International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 5, Issue 6, November - December 2016 **ISSN 2278-6856**

GARCH (1, 1) Outlier Detection Technique for Review Spam Detection

Siddu P. Algur[1], Jyoti G.Biradar[2], Prashant Bhat[3]

[1]Department of Computer Science, Rani Channamma University,
Belagavi-591156, Karnataka, India

[2] Department of Computer Science, Rani Channamma University,
Belagavi-591156, Karnataka, India

[3] Department of Computer Science, Rani Channamma University,
Belagavi-591156, Karnataka, India

Abstract

With the recent development of web 2.0, the volume of on-line sales has been increasing in a remarkable pace and has generated plentiful of user-created content. Among various types of user-generated data on the web, reviews about stores, products, businesses, or services written by users are becoming more and more important due to the word-of-mouth effect and their impact on influencing customers purchase decisions. These reviews are important source of information for the potential customers before deciding to purchase a product. As a result, websites containing customer reviews are becoming targets of opinion spam. Spam reviews corrupt the online review system and confuse the consumers. Hence, a novel and effective technique GARCH (1,1) model is used in this work to find review spamicity in the multidimensional time series reviews of the stores, extracted from review website resellerratings.com. The experimental results demonstrates that the method proposed is effective in detection of review spamicity.

Keywords:- Review spam, outliers, time series, multidimension, GARCH.

1. INTRODUCTION

Currently users of web 2.0 contribute content actively in product review websites, blogs and social media and web-forums. Opinions and reviews can now be found almost everywhere-blogs, social media like Facebook and Twitter, web-forums, e-commerce sites, etc. These opinions are helpful for both business organizations and individuals [4].Most users will spend a good amount of time reading user reviews if they are available. As more than 70% of online shoppers said they frequently rely on online product or book reviews, which were stated in a survey report in faves.com [3]. The consumers value the feedback given by other users as do the companies that sell such products. Blogs, websites, discussion boards etc, are repository of customer comments which are valuable and are rich sources of textual data. Therefore individuals rely extensively on the reviews available online. Due to this reason, the review system has become a target of spammers who are usually hired or enticed by companies
spammers who are usually hired or enticed by companies
to write fake reviews to promote their products and services, and/or to distract customers from their competitors. Driven by profits, there are more and more spam reviews in major review websites, such as PriceGrabber.com, Shopzilla.com, and Resellerratings.com etc. Spammers aims to corrupt the online review system and confuse the consumers [9].Detecting review spam is challenging task as no one knows exactly the amount of spam in existence. Spam reviews usually seems to be perfectly normal until one compares them with other reviews of same products so as to identify that the review comments are not consistent with the other reviews[12]Assessing the trustworthiness of reviews is a key issue for the opinion sites. As we classify reviews as spam and non-spam reviews with different techniques based on the behaviour of the reviewers. One can consider spam reviews as outliers from other reviews. Hence, in the proposed work review spamicity detection approach is based on outlier detection technique from the extracted reviews of the stores.

Outlier detection has become an important part of time series analysis. Outlier detection is the process of finding data objects with behaviours that are different from expectation. An outlier is an extreme observation that may have a severe effect on data analysis. Most real-world data sets contain outliers that have unusually large or small values when compared with others in the data set. Imprecision in data is one of the facts that cause the parameter estimations to be subjective. If the erroneous case is proved statistically, then these cases are called outliers [5]. Outliers are defined as the few observations or records which appear to be inconsistent with the rest of the group of the sample and more effective on prediction values. Outliers can be classified into three categories, namely global outliers, contextual (or conditional) outliers and collective outliers. The proposed work is categorized as contextual outlier which is also known as conditional outlier. As in GARCH (1,1) model, a condition is stated, that the values above the test statistics i.e above the threshold values are considered as outliers. Therefore, in this work, outliers are identified using a statistical approach GARCH (1,1) model for the four dimesions: average positive score, average negative score, average rating and average number of reviews identified from the reviews of stores Auto-parts_warehouse.com, Dhgate.com and neweggs.com extracted from review website resellerratings.com. Further, among the four dimensions, common dates found for three or four dimensions above the threshold value (outliers) days/dates are marked and these days/dates reviews are suspected as spam reviews. Review spamicity is measured considering total number of spam reviews detected using outlier detection model GARCH (1,1) by total number of reviews of the stores for the duration of 623 days. The results show that the method proposed in this work is effective in detecting outliers, and to find review spamicity measure for reviews of the stores namely, Auto_parts_warehouse.com, Dhgate.com and Neweggs.com. The present paper depicts about the trends of detection of review spamicity with respect to multidimensional time series. Section 2, introduces about the related work. Section 3, gives an overview of the proposed technique used to find review spamicity. Section 4, describes the working and experimental results for detecting review spam. And section 5 presents conclusion and future work. .

