2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

An Extended Neuro-Fuzzy Approach for Efficiently Predicting Review Ratings in E-Markets

Giovanni Acampora, Georgina Cosma and Taha Osman

***Abstract*— Internet has opened new interesting scenarios in the fields of commerce and marketing. In particular, the idea of e-commerce has enabled customers to perform their transactions in a faster and cheaper way than conventional markets, and it has allowed companies to increase their sales volume thanks to a world-wide visibility. However, one of the problems that can strongly affect the performance of any e-commerce portal is related to the quality and validity of ratings provided by customers in their past transactions. Indeed, these reviews are used to determine the extent of customers acceptance and satisfaction of a product or service and they can affect the future selling performance and market share of a company. As a consequence, an efficient analysis of customer feedback could allow e-commerce portals to improve their selling capabilities and revenue. This paper introduces an innovative computational intelligence framework for efficiently learning review ratings in e-commerce by addressing different issues involved in this significant task: the dimension and imprecision of ratings data. In particular, we integrate the techniques of Singular Value Decomposition (SVD), Fuzzy C-Means (FCM) and ANFIS and, as shown in experimental results, this synergetic approach yields better learning performance than other rating predictors based on a conventional artificial neural network and FCM algorithm.**

I. Introduction

RAPID expansion of Internet and e-commerce portals is enabling several companies to improve their selling capabilities through a world-wide visibility and, at the same time, it is allowing more people to buy products in a much faster and cheaper way than conventional face-to-face markets.

In order to enhance customer satisfaction and shopping experience, it has become common practice for businesses and online traders to enable their customers to review or to express opinions on the products that they have purchased. Typically, a customer expresses his/her review about a product by typing a set of sentences and selecting an overall numerical rating for the transaction. Companies use customer feedback to provide new potential buyers with the opportunity to get an idea about the extent of acceptance and satisfaction of a given product or service. Consequently, reviews have an important impact on businesses performance and they can affect both sales volume and market shares of a company. For this reason, frameworks for the sentiment analysis of customers' ratings are becoming an important component of each e-commerce portal to extract customers' emotions and classify the related products or services as recommend or not recommend.

G. Acampora, G. Cosma and T. Osman are with the School of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS e-mails: {giovanni.acampora, georgina.cosma, taha.osma}@ntu.ac.uk

However, these systems suffer from several design drawbacks related to the nature of the information that characterises users' feedback: they can be vague, imprecise and, often the textual customers' feedback does not match the numerical rating. Moreover, the size of database storing customers' reviews could be too huge to be analysed by conventional machine learning algorithms such as neural networks. As a consequence, these systems should provide additional functionality to extract the most significant features from reviews and automatically compute a corresponding rating for given textual reviews. In this way, online traders can perform their future marketing activities on a collection of refined data that is more suitable and accurate than raw information directly provided by customers after their purchases. As shown in Section II, several intelligent systems have been designed to face this significant challenge, but none of them addressed the rating prediction issue by considering all of the involved difficulties: the *dimensionality of data* and *accuracy in prediction*.

This paper is devoted to introduce an innovative computational intelligence framework that, by integrating different intelligent methodologies, is able to efficiently reduce the size of textual reviews databases and analyse the human sentiments hidden in reviews. Due to this architecture it is possible to deploy an efficient predicting system supporting online companies in their marketing activities, for instance product/service evaluation, promotion of popular products or improvement of poor rated products. In particular, the proposed framework uses Singular Value Decomposition (SVD) and Dimensionality Reduction to extract the important and semantic features from each review and, consequently, reduce the size and complexity of the entire reviews database; successively, a clustering algorithm as Fuzzy C-Means (FCM) is used to classify the refined reviews into fuzzy clusters in order to create an initial collection of rating prediction rules; finally, this preliminary set of rules are used to start an ANFIS learning aimed at creating an optimised reviews rating prediction system.

As shown in experiments, where the framework has been tested on a well-known dataset[1], the proposed approach yields better performance than conventional neural networks and FCM algorithms.

