SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5, ISSUE 1
ISSN 2054-7390

Stock Recommendations using Bio-Inspired Computations on Social Media

[1]Sophia Swamiraj and [2]Rajkumar Kannan
Department of Computer Science, Bishop Heber College (Autonomous),
Tiruchirappalli, India.
[1]ssophiababu@gmail.com, [2]rajkumar@bhc.edu.in

ABSTRACT

The tremendous growth of the social networks has paved way for social interactions of investing communities about a company's stock performance. Investors are able to share their comments on stocks using social media platforms. These interactions are captured and mined to produce advice on investing which helps retail investors to do prospective investments to increase profits. In this paper, we propose a novel stock recommendation methodology using Ant Colony Optimization (ACO). This method extracts sentiments from the investor's stock reviews and performs the sentiment analysis, which is optimized by the ACO. This method helps to find the correlation between sentiments and stock values, to make future stock predictions and to give stock recommendations to the retail investor.

***Keywords*:** Stock micro blogging, stock investment, recommendations, user generated content, opinion mining, swarm intelligence.

1 Introduction

Investors require as much information as possible about what's going on in the stock market. They need to do a lot of investigation before making an investment. They have to make the right choice of their investment depending upon their situation and requirements. A lot of small investors do it alone. They do their own research. Novice investors don't know where to begin or more particularly how to screen for stocks. In order to select an individual stock as an investment, investors first need a good source of potential investments. Investors need advice or trading recommendations on which stocks to buy and sell which accomplishes all the above.

The Internet, as a whole, has turned out to be an enabler that aggregates crucial information for stock investor decision making. It is altering how information is passed on to investors and the ways in which investors can take action upon that information [1]. In essence, it changes the way that investors invest trade, obtain and share information [2]. Initially, it was more of combining public information such as public news, financial data, and market updates. More recently, with the arrival of WEB 2.0 and social media [3], user generated content (UGC) are integrating private information in addition to public information [4]. Consequently we study how such virtual investing communities (VIC) as Yahoo Finance and Raging Bull are issuing relevant and valuable UGC data such as investment advice and proprietary analysis. UGC in these channels enriches investors' capability in making better investing decisions by

DOI: 10.14738/tmlai.51.2537
Publication Date: 16th January, 2017
URL: http://dx.doi.org/10.14738/tmlai.51.2537

letting investors to observe the thought process and decision makings of others. Thus it is essential for researchers and practitioners to comprehend how individuals in virtual communities interact with one another and how these behaviours relate to future predictive outcomes.

Social networking sites such as Facebook and Twitter collectively have hundreds of millions of users around the world, and are therefore an excellent venue for investors to disseminate their stock tips.

Recently social investing communities are slowly emerging as a platform to unite investors, help them to share information about stocks, and get them to work collectively to make more informed investing decisions. Several such investing decisions are aggregated to form a collective decision. Investment decisions are shared in the form of opinions. This research proposes an approach that takes investors' collective intelligence through their interactions with the contents, their contributions and finally suggests best investment recommendations.

Internet today is becoming more and more interactive and networked. Web 2.0 platforms such as social networks, wikis, and weblogs empower the people and transform their way of organizing within communities. Without any central control, self organizing groups arise in which members work together by following easy rules of communication and interaction. This new form of cooperation emerging from collective intelligent behaviour changes the process of information sharing and opinion formation [5]. In contrast to the top-down approach of information dissemination by classic media [6], opinions are transferred and formed in the bottom-up approach of Web 2.0 by social swarming.

Understanding the process of opinion formation in human swarms provides great potential for opinion research. Opinion development in swarms can be predicted or might even be manipulated. A new approach is introduced which by using text mining techniques identifies the opinions of single swarm members and analyzes opinion formation with regard to the underlying swarming behaviour by applying methods associated with swarm intelligence. Swarm intelligence being a discipline of artificial intelligence and intends at developing algorithms based on the swarming behaviour of social insects [7]. A new algorithm inspired by the commonly known ant colony optimization meta-heuristic is presented which allows the prediction of opinion development in human swarms. The approach is demonstrated by an exemplary online community in which opinions on stock market are exchanged and discussed.

2 Related work

2.1 UGC in Marketing

Our study closely associates an active area of research from the marketing discipline relating consumer behaviour to economic outcome. It is popular for scholars in this area to study consumer behaviour in the forms of customer ratings or electronic Word-of-mouth (eWOM), user reviews and blogs are very popular areas of study for scholars. The Internet's capacity to reach out to vast audience at low cost has presented a new significance for Word-of-mouth (WOM) as a means to influence and build trust [8].

2.2 UGC in Virtual Investing Communities

Virtual investing communities (VIC) are a reputable social media for online investors. It has blossomed with the growth of the Internet and its reputation stems from offering an environment where investors can collaborate and discuss, monitor what others are doing, or simply to seek out fellowship [9]. We peruse a few samples of studies undertaken to understand the relationship between behaviour of community participants and stock market outcomes.

One of the earlier studies in this area is from [10], which used a sample of 3,000 stocks on Yahoo! message board, and found that earlier day returns, changes in trading volume, and changes in previous day postings have no predictive capacity on stock returns. He found that an increase in volume of overnight postings correlated to a 0.18% average abnormal return. In adding up, he concluded that total posting volume is higher for firms with high short-seller action, accounting performance, excessive past stock returns and, higher price earnings and book-to-market ratios, higher past volatility and trading volume, higher market analyst following, and lower institutional holding [10]. In another study [4], using 181,000 postings from RagingBull.com found that, in general, message board activity does not forecast industry-adjusted returns or abnormal trading volume. However they found that it is likely to predict the number of postings using earlier day's trading volume, number of postings and weighted opinion [4].