2 RELATED WORK

A significant amount of work has been performed in the area of time series outliers. In[1],the first work on outlier detection for time series data model for time series outliers was proposed, in this work, several models were proposed like autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), vector autoregression (VARMA), CUmulative SUM Statistics (CUSUM).In[13],the two main types of outlier detection techniques for time series in the data mining community is proposed. The first part concerns techniques to detect outliers over a database of time series, whereas the second part deals with outliers within a single time series. In[14], an extensive overview of outlier detection techniques are given. In this work, outlier detection has been studied in a variety of data domains including high-dimensional data. In[7],an ST-Outlier detection algorithm called Outstretch is proposed, which discovers the outlier movement patterns of the top-K spatial outliers over several time periods. The top-K spatial outliers are found using the Exact-Grid Top-K and Approx-Grid Top-K algorithms. In[8],Outlier detection is used in industrial applications, with many software tools, such as R-packages, SAS, Rapid Miner, and Oracle dataminer. In[6],outliers for data streams using prediction models are used, in this approach, one can also compute distance based outliers for data streams at any time instant.In[15], outliers using the subspace method is used and multivariate

time series measuring the number of bytes, packets and IP-level flows is also used to discover anomalies such as high rate point to-point byte transfer, denial of service, distributed denial of service attacks, etc. In [2], distance based outlier detection for distributed temporal data is used. In this work, Shifted Wavelet Trees method is used to find outlier from the NYSE IBM stock time series data Our method aims at extracting the reviews from review website from the three stores and to find review spamicity using outlier detection model GARCH (1,1) based on multidimensional time series.

3 METHODOLOGY

In the proposed work, an outlier detection technique namely, GARCH (1,1) model is used to detect review spamicity based on constructing multidimensional time series for the extracted reviews of the three stores namely Auto_parts_warehouse.com, Dhgate.com and Neweggs.com. From these reviews of stores, four dimensions [11] are identified and used. The identified multidimensions are positive word length score, negative word length score, review rating and no of reviews. The dimensional values are normalized in the range of 0-1. Each dimensional values are assigned to the GARCH (1,1) model month wise,i.e for each month all the four dimensional values are given to detect outliers for the duration of 623 days from 1^{st} January 2014 to 15^{th} September 2015.

The various steps of the proposed method include:

- Review Extraction.
- Identifying dimensions.
- Time series construction.
- Review Spamicity
  - Outlier detection using GARCH (1, 1)
  - Identifying review spam and to measure spamicity of reviews

- **Review Extraction**:
  Reviews are extracted from review website for the stores Auto_parts_warehouse.com, Dhgate.com and Neweggs.com using review exactor tool (import.io) and are stored in raw review database for all the three stores separately.

- **Identifying dimensions**:
  There are various dimensions which are used to support detection of spamicity of reviews like review similarity spam score, rating similarity score, rating deviation score, positive word length score, negative word length score, review word length score, rating , average rating, total number of reviews, ratio of singleton reviews etc. Among these dimensions, four dimensions are identified and are used in the proposed work, they are positive word length score, negative word length score, review rating and total number of reviews. The specific of these four dimensions has been discussed in [11].

- **Time series construction:**
  A time series is a sequence of numerical data points in successive order, usually occurring in uniform intervals. It's a sequence of numbers collected at regular intervals over a period of time.In the proposed work, review spamicity detection approach is based on multidimensional time series construction. The description of time series has been given in the work [11].

- **Review spamicity:** Reviewspamicity is the degree or measure of spam reviews identified from the given dataset of reviews. In the proposed work, detection of review spam and measure of review spamicity is given in two steps as follows:

  - Outlier detection using GARCH(1,1) model
  - Identifying review spam days and to measure spamicity of the reviews.
  - Outlier detection using GARCH(1,1) model

GARCH is an abbreviated form of Generalized Autoregressive Conditional Heteroskedasticity. This model is generalized form of the ARCH process i.e Autogressive Conditional heteroskedasticity; here the 'autoregressive' property in principle means that old events leave behind a certain time after the actual time of the action. The process depends on its past. The terms 'Conditional Heteroskedasticity' means that the variance (condition on the available information) varies and depends on old values of the process. The GARCH models are mean reverting and conditionally heteroskedastic but have a constant unconditional variance [5]. As unconditional variance have a constant value, it is set as a threshold in GARCH (1, 1) model.
The standard equation used in GARCH (1,1) model , is given below:

$$h_{t+1} = \omega + \alpha CR_i^2 + \beta h_t \text{ ------}1$$

Where h is variance, $CR_i^2$is theresidual squared,t denotes time, ω, α and β are empirical parameters determined by constant values assigned. Residual is the difference between the actual and predicted values. In the proposed work, residual is the difference between next day's (tomorrow's) variance 'h_{t+1}' and the present day's (today's) variance 'h_t' of the four dimensional values used.

The equation tells us that next days's (tomorrow's) variance (h_{t+1})is a function of

- today's squared residual (αCR_i^2)
- today's variance (βh_t)
- the weighted average long-term variance

For the GARCH (1, 1) model, the parameter values α, β and ω values are initialized as α = 0.1 ,β = 0.8, and ω = 0.1*V (where V is unconditional variance), by the variance of the series adjusted by the persistence [5]. In the model used, the argument list of the fitting function GARCH (1,1) requires only the time series dimensional

values, the rest will be computed step by step based on the standard equation of GARCH(1,1) given in equation –1 as:

Step 1 : Series initialization-In this step ,the four dimensional values (average positive score, average negative score, average review rating and average number of reviews) are given in the standard equation of GARCH(1,1) model .