The paper is organised as follows: Section II describes the most relevant research, Section III describes the proposed Computational Intelligence Architecture, Section IV

[1]Available at http://www.cs.cornell.edu/people/pabo/movie-review-data, Sentiment Scale datasets

discusses the evaluation of the proposed approach and the results, and a conclusion and outline of future work is provided in Section V.

II. Related Works

The task of rating a product or service using a rating scale is referred to as rating-inference, belonging to the research area of Opinion Mining. Machine learning approaches (ML), and 'supervised learning' approaches in particular, have been applied to review rating prediction, where these models are trained using a set of reviews and corresponding ratings (i.e. target outputs). The performance of these predictors is evaluated by comparing the target outputs with the actual outputs of the predictor [1], [2], [3]. Most recent predictors are mainly based on Artificial Neural Network (ANN) approaches.

A broad overview of existing literature on the topic of Opinion Mining (or sentiment analysis) is presented in Tang et al. [1], Pang and Lee [2], and Prabowo [3]. ML approaches have been applied to opinion mining classification and such techniques can achieve a level of accuracy comparable to that achieved by human experts [4]. Supervised ML approaches are known to be more suited for classification and prediction tasks returned better results than unsupervised approaches. Prabowo and Thelwall [3] propose a method of combining the rule-based classification and supervised learning approaches into a new combined approach. Ye et al. [5] compared three supervised ML algorithms (Naive Bayes, Support Vector Machine (SVM), and character-based N-gram model) for sentiment classification of online reviews on travel blogs. Their results revealed that the SVM and N-gram approaches outperformed the Naive Bayes approach. Moraes [6] compared ANNs with SVMs for the task of classifying reviews as positive or negative. Their results revealed that ANNs outperformed the SVMs, but also raise the potential limitations of these techniques such as the computational cost of SVM at the running time and ANN at the training time. Ganu [7] proposed methods for deriving a text-based rating from reviews and clustered similar users together using the topics and sentiments that appear in the reviews.

In order to try to overcome the performance of current systems for rating prediction, we propose a computational intelligence framework that is able to simultaneously tackle two main issues characterising this application: the dimensionality of data and accuracy in prediction. In particular, this goal is achieved by integrating different methodologies: SVD, dimensionality reduction, FCM clustering and ANFIS.

III. A Computational Intelligence Predictor for Customer Ratings in E-Markets

This section is devoted to introduce a computational intelligence framework for learning and predicting customer ratings in e-commerce scenarios. As aforementioned, a real design and implementation of a system for predicting ratings based on past reviews is affected by some main issues: the size, imprecision and noise in data. Indeed, as proved by some real case scenarios, a reviews database could contain many terabytes of information some of which maybe imprecise due to intrinsic human vagueness and errors. As a consequence, the proposed framework integrates some techniques to efficiently tackle these issues. In particular, as shown in Figure 1, the framework is composed of three different modules, each addressing a different problem involved in the task: the *Data Preparation module* which prepares the data, the *Input Selection module* which removes noise from the data to represent customers' reviews in a much reduced dimensional space, and the *Neuro-Fuzzy Predictor module* which efficiently applies a learning phase on the reduced dimensional space. Hereafter, a detailed description of each module is provided.

A. Data Preparation Module

Popular e-commerce portals, such as Amazon or e-Bay, are accessed and used on a daily basis by millions of people and this impacts the size of their reviews database repositories in a significant way. For this reason, the Data Preparation Module is aimed at preliminarily reducing the database size in order to decrease the computational complexity of mining the data. The first step towards reducing the size of data involves the application of Natural Language Processing techniques of tokenisation and stemming [8] to the collection of textual reviews contained in an e-commerce portal database. In particular, for each textual review, *Tokenisation* involves separating terms, and *Stemming* involves transforming variants of terms with the same root into the same term. In addition, terms that are solely composed of numeric characters, syntactical tokens (i.e. semicolons, and colons) and punctuation marks, terms that occur in only one review, and terms with length equal to one are all removed. Finally, upper-case letters are converted to lower case.