A well-referenced paper [11], uses 1.5 million postings from Yahoo! Finance and RagingBull.com message boards, found important but negative simultaneous correlation between number of postings and stock returns on the next day. The return, however, is reasonably very small in comparison to transaction costs. Nevertheless, message posting actions do help to predict volatility and trading volume. In addition, the authors concluded that volume of postings is positively linked with volatility and bullishness. Similarly in [12], apart from verifying that day traders are noise traders, also found that day-trading volume increases volatility but concluded no predictive relationship with stock returns. Das and Chen [13] developed a methodology using five classifier algorithms to mine sentiment from stock message boards but found no significant predictive relationship between sentiment and stock prices. However, consistent with result of [11], [13] reaffirmed the reality of a noteworthy correlation between posting volume and volatility but asserted that sentiment does not predict stock movements. Interestingly, Das et al. [14], found that sentiment does not predict returns but instead returns drive sentiments. They inferred that members of virtual community are more likely to extrapolate past returns rather than to be contrarian, which ultimately leads to a behaviour consistent with the representativeness heuristic [14], [15], [16].

Sabherwal et al. [17] downloaded 160,000 postings from TheLion.com stock message board and conducted an event study to assess daily abnormal returns. The authors found that posting volume positively correlates with stock's abnormal returns on the same day and also forecast next day's abnormal returns. They concluded that online investors focused on sparsely traded micro-cap stocks with low institutional assets and low analyst coverage.

Overall, although significant relationship exists amid VIC activities such as posting volume and stock market movements, previous literature has yet to establish subsistence of any predictive power between sentiment and stock market outcome. This is the research gap we get to answer through investigating the relationship between sentiments of stock micro blog postings with prospective stock price movements.

2.3 UGC in Micro blogging

Although micro blogging is an emerging UGC channel, a few scholars have attempted to examine the relationship among predictors mined from micro blogs with future outcomes such as movie revenue, events and stock prices. For instance, Bollen et al. [18] extracted six mood dimensions from over 9 million Twitter postings using an extended version of Profile of Mood States (POMS). They combined mood components on a on a daily scale and evaluated them to the timeline of cultural, social, economic and political events in the same time period. They found significant relationship between extracted mood dimensions and those happening events. Bollen et al. [19] further widened their prior study in Bollen et al. [18] specifically towards forecasting DJIA index over the same time period and concluded an accuracy of 87.6%. Another instance is Asur and Huberman [20] which extracted sentiment from 6 million Twitter postings to predict box office income for movies. They benchmarked versus Hollywood Stock Exchange (HSX) and achieved an accuracy of 0.94.

Research correlating predictors in micro blogging with future outcomes is still in its early years. In this study we seek to understand the relationship between sentiments of stock micro blogs with future stock price performances.

2.4 Agent based models

Agent-based models offer computational models to simulate how communications among individuals lead to the emergence of a group organization [21]. It is based on the findings that the behaviour of a group cannot be explained by the independent behaviour of individuals ([22], [23]). Interaction between the individuals results in a high-level organization which crystallizes without awareness of the individuals. Often, they are even unable to assess their own actions and opinions ([24], [25]). Agent-based models are applied to explain the cooperative behaviour in the fields of organization, contagion and cooperation [21]. For instance, Schelling [26] simulates how individual movements according to neighbourhood similarity lead to the development of segregated groups. Axelrod [27] shows how the individuals' adoption of neighbourhood behaviour brings forth global polarization. Berger and Heath [28] study how ideas are spread during discussion boards depending on environmental signals. Rosenkopf and Abrahamson [29] express how innovations diffuse across groups with regard to reputational and informational influences. Sakamoto et al. [30] examine how individual choices develop under varying group influences within online communities. The method presented in this paper can be considered as an agent-based model which simulates opinion formation by social swarming. The method of opinion formation is also based on the interactions among individuals and the orientation towards neighbouring discussion partner. Alternatively, our approach differs in aim and method. We aim to address the phenomenon of opinion formation by using an ant algorithm from swarm intelligence.

2.5 Swarm intelligence

Swarm intelligence offers problem-solving algorithms which are inspired by the swarming behaviour of animals and which can be transcribed to human behaviour. Swarm intelligence consists of two major meta-algorithms: Ant colony optimization and Particle swarm optimization [7]. Ant colony optimization is motivated by the foraging behaviour of ants. It enables the incremental explaining of discrete optimization problems. Digital ants find solutions by following pheromone trails which indicate the quality of an uncertain solution. Particle swarm optimization imitates the behaviour of birds searching for food. It allows population-based solving of continuous optimization problems. Solutions correspond to birds (particles) which are flying around the solution space by following the best birds so far. Two types of algorithms have been developed in the past for solving conventional optimization problems such as time scheduling and route planning. These days new challenges such as data mining and Web mining are being faced. Lots of papers describe ant colony optimization algorithms and particle swarm optimization algorithms for data clustering [31], data classification [32], feature selection [33] and fuzzy-rule induction [34]. Web mining involves the development and application of such algorithms in a few researches. Abraham and Ramos [35] present a cluster algorithm for identifying Web usage patterns. Ujin and Bentley [36] propose an algorithm which leads online shoppers and visitors to interesting Web sites by personal recommendations based on their preferences. Jensen [33] describes an algorithm for categorizing Web pages based on their topic. Palotai et al. [37] introduce an algorithm which finds news on the Internet. On the other hand, so far there are no algorithms for analyzing the swarming behaviour of Web users during the evolutionary process of opinion formation.