Step 2 : Parameter initialization- The set of parameters α, β and ω are initialized as α =0.1, β=0.8 and ω = 0.1 * V, where V is unconditional variance.

Step 3: To compute variance and unconditional variance ,for four dimensional values, based on the equation -----1, residual 'CR_i', squared residual 'CR_i^2', variance 'h' and unconditional variance 'V' are computed.

Step 4 : To detect outliers – Outliers are detected based on the output of variance 'h' and unconditional variance 'V'. The unconditional variance 'V' is considered as threshold value, as its value remains constant. The dimensional values found above the threshold value i.e unconditional variance are considered as outliers. In the proposed work, the outliers detected dimensional values review dates/day are suspected as spam reviews.

In the proposed work, the procedure to detect review spam days is given in the ALGORITHM1 'Review_Spamicity model ( )' and the procedure to detect outliers using GARCH (1,1) model is given in the ALGORITHM 2 'GARCH (α, β, D_i)'.

ALGORITHM1Review_Spamicity model ( )
// To detect review spam days
// Input parameters: Dimensions D [ ]
// Output: Spam_days
$\alpha \leftarrow 0.1$, $\beta \leftarrow 0.8$ //Initialize α, β values
For each dimension D_i do
Outliers[i]←GARCH (α, β, D_i) // calling GARCH(α, β, D_i) function
End for
Spam_days←(Outliers[1]∩Outliers[2]∩,Outliers[i])
Print (Spam_days)

Algorithm 1 : Computes Review spam days.
The Review_Spamicity model ( ), describes the dates/days where outliers are identified from each dimension. Among four dimensions, common dates found from three or four dimensions wherein identified outliers are considered and those days reviews are suspected as spam day reviews.

ALGORITHM2 GARCH (α, β, D_i)
// To detect outliers using GARCH model
// Input : R[ ] = { $R_1, R_2, R_3 \ldots\ldots\ldots\ldots\ldots R_i$ }set of reviews
// Output: Outliers
For each review R_iDo
$CR_i \leftarrow h_{t+1} - h_t$// Compute residual
CR_isqr ← CR_i^2// Compute square of Residual
V←VAR(CR_1, CR_2, CR_3,......CR_i)
//Compute unconditional variance of (CR_1, CR_2, CR_3,......CR_i)
ω ←0.1*V // Compute ω value
Threshold ← √V // Compute square root of V
End for
For each dimension D_i do
if (ω($D_{i[Day]}$)) > Threshold // To detect outliers
Outlier[i]←$D_{i[Day]}$
End for
Return Outlier [1,2,3,.....i]

Algorithm 2 : Computes the Outliers using GARCH(1,1)
The‘GARCH (α, β, D_i)’ algorithm, describes detection of outliers using GARCH (1,1) model. In this model, the values of unconditional variance remains constant. Hence, this value is considered as threshold value. From each dimension, reviews found above the threshold values are suspected as outliers.

Identifying review spam days and to measure spamicity of the reviews :
Review spam days are identified using the outlier detection technique. Among the four dimensions considered for the proposed work, common dates are found for three or four dimensions above the threshold value in the GARCH (1,1) model.The reviews found for those dates/days which are above the specified threshold value will be suspected as spam days reviews.

Spamicity of the reviews are measured by considering the number of spam days reviews found, by the total number of reviews of the stores for the entire duration.

In the proposed work, review spamicity is computed for the three stores namely Auto_parts_warehouse.com, Dhgate.com and Neweggs.com. The spamicity of these stores are computed by considering four dimensions which are mentioned earlier.

4. EXPERIMENTAL RESULTS

Experimental results are presented to demonstrate the effectiveness of the proposed method. Experiments are carried from extracting reviews from review website resellerrating.com for the three stores Auto_parts_warehouse.com, Dhgate.com and Neweggs.com. The review website contains 49, 49,284 reviews for 1,96,640 stores as on 15th September 2015. There are 27,522 reviews from Auto_parts_warehouse.com, 12,513 reviews from Dhgate.com and 3,281 reviews from Neweggs.com. A total of 43,316 reviews are taken from all the three stores. The data consists of reviews, along with information about stores and reviewers. For each review following information is considered: reviewer’s name, its rating (ranging from 1 to 5), the posting date and content of the review. Detection of review spamicity is constructed on multidimensional time series analysis from the extracted reviews of the three stores based on the average of positive word length score, negative word length score, review ratings and number of reviews. For these four dimensions, outliers are identified month wise from 1st January 2014 to 15th September 2015 using an outlier detection model GARCH (1,1). The size of the time window is set to ‘one day’. For all the four dimensions, outliers detected dates/days are marked month wise. And the dates/days found in the outliers marked are suspected as spam reviews. Spamicity of the reviews of the three stores are measured, considering the total number of spam reviews found using outlier detection technique GARCH (1,1) by the total number of reviews of the stores for duration of 623 days.

In the Figure 1, two graphs are plotted, graph on the upper part depicts total number of reviews found and graph on the lower part depicts spam reviews identified from the four dimensional values for the store Auto_Parts_warehouse.com of the month January 2014.