Once a collection of refined terms have been extracted form the original textual review, the Data Preparation Module creates the *Vector Space Model*, i.e. a term-by-review matrix $A_{m\times n} = [a_{ij}]$, in which each row i holds the frequency of refined term in textual reviews, and each column j represents a textual review. Hence, each cell a_{ij} of A contains the frequency at which a term *i* appears in a review *j*.

The next goal of Data Preparation module is to normalise the term frequency in matrix A by a different point of view. In particular, a *global weighting* function is used to adjust the frequency of textual review terms in respect to the entire collection of reviews. At the same way, a *document length normalisation* is applied to tune the frequencies based on the length of each review. In detail, the Data Preparation module uses the *normal global weighting* function named g_i and the *cosine document length normalisation* named n_j:

$$g_i = \frac{1}{\sqrt{\sum_j a_{ij}^2}}$$
$$n_j = (\textstyle\sum_i (g_i \cdot a_{ij})^2)^{-1/2}$$

where $a_{ij} = A[i, j]$.

After the Data Preparation module computes the normalisation step, each entry of the matrix A is updated as follows:

$$a_{ij} = a_{ij} \times g_i \times n_j$$

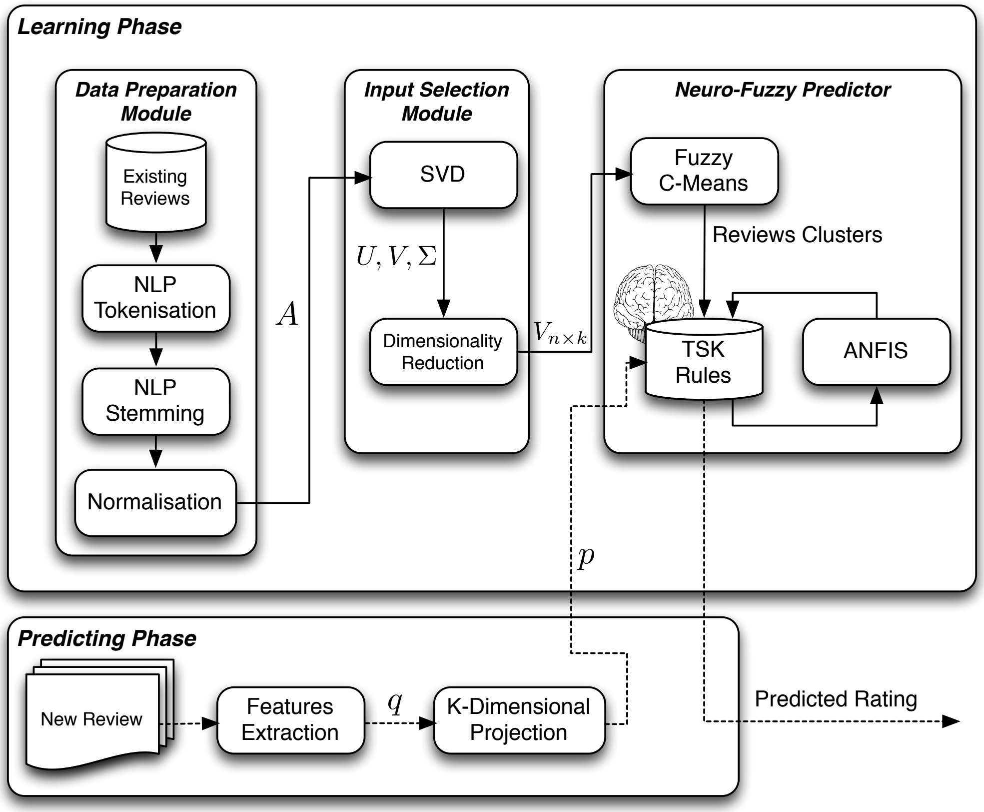

Fig. 1. Computational Intelligence System Architecture for Rating Prediction

with $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

The role of normalisation is crucial in the Data Preparation module because it enables the framework to capture information about the importance of each term in describing each textual review. In particular, if $a[i_1, j] \geq a[i_2, j]$ then the term associated to the row i_1 is more significant in describing the j^{th} textual review than the term related to the row i_2.