3 Proposed methodology

The proposed approach aims at optimizing various opinions from the individual investors using ACO for stock recommendation where the accurate predictions and recommendations are accomplished by extracting sentiments from the opinions by classifying the individual opinions and optimizing them. The architecture of the proposed approach is shown in figure 1.

The proposed approach consists of two important steps:

1. Opinion classification
2. Opinion optimization using ACO

The algorithmic steps of the proposed approach are as follows.

1) Collecting the tweets
2) Pre-processing the text - tokenize into words, removing special chars, stopwords, stemming, bigrams, trigrams, n-grams
3) Sentiment classification of the pre-processed tweets based on word features.
4) The classified tweets are given as inputs to Ant algorithm.
5) The ant algorithm optimizes the classified tweets to give "buy", "sell" and "hold" signals.

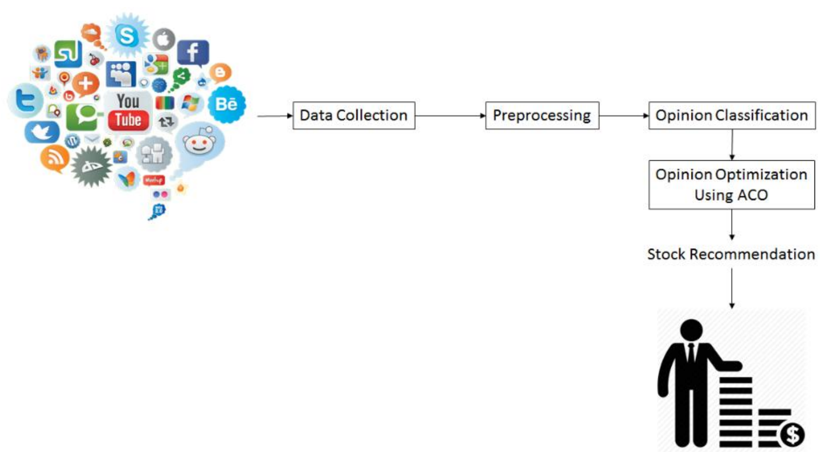

Figure 1. Architecture diagram of the proposed approach

3.1 Opinion classification

The goal of opinion mining is to identify the attitude of single swarm members towards an entity mentioned in postings. Attitudes are classified according to their polarity as “positive”, “negative”, or “no opinion”.

There are two steps involved in opinion mining, the extraction of features from the text and the application of a learning algorithm to identify the polarity of the text [38]. The extraction of features comprises a collective linguistic and statistical analysis. First a posting is decomposed into single words. After removing insignificant words (e.g. “the”), the remaining words are reduced to their stem and their frequency is calculated. Those word stems which are especially typical for each of the classes (meaning that they appear often in one class but not often in others) are used as the features of the postings. Based on the extracted features, the postings are classified as positive or negative by a learning algorithm. In general, machine learning provides three categories of learning algorithms [39]. Supervised learning algorithms use input and required output data for learning to produce the accurate output data. Reinforcement learning algorithms are trained to generate actions by getting rewards and punishments [40]. Unsupervised learning algorithms, in contrast, accept no feedback and find patterns within the input data which can be used to produce output data [39]. Supervised and unsupervised learning methods are often used for text classification [41]. Supervised learning requires more effort for pre-classifying texts (desired output) but enables enhanced classification results. It is, hence, employed in this approach.

Different supervised learning methods such as Naïve Bayes or Maximum Entropy can be used for text classification [42]. Support Vector Machines [43] are applied because of their ability to process a large number of features and their success in related projects [42]. Their input are sample data records which consist of various discussion postings with their features and classes. Support Vector Machines learn the constraints of a rule by analyzing the sample data which categorizes the postings best. The rule allows a binary classification. If there are three classes, three rules must be learned: "positive" versus "not positive", "negative'" versus "not negative", and "opinion" versus "no opinion". A posting will be assigned to the class which has the highest probability. In the simple two-dimensional case the rule can be described as a straight line (linear rule). Postings lying on one side of the line belong to the first class and those lying on the other side belong to the second class.

3.2 Opinion analysis using ant colony optimization approach

Complex tasks such as picking up objects or finding food can be achieved by colonies of social insects like ants, bees, wasps, and termites by means of cooperation. Swarm intelligence is the collective intelligent behaviour emerging from relatively simple interactions among colony members [44].

In general terms, swarm intelligence can be defined as an occurrence which arises from the social structure of interacting agents over a period of time if the sum of the problems solved collectively is higher than the amount of the problems solved individually [45]. Two preconditions must be satisfied in order for swarm intelligence to develop: The agents must interact with each other and must be capable of problem-solving [45]. Characteristics of emerging swarm intelligence are self-organization, robustness, and flexibility [44]. The members of the swarm interact without supervision or centralized control. The swarm is capable of achieving its task even if some members fall short and is able to adapt to a changing environment.

This phenomenon of collective intelligence is observed not only in the colonies of social insects but also in collaborative groups of humans. By, inspiring one another, correcting mistakes, and exchanging experiences collaborative groups are in a better position of solving problems than individuals [46]. Collaboration can be understood as an act of collective information processing [47]. Discussion is one of its basic forms [45]. Due to the collective process of exchanging information and opinions during a discussion, the total of the combined knowledge of the community becomes more valuable than the sum of the knowledge of all individual community members [48]. Web 2.0 platforms increase this effect of knowledge enhancement [49]. A wider range of people can connect more easily and more swiftly to reach a common opinion in an online discussion.

The Web also offers a benefit for opinion research. The process of opinion formation can be traced by applying mining techniques. A novel approach based on text mining and swarm intelligence is presented which is capable of analyzing the evolutionary method of opinion formation by social swarming. Text mining enables the recognition of opinions of single community members. An algorithm connected with swarm intelligence, especially the commonly known ant colony meta-heuristic, permits the prediction of the opinion trend during the collective intelligent process of opinion formation in online communities.

