*	-

(b)

Note: (a) Basic statistics; (b) Student *T* test

The result shows that the proposed method with the differential weighting mechanism alone significantly outperformed the state-of-the-art method, CWM. In

addition, the complete CrowdIQ method also significantly outperformed CrowdIQ-beta. It shows that accounting for judges' dependence also helped improve the performance significantly.

4.4. Simulation for Stock Price Prediction

The quadratic scoring rule provides a mathematical method to measure the performance of opinion aggregation methods, but it does not directly compare the crowd wisdom models' capacity of stock price prediction. Hence, in this paper, we implement a simple stock trading strategy that takes advantage of valuable and predictive crowd wisdom on individual stocks. Our strategy builds a portfolio by distributing funds evenly to purchase 11 stocks at the beginning. Trading decision on each stock is made dependently on a daily opinion aggregation. For each stock in our portfolio, we sell all shares of the stock at the opening next day if the daily aggregated opinion is bearish. When the daily aggregated opinion is bullish, we hold the stock or buy it back if we sold it earlier. If the aggregated opinion is neutral, no action is taken. The return of our portfolio is the summation of all 11 stocks' returns.

In Figure 4, x -axis represents the trading day, and y -axis is the overall investment return rate of the 11 stocks. We simulate three crowd wisdom methods, CrowdIQ, CWM, and randomly trading strategy. To avoid the small probability event, we generate the data of randomly trading strategy by using the average net profit rate of 1,000 rounds calculations for each date. Unlike in the weighting procedure, our trading strategy simply uses $t+1$ policy in our simulation instead of $t+3$ policy. As shown in Figure 4, CrowdIQ outperforms CWM and a random trading strategy in terms of overall investment returns. Specifically, CrowdIQ, CWM, and the random trading strategy has a net profit rate of 46.28%, 40.59%, and 6%, respectively.

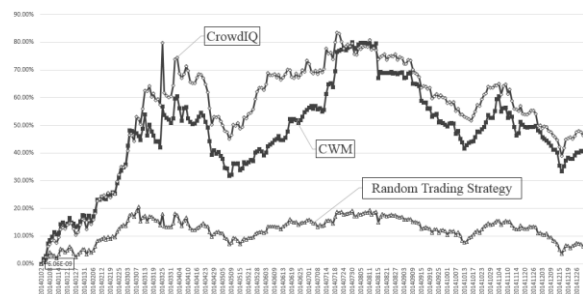


Figure 4. CrowdIQ vs. CWM vs. Random Trading Strategy

5. Conclusions and Discussions

In this study, we proposed a new crowd opinion aggregation model, namely CrowdIQ, which has a differential weighting mechanism and accounts for individual dependence. We empirically evaluated CrowdIQ in comparison to four baseline methods using real data collected from StockTwits. The results show that, CrowdIQ significantly outperformed all baseline methods in terms of both a quadratic prediction scoring method and simulated investment returns.

This study has inherent limitations. First the research sample is relatively small, and we will include more stocks and postings to improve the generalizability of our results. Second, although CrowdIQ has a significantly positive effect for crowd wisdom, there is still some room for further improvement. We plan to conduct more sensitivity analyses on the parameters used in CrowdIQ in future studies.

6. Acknowledgements

This research was supported by both Center for Business Intelligence and Analytics (CBIA) at Virginia Tech and the Natural Science Foundation of China (Grant# 71531013).

7. References

- [1] J. S. Armstrong, "The Seer-sucker Theory: The Value of Experts in Forecasting", *Technology Review* 82 (1980), pp. 16-24.
- [2] W. Aspinall, "A Route to more Tractable Expert Advice", *Nature*, 463 (2010), pp. 294-295.
- [3] J. R. Bettman, M. F. Luce and J. W. Payne, "Constructive Consumer Choice Processes", *Journal of Consumer Research*, 25 (1998), pp. 187-217.
- [4] H. Chen, P. De, Y. Hu and B.-H. Hwang, "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media", *Review of Financial Studies*, 27 (2014), pp. 1367-1403.
- [5] R. T. Clemen, "Combining Forecasts: A Review and Annotated Bibliography", *International Journal of Forecasting*, 5 (1989), pp. 559-583.
- [6] R. T. Clemen and R. L. Winkler, "Combining Probability Distributions From Experts in Risk Analysis", *Risk Analysis*, 19 (1999), pp. 187-203.
- [7] R. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, New York, 1991.
- [8] B. De Finetti, "Does it Make Sense to Speak of 'Good Probability Appraisers'", *The Scientist Speculates: An Anthology of Partly-baked Ideas* (1962), pp. 257-364.
- [9] T. Evgeniou, L. Fang, R. M. Hogarth and N. Karelaia, "Competitive Dynamics in Forecasting: The Interaction of

- Skill and Uncertainty", *Journal of Behavioral Decision Making*, 26 (2013), pp. 375-384.
- [10] S. French, "Expert Judgment, Meta-analysis, and Participatory Risk Analysis", *Decision Analysis*, 9 (2012), pp. 119-127.
- [11] B. Gu, P. Konana, R. Raghunathan and H. M. Chen, "Research Note—The Allure of Homophily in Social Media: Evidence from Investor Responses on Virtual Communities", *Information Systems Research*, 25 (2014), pp. 604-617.
- [12] R. Hastie and T. Kameda, "The Robust Beauty of Majority Rules in Group Decisions", *Psychological Review*, 112 (2005), pp. 494-508.
- [13] R. P. Larrick, A. E. Mannes and J. B. Soll, *The Social Psychology of the Wisdom of Crowds Krueger II, ed. Frontiers in Social Psychology: Social Judgment and Decision Making*, Psychology Press, New York, 2011.
- [14] J. Lorenz, H. Rauhut, F. Schweitzer and D. Helbing, "How Social Influence can Undermine the Wisdom of Crowd Effect", *Proceedings of the National Academy of Sciences of the United States of America*, 108 (2011), pp. 9020-9025.
- [15] B. Pang, L. Lee and S. Vaithyanathan, *Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, 2002, pp. 79-86.
- [16] J. P. Simmons, L. D. Nelson, J. Galak and S. Frederick, "Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds", *Journal of Consumer Research*, 38 (2011), pp. 1-15.
- [17] J. B. Soll and R. P. Larrick, "Strategies for Revising Judgment: How (and How well) People Use Others' opinions", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35 (2009), pp. 780-805.
- [18] J. Surowiecki, *The Wisdom of Crowds*, United States by Anchor Books, New York, 2005.
- [19] G. Thomas, D. Griffin and D. Kahneman, *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge, 2002.
- [20] G. Wang, S. R. Kulkarni, H. V. Poor and D. N. Osherson, "Aggregating Large Sets of Probabilistic Forecasts by Weighted Coherent Adjustment", *Decision Analysis*, 8 (2011), pp. 128-144.