B. Input Selection Module

After the Data Preparation Module has performed an initial reduction of database size and it has computed the vector space model A, the Input Selection Module further reduces the space complexity by removing noise and irrelevant data. This task is accomplished by using some methodologies such as the SVD and the Dimensionality Reduction Statistical technique both described in [9]. The joint exploitation of these techniques enables the Input Selection Module to further reduce the computational time for training the Neuro-Fuzzy predictor module. Principal Component Analysis (PCA) could be another suitable technique for capturing the semantic information from reviews, and both PCA and SVD are eigenvalue methods used to reduce a high-dimensional dataset into fewer dimensions while retaining important information. The reason for using SVD over the PCA is because the resulting matrices upon decomposing using these two techniques are different, and SVD provides outputs which are more useful to the particular data mining problem of reviews analysis. In particular, SVD can be applied to a term-by-review matrix and return useful matrices including a term-by-dimension matrix, and a reviews-by-dimension matrix. These matrices can then be used for projecting new terms, and reviews into a reduced dimensional space, and also for computing the term-term, review-review, as well as query-review similarities. On the other hand, PCA returns the principal component coefficients for a term-by-review matrix, which is not exactly what is required for the particular data mining process.

As a consequence, the data stored in our computational intelligence predictor system uses SVD during data mining in order to allow efficient prediction using machine learning approaches appropriate for large data repositories.

SVD decomposes the normalised $m \times n$ matrix A into the product of three other matrices:

$$A_{m \times n} = U_{m \times r} \cdot \Sigma_{r \times r} \cdot V_{r \times n}$$

where U is an $m \times r$ term-by-dimension matrix, Σ is an $r \times r$ singular values matrix and V is an $n \times r$ reviews-by-dimension matrix and r is the rank of the matrix A.

The Input Selection Module completes its task by providing a rank-k approximation to matrix A, where k represents the number of dimensions (or factors) retained, and $k \leq r$. In order to achieve this goal, the Input Selection Module uses a process known as *dimensionality reduction*, which involves truncating all three matrices to k dimensions. The dimensionality reduction technique applies the Cattell's graphical Scree test [10] on the singular values contained in the Σ matrix in order to determine the optimal number of the value of k.

Because the Input Selection Module reduces the dimensionality of the original Vector Space Model A, when a new review is input into the system to be classified, it needs to be transformed into a term-by-review vector and opportunely projected to the reduced dimensional space V. Thus, given a review vector q, whose non-zero elements contain the normalised term frequency values of the terms, a new review vector p can be obtained from the projection of q to the k-dimensional space as follows [9]:

$$p = q^T \times U_k \times \Sigma_k^{-1} \tag{1}$$

Once projected, a review is represented as a review vector p of size k. This means that SVD and dimensionality reduction need only be recomputed once the size of the reviews dataset stored in the database increases significantly.

C. Neuro-Fuzzy Predictor Module

This module takes as input the reduced review-by-dimension matrix V and applies a neuro-fuzzy learning algorithm to opportunely train the rating predictor component of the proposed framework. The learning algorithm works in two sequential steps. In the first step, the FCM clustering is applied to generate a collection of textual review clusters where each cluster contains the review characterised by a similar collection of qualitative terms. Successively, by using the approach proposed by Sugeno and Yasukawa [11], a collection of TSK rules, one for each cluster, are generated for determining the membership of a review to a particular cluster. The second step uses this collection of rules as input to the ANFIS algorithm. ANFIS opportunely tunes the fuzzy rules and the related fuzzy membership functions in order to generate an optimised predictor model for textual reviews.