The aim of opinion analysis is to gain a better understanding of opinion formation in social swarms. With this knowledge opinion trends can be predicted and the process of opinion formation might even be manipulated.

Opinion analysis is inspired by the collective intelligent behaviour of living ants finding the shortest path between their nest and their source of food. This intelligent behaviour emerges from the ants' indirect manner of communicating by leaving and following pheromone trails in their environment – a phenomenon called stigmery [50]. While ants are moving about they drop chemical substances called pheromones on their paths. The more the same path is frequented, the more the pheromone intensity increases and the more likely this path will be followed by other ants. If the same path is only followed by a few ants, the pheromone intensity of the path decreases due to evaporation. As a result of this feedback loop, the probability that the path will be followed by an ant depends on the number of ants having taken this path before.

The ants' behaviour seems to resemble in some ways the behaviour of human swarming within online communities and can be used as a simplified model to simulate the process of opinion formation. Members of online communities communicate indirectly with each other by posting messages to a discussion thread. In their postings they can express positive or negative opinions. The more messages of the same opinion are posted, the more other people are attracted by this opinion and the more likely they are to follow this opinion.

The collective intelligent behaviour of ant colonies is also the basic idea of the ant colony optimization meta-heuristic, from which an algorithm for simulating the process of opinion formation can be derived. In ant colony optimization algorithms, possible solutions for a given problem are represented by paths [51]. If a path is followed by an ant, a certain amount of pheromones is deposited on it depending on the quality of the solution. Evaporation gradually decreases the pheromone amounts on those paths which are not traversed frequently. This means that the corresponding solution is not particularly appreciated.

When simulating the process of opinion formation the problem is to predict the polarity of the next posted opinion in a discussion thread. Possible solutions are represented by two different paths: one for positive and one for negative opinions. An ant predicts the next posted opinion by following the corresponding path and drops a certain amount of pheromones on this path depending on the correctness of the prediction. Evaporation depends on the sequence of postings in the thread and leads to a reduction of the pheromone amount on the path of the less frequently mentioned opinion. The ant is more likely to predict the opinion class whose corresponding path has a higher amount of pheromones.

In general, the ant colony optimization meta-heuristic comprises the following components [44]:

- A heuristic function which evaluates the quality of the solution found by an ant
- A rule for pheromone updating which describes how to reinforce pheromones on paths

- A rule for pheromone evaporation which specifies how pheromones on paths diminish over time
- A decision function which finds solutions by considering the value of the heuristic function and the amount of pheromones on paths.

In order to build up an algorithm based on ant colony optimization for predicting opinions in online swarms, these components must be specified and integrated into a procedure.

Figure 2 shows the flowchart of the developed procedure. First, all variables are initialized. The pheromone values of both paths representing the positive and negative opinions are given equal amounts of pheromones. While the discussion is going on, the opinion trend in terms of the next posted opinion class is predicted by an ant. The ant predicts the opinion class by choosing a path according to the decision function. As soon as the next message is posted to the thread its content is checked. If no opinion is expressed in the posting evaporation takes place. However, if the posting contains an opinion the correctness of the ant's prediction is evaluated. In case the predicted opinion differs from the posted opinion, the average error ratio is increased. Otherwise the average error ratio is decreased. Thereafter, the heuristic values of the decision function are adjusted depending on the dynamics of the opinion discussion. In addition, the pheromone value of the path representing the predicted opinion is updated. Finally, evaporation takes place which decreases the pheromone values of all paths.

According to the procedure, the functions for decision making, pheromone updating and evaporation have to be defined. The decision function determines the predicted opinion at time *i* by comparing the weighted sum of pheromones on the positive (τ_p^i) and negative path (τ_n^i). It is implemented as a signum function based on the difference of the weighted pheromone sums of both paths (positive and negative opinions), By this means the opinion of the path with the highest weighted sum of pheromones is predicted.

$$di = sgn\left(n_i^p \tau_i^p\right) - \left(n_i^n \tau_i^n\right) \tag{1}$$

If the decision function yields the value 1, a positive opinion is predicted by the ant. In the case of value -1 a negative opinion is forecasted. The pheromone sums of both paths are weighted by the iteratively computed heuristic values η_p^i and η_n^i. The heuristic values are incremented by a certain factor x if the actual opinion (O_i) equals the previous one $(O_i - 1)$ and decremented by x if both opinions are different. By doing so, the dynamics of opinion changes during opinion prediction is taken into account.

$$\eta_i = \begin{cases} \eta_i - 1 + \ x, if\ O_i = O_i - 1 \\ \eta_i - 1 - \ x, if\ O_i \neq O_i - 1 \end{cases} \tag{2}$$

If a lot of consecutive messages with no opinion are posted to the thread, the influence of the last posted opinions on the future opinion trend becomes insignificant. In such a case, the prediction should be based on all opinions posted to the thread. In the ant algorithm evaporation leads to a rapid diminution in pheromone values of the negative and positive path if there is a sequence of postings without opinions. A minimum-rule derived from the Max-Min Ant System of Stützle and Hoos [52] is implemented to change the prediction basis in this case. According to this rule, the decision function predicts the future opinion trend based on the opinion class most frequently mentioned in the thread, if the pheromone values of both paths fall below a minimum value. The rule is also inspired by the biological archetype of behaviour of real ants. If the amount of pheromones on a path cannot be smelled any more, the ants rely on their instinct when choosing a path.