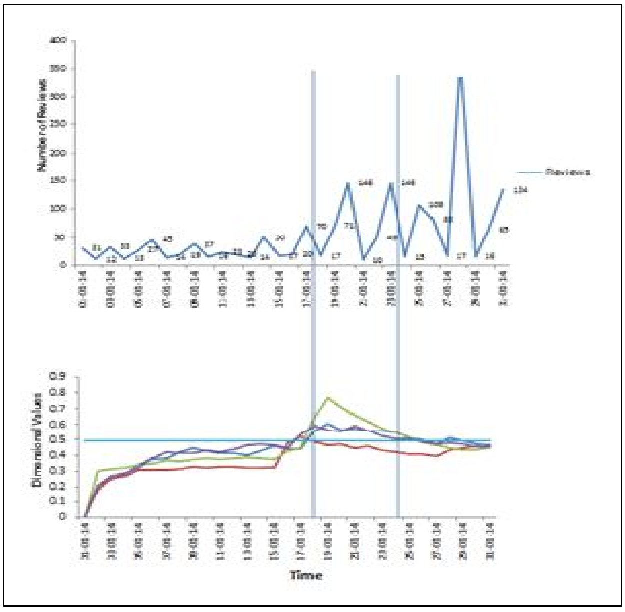

Figure1. Total Number of reviews and spam reviews identified from four dimensions for the store Auto_parts_warehouse.com of January 2014.

From the Figure 1, the common dates found from three/four dimensions above the threshold value i.e 0.49 are from 18-01-14 to 24-01-14(vertical lines are used in the Figure1 to highlight the common dates found). The numbers of reviews for these days are 454, these reviews are suspected as spam reviews. Review spamicity measure is calculated considering total number of spam reviews identified i.e, ‘454’ reviews by total number of reviews of the store for the month January 2014 i.e ‘1701’ reviews. The spamicity measure for the month January 2014 of the store Auto_parts_warehouse is 27%. For the month February 2014, the common dates found above the threshold values are from 02-02-14 to 8-02-14 and 15-02-14, the reviews found above the threshold values are ‘733’. The total number of reviews for this month are 2536 reviews. The spamicity measure for the month, February 2014 is 29%. Similarly, spamicity of the reviews is calculated for the remaining months for duration of 623 days is shown in the Table 1.

The total number of spam reviews found are ‘5432’ reviews, is illustrated in the Figure 3.Total number of reviews of this store for entire duration are ‘27522’ (is depicted in the Figure 2). Hence, review spamicity measure of the store Auto_parts_warehouse.com is 19.73%.

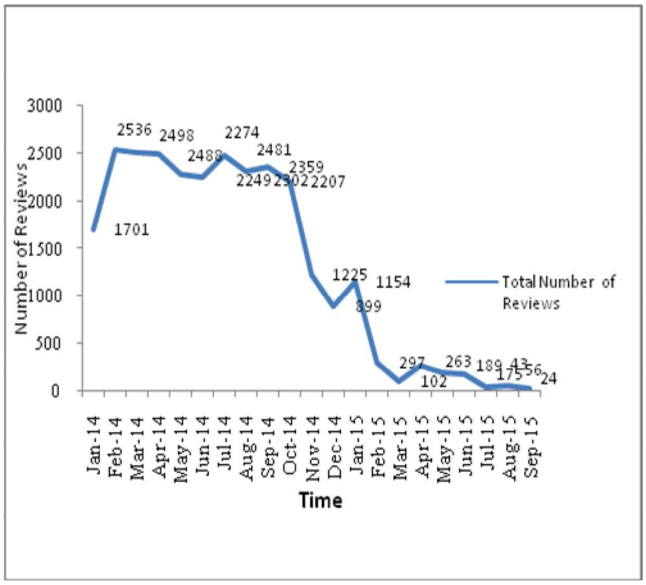

Figure2. Total number of reviews for the store Auto_parts_warehouse.com from 1st January 2014 to 15th September 2015.

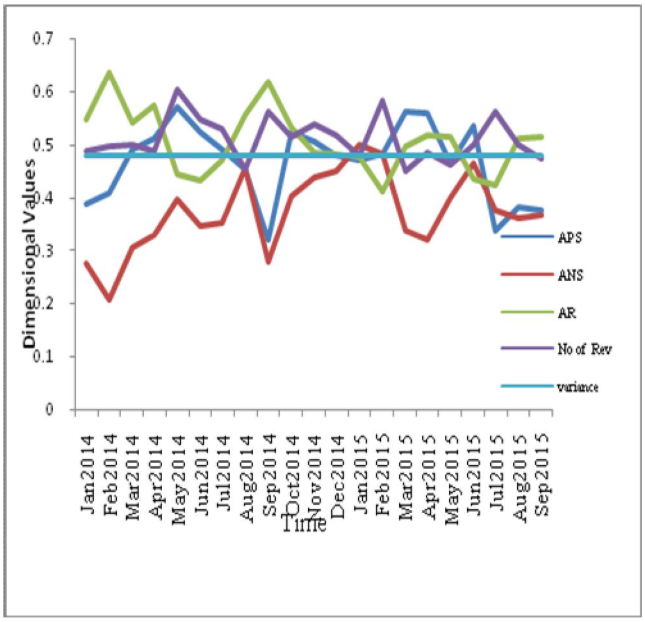

Figure 3.Spam reviews detected from four dimensions for the store Autopartswarehouse.com from 1st January 2014 to 15th September 2015.