IV. A Case Study: Classifying Movie Reviews

This section explains the experiment carried out for testing the performance of the computational intelligence predictor. Our experiments involve classifying movie reviews extracted from the databases of an online movie seller website. These reviews are classified on a 4-point numerical rating scale.

A. The Dataset

The dataset[2] used for experimentation consists of movie reviews and numerical ratings. This dataset is widely used by researchers conducting opinion mining and sentiment analysis experiments. Experiments were conducted using the largest dataset within that collection, which consists of 16409 dictionary terms and 1770 reviews. Figure 2 shows the distribution of terms across the collection of reviews.

The dataset was accompanied by sets of ratings of different classes (see Table I). Initially pre-processing was applied to the dataset, as specified by the *Data Preparation* module of the proposed Computational Intelligence Predictor system. This module reduces the corpus size to 9848 dictionary terms and 1770 reviews and produces the normalised term-by-review matrix A.

[2]Available at http://www.cs.cornell.edu/people/pabo/movie-review-data, Sentiment Scale datasets

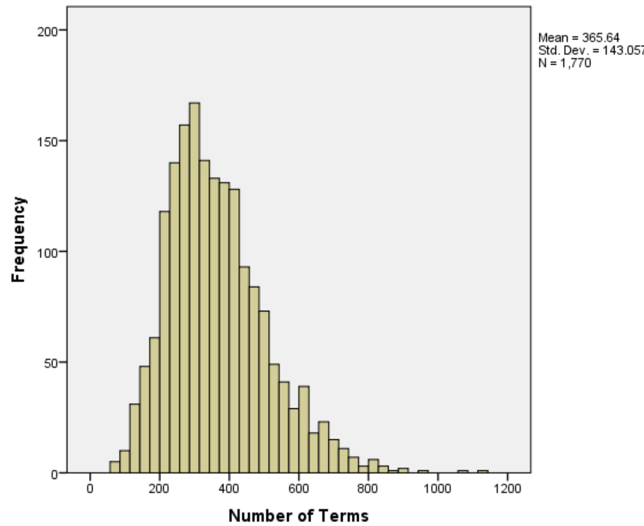

Fig. 2. Average Number of Terms per Review Histogram

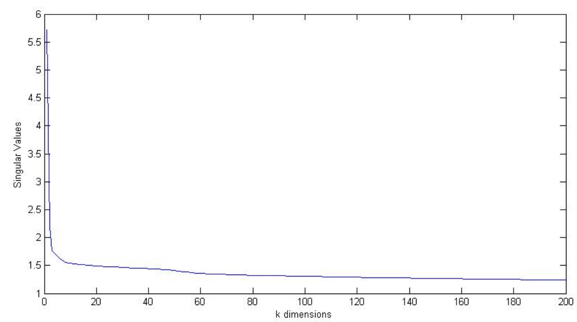

Fig. 3. Scree test

TABLE I
DATASET RATING CLASS CHARACTERISTICS

Rating	Average number of terms	Number of Reviews	Std. Deviation
0	367.55	191	130.531
1	336.36	526	133.704
2	359.77	766	141.192
3	433.74	287	151.000
Total	365.64	1770	143.057

This matrix A is then input to the *Input Selection* module. SVD was applied to the term-by-review matrix A. SVD was initially computed with the Dimensionality Reduction parameter k set to $k = 200$ dimensions in order to ease interpretation of Scree test output and to better identify an appropriate number of dimensions for matrix A. For quickly identifying the optimal number of components to retain, we have used Cattell's graphical Scree test [10], which is one of the most popular tests used for this purpose. The Scree test was performed by plotting matrix $\Sigma_{200\times 200}$, and the results are shown in Figure 3.

The results illustrated in Figure 3 indicate that selecting 20 features (i.e. setting the value of k to 20) is sufficient for capturing enough information to perform the prediction task. Consequently, Dimensionality Reduction was performed by using $k = 20$ dimensions and the derived review-by-dimension matrix $V_{1770\times 20}$ was used as input to the Neuro-Fuzzy Predictor module, described in Section III-C. The output of this module was a set of predicted values, which comprised one numerical rating for each review.