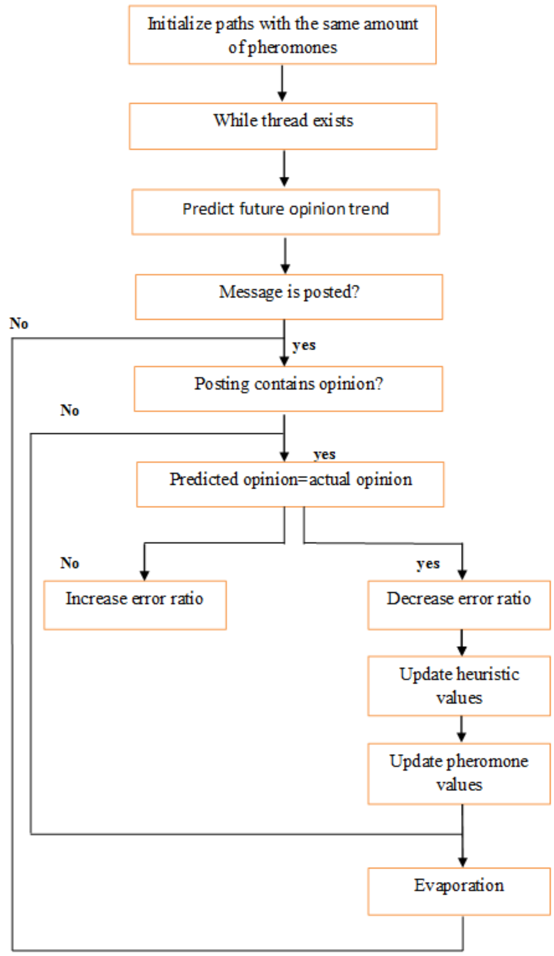

Figure 2. ACO Algorithm Flowchart

In order to reinforce predicted opinions the pheromone values of the corresponding paths are updated. Pheromone updating is realized by adding a predefined amount of pheromones ρ to the current pheromone value of the selected path.

$$\tau_i = \tau_i + \rho \tag{3}$$

Since the opinion trend in a discussion can change from time to time more weight should be added to recent opinions than to past opinions. Pheromone evaporation enables this weighting. It is realized by multiplying the pheromone value τ_i with a certain factor e<1.

$$\tau_{i+1} = \tau_i e \tag{4}$$

The above equations are followed from [53] which incorporate an algorithm based on ant colony optimization for predicting opinions in online swarms.

4 Experiments and Results

4.1 Data

The primary data for this study was downloaded from Stocktwits.com (http://www.stocktwits.com) for the period May 21 2016, to April 30, 2016 (22 days). We obtained over 7,140 stock micro blog postings for the company NETFLIX.

Stock micro blog postings were pre-processed; those without any ticker, more than one ticker, or not in NASDAQ exchange were removed leaving 2,140 valid postings for testing and 5000 postings for training. A list of top 6 stock tickers with corresponding number of postings is shown in table 1 while a description of all attributes is in table 2. Interestingly, top 10% of all the stock tickers are responsible for over 70% of all postings. These are popular stocks, consistent with the finding that people invest in the familiar while often ignore principles of portfolio theory [54].

Table 1. Distribution of postings by top 6 tickers.

S. No	Ticker	Total	Exchange
1	AAPL	7212	NASDAQ
2	AMZN	6220	NASDAQ
3	BBRY	2502	NASDAQ
4	GOOG	1803	NASDAQ
5	MSFT	1576	NASDAQ
6	NFLX	7140	NASDAQ

Table 2. Description of posting attributes [55]

Variable	Description
Sentiment	0-neutral , 1-bullish, -1-bearish (manually labelled)
Posting	Posting id, post date, day of the week, time of the day, market hours, text of posting.
Author	Expert, bio, url, location, follower, following, total postings, posting per day, retweet, direct, mention, etc.
Ticker	Exchange, volume, past 7 days closing prices and volumes.
Market	Past 7 days NASDAQ index.

4.2 Experiments on opinion classification

For validation, opinions posted to the stock investment community of Stocktwits.com were classified. The result of opinion classification is shown in the table 3. Stocktwits.com is the online platform of stock investment community. 5000 postings were extracted and assigned to the three classes “positive”, “negative”, and “no opinion” by a human annotator.

In order to examine the classification results a stratified ten-fold cross validation is applied. This means that all postings are divided into ten equally sized parts containing the same proportions of class labels. There are ten validation loops. In each loop nine parts are used for learning the classification rules and the remaining part for testing the classification rules learned. After ten runs the average precision and recall are calculated. While precision describes how many of the recognized opinions are correct, recall shows how many of the opinions are really recognized. The results of validation are shown in table 4. They indicate that learning was more successful for positive opinions and negative opinions than for neutral opinions. Lessons learned from misclassification show that quite often postings are not recognized as neutral if they contain several positive arguments or negative information but a neutral introduction or a neutral conclusion. This problem is planned to be solved by attaching more weight to the words at the end and the beginning of the postings.

Table 3. Results of opinion classification

Class	No of tweets	Accuracy
Positive	1379	100%
Negative	170	86%
Neutral	591	67%

Table 4. Precision and recall

Class	Precision	Recall
Positive	0.586	0.898
Negative	0.724	0.104
Neural	0.576	0.368

4.3 Experiments on opinion analysis and recommendations

In order to validate the ant algorithm for opinion prediction the financial communications platform for the investing community stocktwits.com was analyzed. Three companies APPLE, NETFLIX and GOOGLE in which community members discussed their opinions on stocks were extracted. All opinions mentioned in the postings were classified as ‘positive’, ‘negative’ or ‘no opinion’. Before applying the ant algorithm the parameters of the functions involved must be determined. Tests with different combinations of parameters revealed the following best set:

- Pheromone update ρ: 0.5
- Heuristic factor x: 0.3
- Minimum threshold: 0.0001
- Evaporation rate e: 0.8

Based on this set of parameters, the average error ratio is measured over the entire period of time. It is calculated as the fraction of all incorrectly predicted opinions to all opinions. Table 5 depicts average ratios for the three company discussions. The low error rate of 17.2% for the company “AAPL” indicates high prediction accuracy. The error rate for the company “GOOG” is higher indicating that prediction was less successful.