In the Figure 4, two graphs are plotted, graph on the upper part depicts total number of reviews found and graph on the lower part depicts spam reviews identified from the four dimensional values for the store Dhgate.com of the month January 2014.

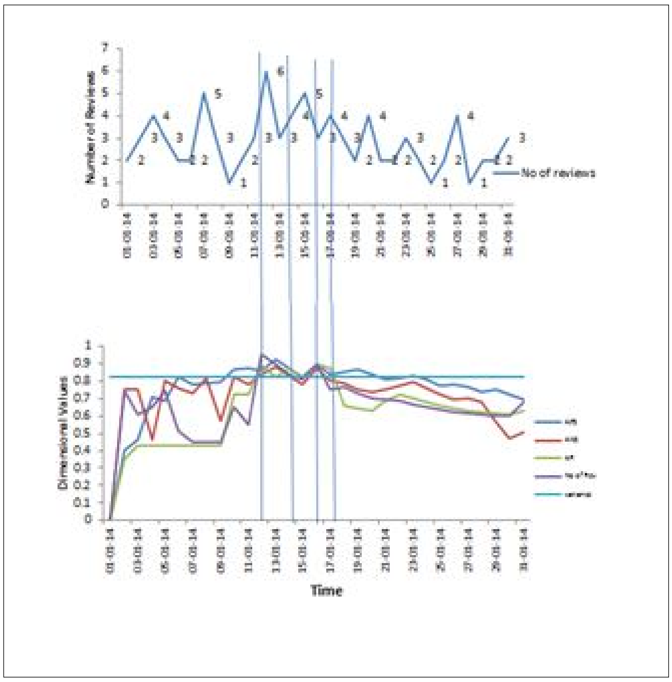

Figure 4. Total number of reviews and spam reviews identified from four dimensions for the store Dhgate.com of January 2014.

From the Figure 4, the common dates found from three/four dimensions above the threshold value i.e 0.82 are from 12-01-14 to 14-01-14 and 16-01-14 to 17-01-14. (vertical lines are used in the Figure 4 to highlight the common dates found) The number of reviews for these days are 20, these reviews are suspected as spam reviews. Review spamicity measure has been calculated by considering the total number of spam reviews identified i.e '20' reviews by total number of reviews of the store for the month January 2014 i.e '88' reviews. The spamicity measure for the month January 2014 of the store Dhgate.com is 23%. For the month, February 2014, the common dates found above the threshold values are from12-02-14 to 15-02-14, the reviews found above the threshold values are '16'. The total number of reviews of this month is 62 reviews. The spamicity measure for February 2014 is 26%. Similarly, spamicity of the reviews is calculated for the remaining months for the entire duration is shown in the Table 1. And the total number of spam reviews identified are '2345' reviews, (is illustrated in the Figure 6).Total number of reviews of this store for the entire duration are '12513' reviews (is depicted in the Figure 5).Hence, review spamicity measure for duration of 623 days for the store Dhgate.com is 18.74%.

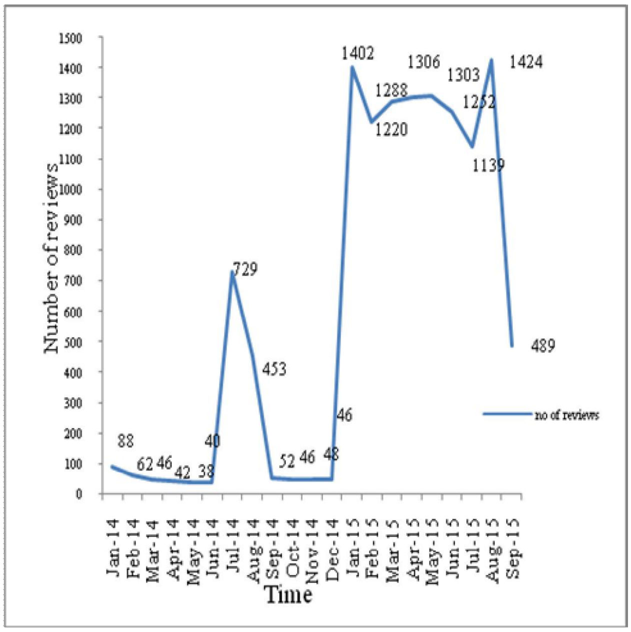

Figure5. Total number of reviews for the store Dhgate.com from 1st January 2014 to 15th September 2015.

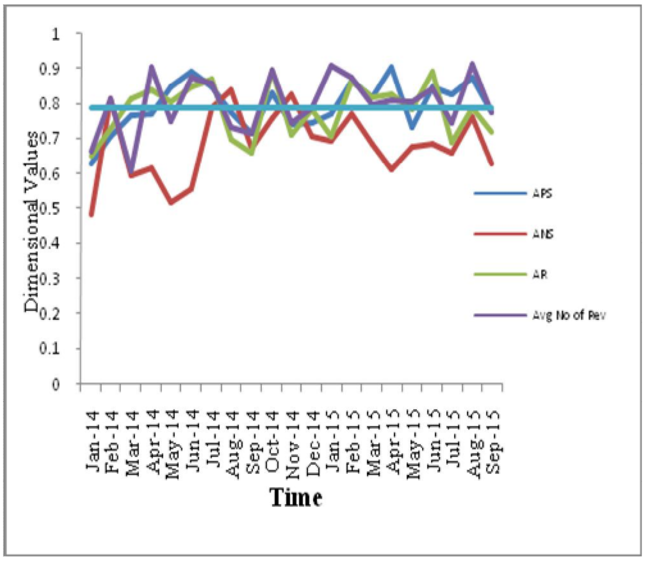

Figure 6. Spam reviews identified from four dimensions for the store Dhgate.com from 1st January 2014 to 15th September 2015.