B. Evaluation Method and Results

The performance of the proposed Computational Intelligence Predictor was compared to two other models: a FCM based model (this will be referred to as SVD-FCM-FIS) and an Artificial Neural Network (ANN) based model (this will be referred to as SVD-ANN). The SVD-FCM-FIS model was constructed by replacing the Neuro-Fuzzy Predictor Module of the proposed Computational Intelligence System by FCM-FIS (completely removing the ANFIS functionality); and for the SVD-ANN based model the Neuro-Fuzzy Predictor Module was completely replaced by an ANN model. The remaining components (i.e. Data Preparation, and Input Selection modules) of the Computational Intelligence Predictor remained unchanged to allow for appropriate comparison of results.

The reason for constructing the aforementioned models is to allow the FCM and ANN techniques to be applied to the large data involved in this experiment. As mentioned in Section IV-A, the size of the pre-processed term-by-review matrix A is 9848×1707 and applying the ANN or FCM models on this dataset without first reducing the dimensionality of the data (i.e. via applying input selection techniques) makes these models unsuitable for the task of prediction due to the inherent computational complexity.

The same data indices were used for performing experiments with all three models. Evaluations were performed using the Root Mean Squared Error (RMSE) and Pearson product-moment correlations between target outputs t and actual outputs a. If the value of the correlation r, is 1.0, this indicates that there is an exact linear relationship between outputs and targets. The closer the value of r is to 1.0, the stronger the correlation between target and actual outputs. If r is close to 0, then there is no linear relationship between outputs and targets. The closer the value of r is to 0, the weaker the correlation between the target and actual outputs.

For the ANN model, the tan-sigmoid transfer function was used at the hidden layer, and the linear transfer function was used at the output layer. The ANN was trained by using the Scaled Conjugate Gradient (SCG) for Fast Supervised Learning suitable for large-scale problems[12]. The number of neurons was set to 4 so as to match the number of rules created by the proposed neuro-fuzzy model. The process of training the ANN involves tuning the values of the weights and biases of the network to optimise network performance measured by the mean squared error network function.

TABLE II
RESULTS OF ACTUAL AND PREDICTED OUTPUTS BETWEEN MODELS

	RMSE	Pearson Correlations (r)
Proposed CI Model	0.0603	0.7247
SVD-FCM-FIS	0.0681	0.6313
SVD-ANN	0.0634	0.6914

TABLE III
PEARSON CORRELATIONS (R) OF ACTUAL OUTPUTS

	Proposed CI model	SVD-FCM-FIS	SVD-ANN
Proposed CI model	1.0	0.786	
SVD-FCM-FIS	0.786	1.0	
SVD-ANN	0.837	0.929	1.0

The performance of the Computational Intelligence Predictor model was compared to that of the SVD-FCM-FIS and the SVD-ANN models. The results are shown in Table II and illustrated in Figs. 4-6. Highest correlations were achieved between the actual and target outputs of the proposed Computational Intelligence Predictor model ($r = 0.7247, p = 0.00, a < 0.01$). Lower correlations were returned by the SVD-FCM-FIS ($r = 0.6313, p = 0.00, a < 0.01$), and the SVD-ANN ($r = 0.6914, p = 0.00, a < 0.01$) models.

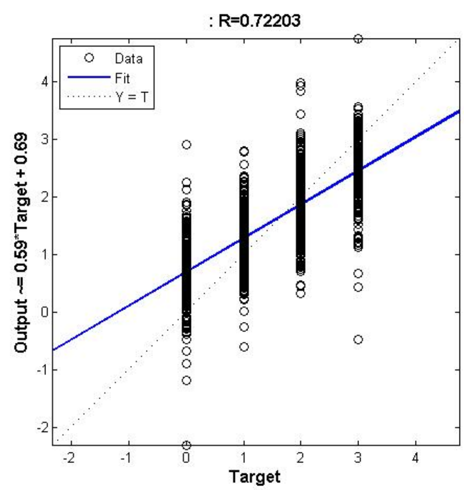

Fig. 4. Computational Intelligence model results: Correlation between actual and target outputs