Table 5. Average error ratios

Discussion on Company	Amount of postings	Average error ratio
$APPL	1242	17.2%
$NFLX	2140	13.26%
$GOOG	2460	41.3%

Besides the average error ratio of the entire period, the development of the error ratio over time is important as well. It reveals whether the decision function enables incremental learning. Figure 3 and 4 depict the error curves associated with the discussion on the company “NFLX”. The descending error curve indicates a successful learning process. This effect results from the heuristic values which adapt incrementally to the dynamics of the discussion.

The necessity of the minimum-rule becomes apparent when looking at the development of the pheromone amounts on the positive and negative paths over time. For example, Figure 5, 6, 7 and 8 show the pheromone curves associated with the discussion on the company "NFLX". 12.5% of the pheromone values fell below the minimum threshold so that prediction was based on the most frequent opinion class.

In addition, the pheromone development shows the underlying swarming behaviour of opinion formation. The amounts of pheromones decrease over time which indicates a levelling in discussion. At the same time opinions are getting more homogeneous. During discussion a clear opinion trend emerges from the initially differing opinions of the swarm members. At the end of the discussion there are only a few opinions differing from the overall opinion trend which have less influence. This can be interpreted as a sign of robustness of the swarming behaviour. Application results of the ant algorithm show that prediction is more successful for discussion threads in which the length of the sequences of equal opinions vary to a high degree.