In the Figure 7, two graphs are plotted, graph on the upper part depicts total number of reviews found and graph on the lower part depicts spam reviews identified from the four dimensional values for the store Neweggs.com of the month January 2014.

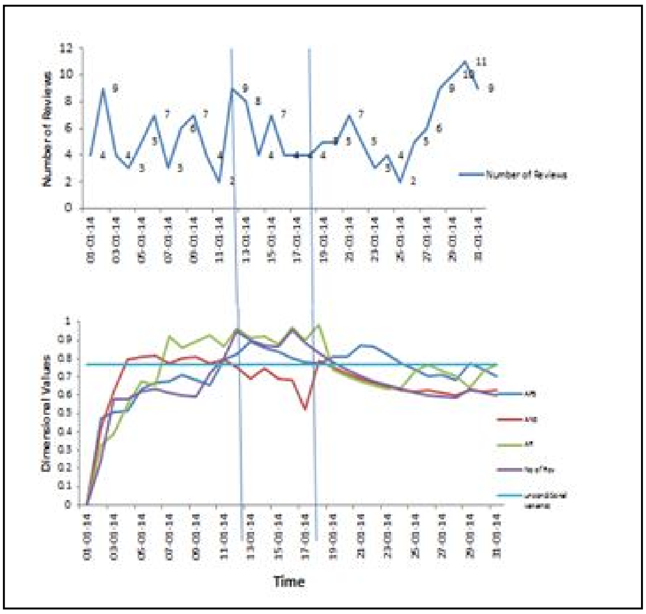

Figure 7.Total number of reviews and spam reviews detected from four dimensions for the store Neweggs.com of January 2014.

From the Figure 7, the common dates found from three/four dimensions above the threshold value i.e 0.77 are from 12-01-14 to 18-01-14.(Vertical lines are used in the Figure 7 to highlight the common dates found) The number of reviews for these days are 40, these reviews are suspected as spam reviews. Review spamicity measure is calculated considering total number of spam reviews identified i.e '40' reviews by total number of reviews of the store for the month January 2014 i.e '175' reviews. The spamicity measure for the month, January 2014 of the store Neweggs.com is 23%. For the month February 2014, the common dates found above the threshold values are from09-02-14 to 14-02-14, the reviews found above the threshold values are '48'. The total number of reviews for this month are '198' reviews. The spamicity measure for February 2014 is 24%. Similarly, spamicity of the reviews is calculated for the remaining months for duration of 623 days is shown in the Table 1. And the total number of spam reviews found are '550' reviews (is illustrated in the Figure 9). Total number of reviews of this store for the entire duration are '3281' reviews (is depicted in the Figure 8).Hence, review spamicity measure for the store Neweggs.com for duration of 623 days is 16.76%.

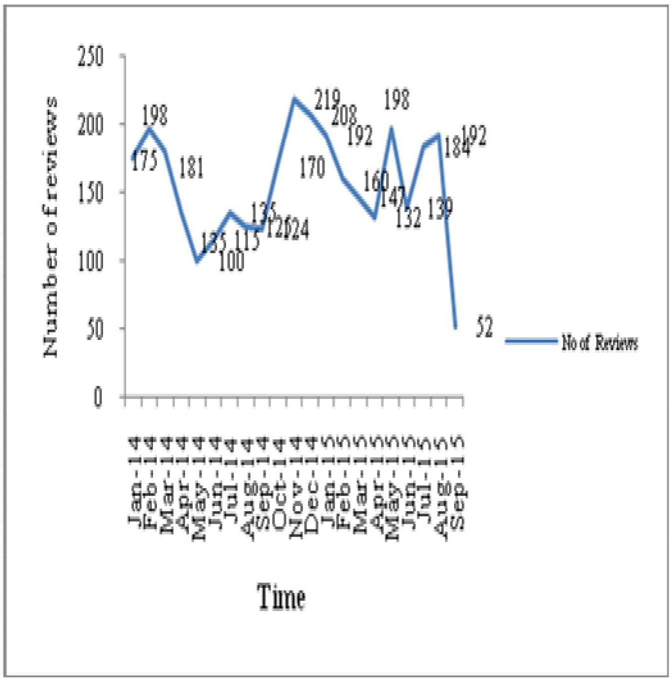

Figure8. Total number of reviews for the store neweggs.com from 1st January 2014 to 15th September 2015.