Further experiments were conducted to determine the degree of similarity of the outputs of the three models. Correlations between the actual outputs of the models were computed and the result of this comparison is shown in Table III and illustrated in Figs. 7-9. Interestingly, when comparing the actual outputs (i.e. predicted outputs) of the models, the results revealed near perfect correlations between the SVD-FCM-FIS and SVD-ANN models ($r = 0.929, p = 0.000, a < 0.01$), strong correlations between the proposed Computational Intelligence Predictor model and the SVD-ANN model ($r = 0.837, p = 0.000, a < 0.01$), and high correlations between SVD-FCM-FIS model and the proposed

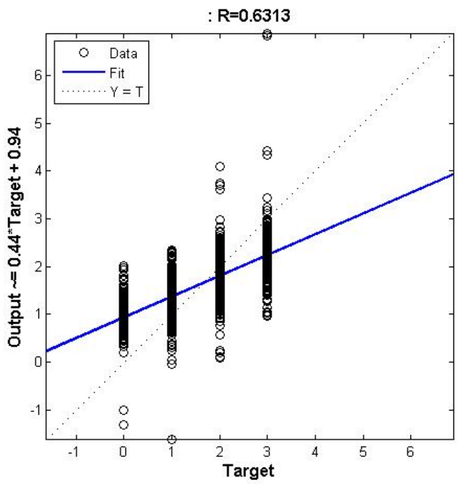

Fig. 5. SVD-FCM-FIS model results: Correlation between actual and target outputs

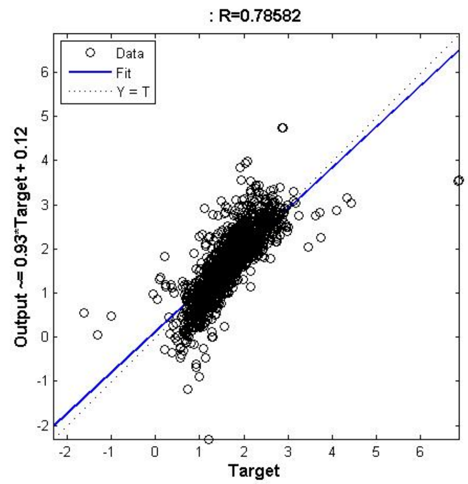

Fig. 7. SVD-FCM-FIS and Computational Intelligence Predictor models: Correlation between the actual outputs of the two models

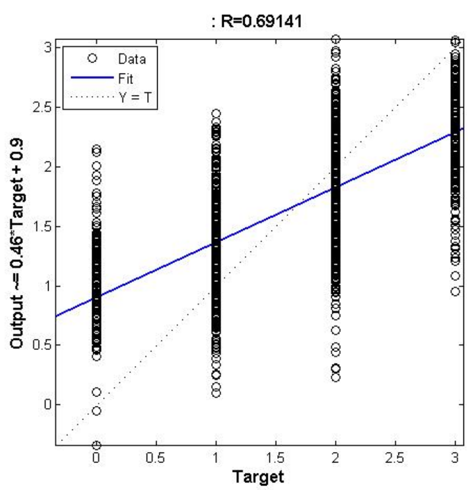

Fig. 6. SVD-ANN model results: Correlation between actual and target outputs

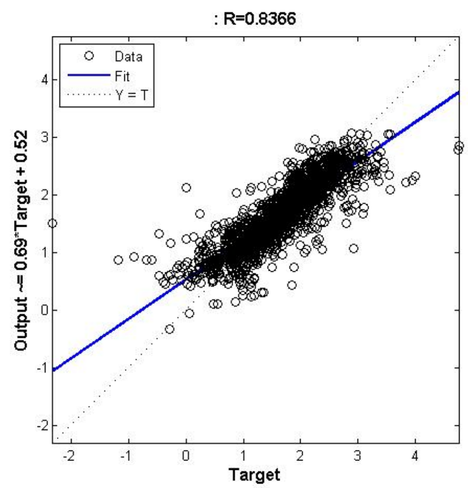

Fig. 8. Computational Intelligence Predictor and SVD-ANN models: Correlation between the actual outputs of the two models