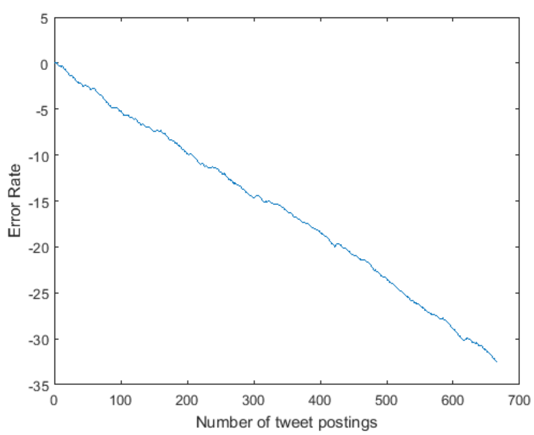

Figure 3. Normal Ant error ratio graph- 666 tweets - $NFLX

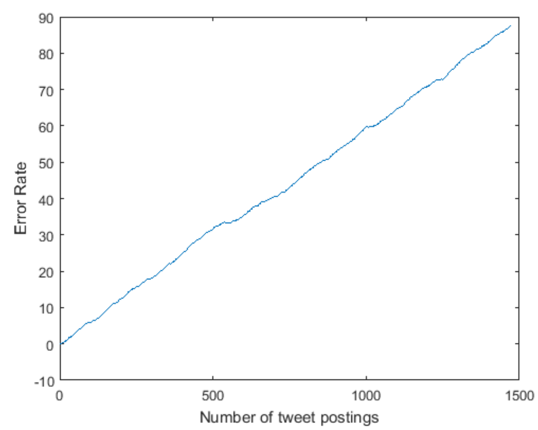

Figure 4. Normal Ant error ratio graph-1475 tweets - $NFLX

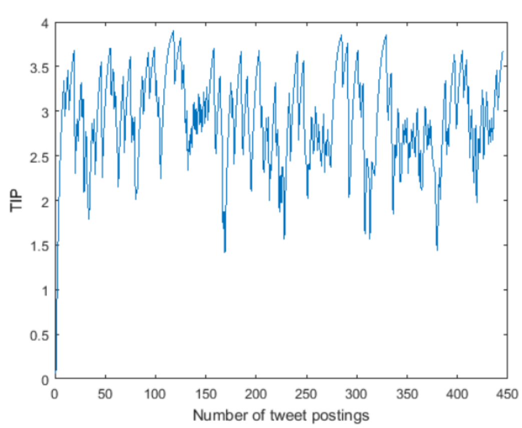

Figure 5. Pheromone graph TIP - 666 tweets

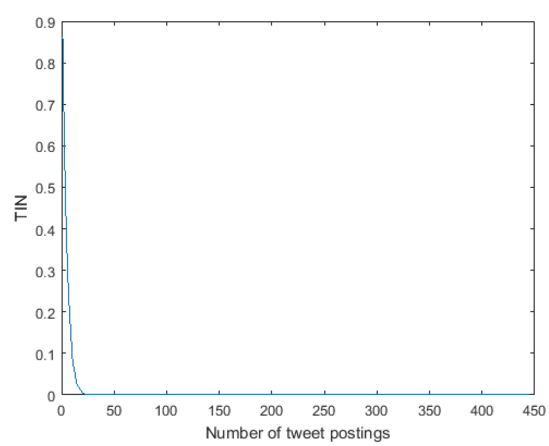

Figure 6. Pheromone graph TIN – 666 tweets

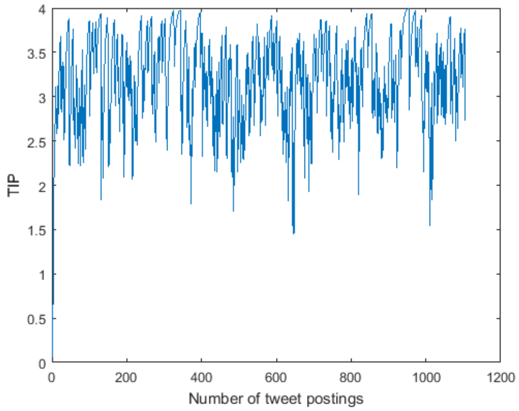

Figure 7. Pheromone graph TIP -1475tweets

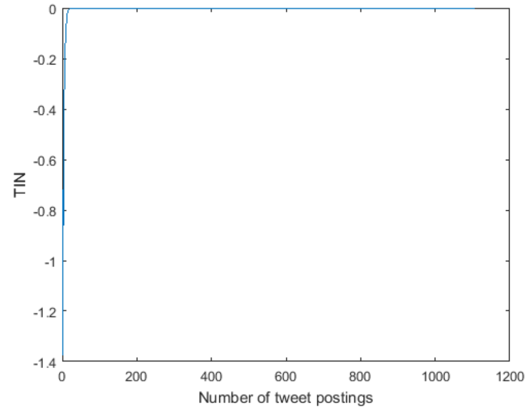

Figure 8. Pheromone graph TIN-1475 tweets

Table 6. Ant colony algorithm results for stock tweets of Netflix Inc.

		Ant algorithm output		
Time	Total tweets	Positive	Negative	Neutral
April 23,24,25	666	444	1	221
April 26,27,28,29	1474	0	1105	370

Table 7. Netflix Historical Stock Prices

Date	Open	High	Low	Close/last	Volume	*Adj close
Apr 29, 2016	90.50	90.56	88.21	90.03	13,968,000	90.03
Apr 28, 2016	91.50	92.67	90.09	90.28	11,474,900	90.28
Apr 27, 2016	92.18	92.50	90.21	91.04	12,218,900	91.04
Apr 26, 2016	93.50	93.55	91.25	92.43	15,330,900	92.43
Apr 25, 2016	95.70	95.75	92.80	93.56	14,985,400	93.56
Apr 22, 2016	94.85	96.69	94.21	95.90	15,806,300	95.90
Apr 21, 2016	97.31	97.38	94.78	94.98	19,919,400	94.98

As we observe from the above two tables 6 and 7 the ant's prediction proved to be accurate. The historical stock prices for the company Netflix are increasing for the dates April 22-25 for which in the above table 6 the ant result gives a positive sentiment prediction. For the dates April 26, 27, 28, and 29 the historical stock prices are declining and the ant accurately gives a negative sentiment prediction. Thus we are able to get the accurate sentiment predictions to give the accurate stock recommendation to "sell" the Netflix company stocks during this period of decline. Hence we prove the sentiments collected from the stock related tweets are able to guide the retail investor's decision making into profitable investments. Stock recommendations thus guided by sentiments are very useful for retail investors and to everyone in the investing community to increase profits and to decrease losses.

5 Conclusion

The outcomes are more accurate recommendations which best suit an individual trader amongst the multiple choices. Recommender systems can maximize investment returns in stock portfolio investments. Investment returns are enhanced with a reduction in trading losses using intelligent recommender systems. This research will result into a recommender system that allows the retail investor to save a lot of time in locating potentially profitable trading opportunities. The capability to scan the universe of stocks and only select the ones that meet their criteria in a matter of seconds is a huge advantage for the active trader. Stock investing recommender system provides the professional help to small investors so that they can select the right stocks for them and get good returns. Investors get ahead in investing; they not only earn great returns in the bull market but also minimize losses in the bear market.

REFERENCES

[1] Barber, B. M., & Odean, T. (2001). The internet and the investor. *The Journal of Economic Perspectives*, *15*(1), 41-54.

[2] Zhang, Y., & Swanson, P. E. (2010). Are day traders bias free?—evidence from internet stock message boards. *Journal of Economics and Finance*, *34*(1), 96-112.

[3] Ullrich, C., Borau, K., Luo, H., Tan, X., Shen, L., & Shen, R. (2008, April). Why web 2.0 is good for learning and for research: principles and prototypes. In *Proceedings of the 17th international conference on World Wide Web* (pp. 705-714).

[4] Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, *57*(3), 41-51.

[5] Rheingold, H. (2007). Smart mobs: The next social revolution. Basic books.

[6] Kolbitsch, J., & Maurer, H. A. (2006). The Transformation of the Web: How Emerging Communities Shape the Information we Consume. *J. UCS*, *12*(2), 187-213.

[7] Blum, C., & Li, X. (2008). Swarm intelligence in optimization. In *Swarm Intelligence* (pp. 43-85).

[8] Dellarocas, C. (2003) "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms" *Management Science,* 49(10):1407-1424.

[9] Wasko, M. & Faraj, S. (2005) "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice" *MIS Quarterly,* 29(1), 35-57.

[10] Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper*, (98025).

[11] Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259-1294.

[12] Koski, J. L., Rice, E. M., & Tarhouni, A. (2004). Noise trading and volatility: Evidence from day trading and message boards. *Available at SSRN 533943*.

[13] Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375-1388.

[14] Das, S., Martinez-Jerez, A. & Tufano, P. (2005) "eInformation: A Clinical Study of Investor Discussion and Sentiment" *Financial Management,* 34(3), 103-137.

[15] Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The journal of finance*, *49*(5), 1541-1578.

[16] Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

[17] Sabherwal, S., Sarkar, S. K., & Zhang, Y. (2008). Online talk: does it matter?. *Managerial Finance*, *34*(6), 423-436.