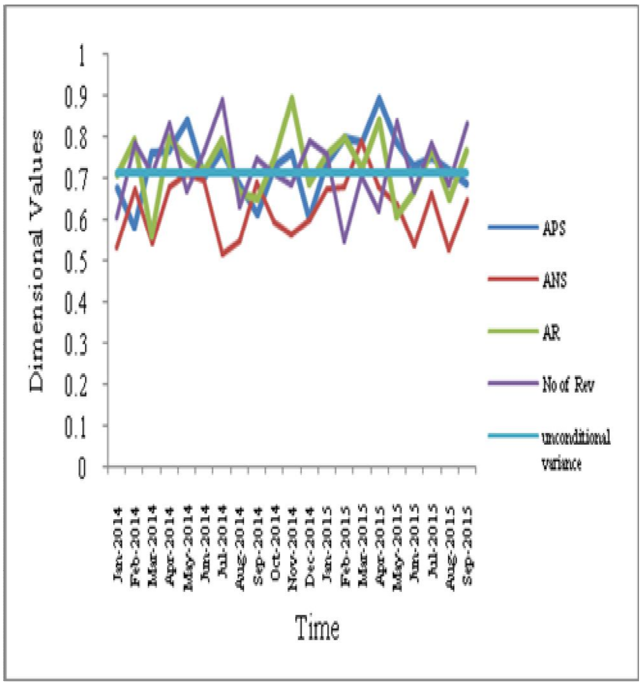

Figure9. Spam reviews detected from four dimensions for the storeNeweggs.com from 1st January 2014 to 15th September 2015.

Table 1. Comparative table of Total number of Reviews, Number of reviews detected as spam and Percentage of reviews detected as spam for the three stores.

Names of Stores	Auto_parts_warehouse. com			Dhgate.com			Neweggs.com		
Date	Total number of Reviews	Number of reviews detected as Spam	% of reviews detected as spam	Total number of Reviews	Number of reviews detected as spam	% of reviews detected as spam	Total number of Reviews	Number of reviews detected as Spam	% of reviews detected as spam
Jan-14	1701	454	27	88	20	23	175	40	23
Feb-14	2536	733	29	62	16	26	198	48	24
Mar-14	2498	380	15	46	11	24	181	22	12
Apr-14	2488	598	24	42	6	14	135	24	18
May-14	2274	917	40	38	4	11	100	26	26
Jun-14	2249	288	13	40	7	18	115	20	17
Jul-14	2481	359	14	729	93	13	135	28	21
Aug-14	2302	520	23	453	54	12	125	23	18
Sep-14	2359	253	11	52	8	15	124	27	22
Oct-14	2207	284	13	46	16	35	170	30	18
Nov-14	1225	120	10	48	8	17	219	38	17
Dec-14	899	63	7	46	9	20	208	32	15
Jan-15	1154	92	8	1402	263	19	192	39	20
Feb-15	297	125	42	1220	220	18	160	14	9
Mar-15	102	31	30	1288	214	17	147	18	12
Apr-15	263	56	21	1303	219	17	132	26	20
May-15	189	91	48	1306	289	22	198	21	11
Jun-15	175	35	20	1252	325	26	139	20	14
Jul-15	43	17	40	1139	236	21	184	27	15
Aug-15	56	10	18	1424	266	19	192	15	8
Sep-15	24	6	25	489	61	12	52	12	23
Total number of Reviews/ Number of reviews detected as spam / % of reviews detected as spam	**27522**	**5432**	**19.73%**	**12513**	**2345**	**18.74%**	**3281**	**550**	**16.76%**

As a result, the spam detection rates found are 19.73%, 18.74% and 16.76% for the stores Auto_parts_warehouse.com, Dhgate.com and Neweggs.com respectively. From the experimental results one can observe that, if the number of review are more, spamicity of the reviews will be less. And if the numbers of reviews are less, spamicity of the reviews will be more. There are evenlarge numbers of non-spam reviews also. Hence, these reviews do not influence the buying decision of the customers and could be regarded as trustworthy as they provide genuine opinion on some or the other sentiment of the store and are often unbiased [10].

5. CONCLUSION AND FUTURE WORK

In this work, a novel evaluation method, Outlier detection technique GARCH (1, 1) is used to find review spamicity with multidimensional time series by using multiple stores. Four dimensions are identified namely, positive word length score, negative word length, review rating and number of reviews. Based on these dimensions, multidimensional time series is constructed. The length of the time window is chosen to be of one day.

The procedure to detect review spam days is given in the ALGORITHM1 Review_Spamicity model ( ) and the procedure to detect outliers using GARCH (1,1) model is given in the ALGORITHM 2 GARCH (α, β, D_i). Experimental results of detecting review spamicity by using review website resellerratings.com for the stores Auto_parts_warehouse.com, Dhgate.com and Neweggs.com for the duration of 623 days from 1st January 2014 to 15th September 2015 demonstrates that the proposed method is effective in detecting review spamicity. Review spamicity measure using multiple criteria of the reviews gives the scope for future work.

REFERENCES:

[1] A. J. Fox, "Outliers in Time Series," Journal of the Royal Statistical Society. Series B (Methodological), vol. 34, no. 3, pp. 350–363., 1972.

[2] A. Lakhina, M. Crovella, and C. Diot, "Characterization of Network-wide Anomalies in Traffic Flows," in Proc. of the 4th ACM SIGCOMM Conf. on Internet Measurement (IMC), pp. 201–206.,2004.