Computational Intelligence Predictor model ($r = 0.786, p = 0.00, a < 0.01$). Clearly the performance of the SVD-FCM-FIS and SVD-ANN models is very similar. The proposed model outperformed both the SVD-FCM-FIS and SVD-ANN predictors.

V. Conclusion and Future work

Internet is changing the way how people approach the markets. Indeed the of e-commerce is enabling customers to accomplish their purchases in a more efficient way than when they use conventional commerce approaches. At the same time, these innovative market environments are enabling companies to increase their sales volume thanks to a world-wide visibility. In this new scenario, the evaluation of customers' reviews become of crucial importance to online merchants because they can use this information for planning their future business activities. In this paper we proposed an integrated computational intelligence approach that addresses some of the most important issues in this field: the dimensionality of data and accuracy. The reported case study shows how our idea outperforms some of the common methodologies used in this areas. In future works, we are aimed at improving the performance of the proposed system by replacing some of its components, as for instance the Input Selection Module, with evolutionary algorithms that could be able to reduce the input dimensionality better than the current approach.

References

[1] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10 760–10 773, sep 2009.

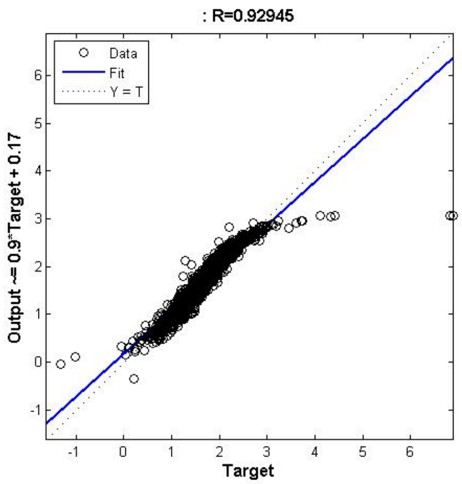

Fig. 9. SVD-FCM-FIS and SVD-ANN models: Correlation between the actual outputs of the two models

[2] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundation Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, Jan. 2008.

[3] R. Prabowo and M. Thelwall, “Sentiment analysis: A combined approach.” *Journal of Informetrics*, vol. 3, no. 2, pp. 143–157, 2009.

[4] F. Sebastiani, “Machine learning in automated text categorization,” *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.

[5] Q. Ye, Z. Zhang, and R. Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert Systems with Applications*, vol. 36, pp. 6527–6535, 2009.

[6] R. Moraes, J. F. Valiati, and W. P. G. Neto, “Document-level sentiment classification: An empirical comparison between svm and ann,” *Expert Systems with Applications*, vol. 40, pp. 621–633, 2013.

[7] G. Ganu, Y. Kakodkar, and A. Marian, “Improving the quality of predictions using textual information in online user reviews,” *Information Systems*, vol. 38, no. 1, pp. 1 – 15, 2013.

[8] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. ACM Press/Addison-Wesley, 1999.

[9] M. W. Berry, S. T. Dumais, and G. W. O’Brien, “Using linear algebra for intelligent information retrieval,” Tech. Rep. UT-CS-94-270, 1994. [Online]. Available: citeseer.ist.psu.edu/berry95using.html

[10] R. B. Cattell, “The scree test for the number of factors,” *Multivariate Behavioral Research*, vol. 1, pp. 245–276, 1966.

[11] M. Sugeno and T. Yasukawa, “A fuzzy-logic-based approach to qualitative modeling,” vol. 1, no. 1, pp. 7–31, 1993.

[12] M. F. Möller, “A scaled conjugate gradient algorithm for fast supervised learning,” *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.