[18] Bollen, J., Pepe, A. & Mao, H. (2010a) "Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena" *Proceedings from 19th International World Wide Web Conference* Raleigh, North Carolina.

[19] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

[20] Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010,* (Vol. 1, pp. 492-499).

[21] Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in cognitive sciences*, *9*(9), 424-430.

[22] Axelrod, R. M. (1997). The complexity of cooperation: Agent-based models of competition and collaboration. Princeton University Press.

[23] Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. S*cience*, *311*(5762), 854-856.

[24] Loewenstein, G., & Schkade, D. (1999). Wouldn't it be nice? Predicting future feelings. *Well-being: The foundations of hedonic psychology*, 85-105.

[25] Bonds-Raacke, J. M., Fryer, L. S., Nicks, S. D., & Durr, R. T. (2001). Hindsight bias demonstrated in the prediction of a sporting event. *The Journal of social psychology*, *141*(3), 349-352.

[26] Schelling, T. C. (1971). Dynamic models of segregation†. *Journal of mathematical sociology*, *1*(2), 143-186.

[27] Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of conflict resolution*, *41*(2), 203-226.

[28] Berger, J. A., & Heath, C. (2005). Idea habitats: How the prevalence of environmental cues influences the success of ideas. *Cognitive Science*, *29*(2), 195-221.

[29] Rosenkopf, L., & Abrahamson, E. (1999). Modeling reputational and informational influences in threshold models of bandwagon innovation diffusion. *Computational & Mathematical Organization Theory*, *5*(4), 361-384.

[30] Sakamoto, Y., Sadlon, E., & Nickerson, J. V. (2008). Bellwethers and the emergence of trends in online communities. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1416-1421).

[31] Ye, F., & Chen, C. Y. (2005). Alternative KPSO-clustering algorithm. In *淡江理工學刊*, *8*(2), 165-174.

[32] Admane, L., Benatchba, K., Koudil, M., Siad, L., & Maziz, S. (2006). AntPart: an algorithm for the unsupervised classification problem using ants. *Applied Mathematics and Computation*, *180*(1), 16-28.

[33] Jensen, R. (2006). Performing feature selection with ACO. In *Swarm Intelligence in Data Mining* (pp. 45-73).

[34] Galea, M., & Shen, Q. (2006). Simultaneous ant colony optimization algorithms for learning linguistic fuzzy rules. In *Swarm intelligence in data mining* (pp. 75-99).

[35] Abraham, A., & Ramos, V. (2003, December). Web usage mining using artificial ant colony clustering and linear genetic programming. In *The IEEE Congress on Evolutionary Computation, 2003. CEC'03.*, (Vol. 2, pp. 1384-1391).

[36] Ujjin, S., & Bentley, P. J. (2003, April). Particle swarm optimization recommender system. In *Proceedings of the 2003 IEEE Swarm Intelligence Symposium, 2003. SIS'03.* (pp. 124-131).

[37] Palotai, Z., Mandusitz, S., & Lórincz, A. (2006). Computer study of the evolution of 'news foragers' on the Internet. In *Swarm Intelligence in Data Mining* (pp. 203-219).

[38] Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems* (No. 1). Oxford university press.

[39] Szuba, T. M. (2001). *Computational collective intelligence*. John Wiley & Sons, Inc..

[40] Hofstätter, P. R. (1986). Gruppendynamik, Kritik der Massenpsychologie; 3. *Auflage, Rowohlt, Reinbeck bei Hamburg*.

[41] Smith, J. B. (1994). *Collective intelligence in computer-based collaboration*. CRC Press.

[42] Kolbitsch, J., & Maurer, H. A. (2006). The Transformation of the Web: How Emerging Communities Shape the Information we Consume. *J. UCS*, *12*(2), 187-213.

[43] Johnson, N., Rasmussen, S., Joslyn, C., Rocha, L., Smith, S., & Kantor, M. (1998). Symbiotic Intelligence: self-organizing knowledge on distributed networks driven by human interaction. In *Proceedings of the 6th International Conference on Artificial Life* (pp. 403-407).

[44] Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). *Text mining: predictive methods for analyzing unstructured information*.

[45] Ghahramani, Z. (2004). Unsupervised learning, Bousquet O., Raetsch G., and von Luxburg U.(Eds.), Advanced Lectures on Machine Learning, LNAI3176.

[46] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, *4*, 237-285.

[47] Lee, C. H., Yang, H. C., Chen, T. C., & Ma, S. M. (2006, August). A comparative study on supervised and unsupervised learning approaches for multilingual text categorization. In *First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06)* (Vol. 2, pp. 511-514).

[48] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86).

[49] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

[50] Martens, D., De Backer, M., Haesen, R., Baesens, B., & Holvoet, T. (2006). Ants constructing rule-based classifiers. In *Swarm Intelligence in Data Mining* (pp. 21-43).

[51] Parpinelli, R. S., Lopes, H. S., & Freitas, A. A. (2002). Data mining with an ant colony optimization algorithm. *IEEE transactions on evolutionary computation*, *6*(4), 321-332.

[52] Stützle, T., & Hoos, H. H. (1996). Improving the Ant System: A detailed report on the MAX–MIN Ant System. *FG Intellektik, FB Informatik, TU Darmstadt, Germany, Tech. Rep. AIDA–96–12*.

[53] Kaiser, C., Krockel, J., & Bodendorf, F. (2010, January). Swarm intelligence for analyzing opinions in online communities. In *43rd Hawaii International Conference on System Sciences (HICSS), 2010,* (pp. 1-9).

[54] Huberman, G. (2001). Familiarity breeds investment. *Review of financial Studies*, *14*(3), 659-680.

[55] Oh, C., & Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement, In *Proceedings of ICIS 2011*. 17.