[3] Cios, K.J., Pedrycz, W., and Swiniarski, R.W." Data mining Methods for Knowledge Discovery. KluwerAcademicPublishers, Norwell, MA, USA.(ISBN0-7923-8252-8)., 1998.

[4] Daya L. Mevada,VirajDaxini "An Opinion Spam Analyzer For Product Reviews Using Supervised Machine Learning Method " Journal Of Information, Knowledge And Research In Computer Engineering,. ISSN: 0975 – 6760 Volume – 03, Issue 02. , 2015.

[5] Diethelm W¨urtz1., Yohan Chalabi., and Ladislav Luksan "Parameter Estimation of ARMA Models with GARCH/APARCH Errors An R and SPlus Software Implementation "Journal of Statistical Software, Volume VV, Issue II., 2015.

[6] E. M. Knorr and R. T. Ng, "Algorithms for Mining Distance Based Outliers in Large Datasets," in Proc. of the 24th Intl. Conf. on Very Large Data Bases (VLDB). Morgan Kaufmann, pp. 392–403., 1998.

[7]E. Wu, W. Liu, and S. Chawla, "Spatio-Temporal Outlier Detection in Precipitation Data," in Proc. of the 2nd Intl. Conf. on Knowledge Discovery from Sensor Data (Sensor-KDD), pp. 115–133.,2010.

[8] H. Cho, Y. jin Kim, H. J. Jung, S.-W. Lee, and J. W. Lee, "Outlierd: An r package for outlier detection using quantile regression on mass spectrometry data," Bioinformatics, vol. 24, no. 6, pp. 882–884., 2008.

[9] Jindal Nitin., and Bing Liu..,. "Opinion spamand analysis". In Proceedings ofthe International Conference on Web search and web data mining, WSDM, New York, ACM ,pp.219–230., 2008

[10] Siddu P. Algur., Amit P.Patil., P.S Hiremath., and S. Shivashankar "Conceptual level Similarity Measure based Review Spam Detection" In IEEE International Conference on Signal and Image Processing . ISBN 978-1-4244-8594-9/10., 2010.

[11] Siddu P. Algur, Jyoti G. Biradar and Prashant Bhat "Multidimensional Time Series Based Review Spam Detection" International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE) " Vol.4, Issue 6,pp.11761-11774., June 2016.

[12] Sushant Kokate1, Bharat Tidke"Fake Review and Brand Spam Detection using J48 "(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (4) ,pp. 3523-3526, 2015.

[13] V. Barnett and T. Lewis, Outliers in Statistical Data. John Wiley & Sons Ltd., 1978.

[14] Y. Zhang, N. Meratnia, and P. J. M. Havinga, "Outlier Detection Techniques For Wireless Sensor Networks: A Survey," Centre for Telematics and Information Technology University of Twente, Tech. Rep. TR-CTIT-08-59, Oct 2008.

[15]Y. Zhu and D. Shasha, "Efficient Elastic Burst Detection in Data Streams," in Proc. of the 9th ACM Intl. Conf on Knowledge Discovery and Data Mining (KDD), pp. 336–345.,20

AUTHORS

Dr. Siddu P. Algur is working as Professor, Dept. of Computer Science, Rani Channamma University, Belagavi, Karnataka, India. He received B.E. degree in Electrical and Electronics from Mysore University, Karnataka, India, in 1986. He received his M.E. degree in from NIT, Allahabad, India, in 1991.He obtained Ph.D. degree from the Department of P.G. Studies and Research in Computer Science at Gulbarga University, Gulbarga. He worked as Lecturer at KLE Society's College of Engineering and Technology and worked as Assistant Professor in the Department of Computer Science and Engineering at SDM College of Engineering and Technology, Dharwad. He was Professor, Dept. of Information Science and Engineering, BVBCET, Hubbli, before holding the present position. He was also Director, School of Mathematics and Computing Sciences, RCU, Belagavi. He was also Director, PG programmes, RCU, Belagavi. His research interest includes Data Mining, Web Mining, Big Data and Information Retrieval from the web and Knowledge discovery techniques. He published more than 49 research papers in peer reviewed International Journals and chaired the sessions in many International conferences.

Mrs. Jyoti. G.Biradar is pursuing Ph.D programme in Computer Science at Rani Channamma University Belagavi, Karnataka, India. She received MCA & M.Phil degrees from Indira Gandhi National Open University and Vinayak Missions University, India in 2005 and 2009 respectively. Her research interest are Data Mining, Text Mining, and Information Retrieval from the web and Knowledge discovery techniques, and published 06 research papers in International Journals. She has attended and participated in her research field in International and National Conferences and Workshops

Mr. Prashant Bhat is pursuing Ph.D programme in Computer Science at Rani Channamma University Belagavi, Karnataka, India. He received B.Sc and M.Sc (Computer Science) degrees from Karnataka University, Dharwad, Karnataka, India, in 2010 and 2012 respectively. His research interest includes Data Mining, Web Mining,web multimedia mining and Information Retrieval from the web and Knowledge discovery techniques, and published 21 research papers in International Journals. Also he has attended and participated in International and National Conferences and Workshops in his research field.