FEATURE SELECTION USING MODIFIED ANT COLONY OPTIMIZATION APPROACH (FS-MACO) BASED FIVE LAYERED ARTIFICIAL NEURAL NETWORK FOR CROSS DOMAIN OPINION MINING

[1]DR.E.CHANDRA BLESSIE, [2] S.GNANAPRIYA
[1]associate Professor, Department Of Computer Science, Nehru College Of Management, Thirumalayampalayam, Coimbatore 641105, India
[2]ph.D Research Scholar, Department Of Computer Science,Nehru College Of Management,Thirumalayampalayam,Coimbatore 641105,India

E.mail: [1]chandra_blessie @yahoo.co.in, [2] gnanapriya_2006@yahoo.co.in

ABSTRACT

Web mining and web usage mining are attracting many researchers to propose new ideas, models, deploying machine learning algorithms and more. Internet usage expands its wings to almost all kind of applications which includes e-commerce. E-commerce facilitates the consumers/customers to buy the products online and at the same time, web analytics helps the website administrators to identify which products sell more. Opinion mining is the key to analytics in many decision-making tasks in the e-commerce arena. This research work aims to propose feature election using modified ant colony optimization approach (FS-MACO) based five layered artificial neural networks for cross-domain opinion mining. Dataset is obtained which consists of reviews about products such as books, DVDs, electronics and kitchen appliances. The features are identified by making use of modified ACO and opinion mining is performed by using ANN. Accuracy and F-measure are the metrics chosen for the evaluating the performance of the proposed work. Comparison of domain-specific and domain – dependent words are presented. Results portray that the proposed work outperforms better than that of the existing work in terms of the chosen performance metrics.

Keywords: *Opinion Mining, E-Commerce, Optimization, Neural Network, Accuracy, F-Measure.*

1. INTRODUCTION

The chance to get the opinion of the overall population about get-togethers, political developments, association methodologies, exhibiting efforts, and item tilt has raised expanding enthusiasm of both mainstream researchers (as a result of the energizing open difficulties) and the business world (because of the surprising advantages for advancing and money related market expectation). Today, sentiment analysis research has its applications in a few unique circumstances. There are a decent number of associations, both substantial and little scale, that emphasis on the investigation of opinions and sentiments as a feature of their fundamental objective. Opinion mining strategies can be utilized for the creation and mechanized upkeep of survey and opinion collection sites, in which opinions are consistently assembled from the Web and not confined to just item reviews, but instead add to more extensive subjects, for example, political issues and brand observation. Sentiment analysis having extraordinary potential as a subcomponent advancement for different frameworks. It can upgrade the abilities of client relationship organization and proposition frameworks; for instance, enabling clients to discover which features clients are especially inspired by or to reject things that have gotten unmistakably negative feedback from suggestion records. Essentially, it can be utilized as a piece of social correspondence for troll separating and to upgrade hostile to spam frameworks. Business knowledge is moreover one of the fundamental

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

factors behind corporate enthusiasm for the field of sentiment examination.

Makers and purchasers, require opinion mining apparatuses to gather opinions about a specific item. The opinion investigation devices can be utilized by makers to choose an advancing system for assessing creation rate. Then once more, buyers can utilize these apparatuses to settle on choice of buying another item or travel to relax areas or select hotel, et cetera. Named opinions are utilized to break down the classifier. For all intents and purposes, named opinions for each domain isn't conceivable, as it delimited by time and cost, while domain adjustment or exchange learning could be utilized to circumvent this imperative. With the above-mentioned problem statement, in this first phase of research work, we expect to propose Feature Selection using Modified Ant Colony Optimization Approach (FS-MACO) based Five Layered Artificial Neural Network (FLANN) for Cross-Domain Opinion Mining.

2. LITERATURE REVIEW

In M. Kolahkaj and M. Khalilian.,2015 hybrid procedure used to be proposed to generate a list of fascinating strategies centered on users view. Khairullah Khan et al.,2014 proposed a systematic background study regarding the computational procedures, models, and algorithms for mining opinion accessories from unstructured reviews, given that Opinion mining was a solution to retrieve knowledge via serps, web blogs and social networks. M. Roccetti et al.,2015 uncovered the role of online social networks for a developing group of power patients: Crohn's disorder patients. Their contribution was twofold: (a) authors represent the data exchanged with the aid of Crohn's sufferers, (b) while analyzing how they handle given subject matters of medical interest. In precise, of great clinical relevance used to be the effect that Infliximab was once the therapy that customarily influences the Crohn's sufferer neighborhood sentiments. X. Gong et al.,2013 have presented a search behavior based latent semantic user segmentation approach and validate its effectiveness on new ads. Marios Belk et al.,2013 all for modeling users' cognitive styles centered on a suite of net usage mining strategies on user navigation patterns and clickstream knowledge. Pei-Ling Hsu et al.,2015 detected quite a lot of semantic relationships (hierarchical and non-hierarchical) between ideas utilizing search logs and social annotations, on the grounds that many experiences have proposed tactics to extract wisdom from these consumer-generated datasets. Mohamed M. Mostafa et al.,2013 examined the outcomes of quite a lot of demographic, cognitive and psychographic motives on Egyptian citizens' use of e-govt services. Data mining makes use of an extensive family of computationally intensive ways that comprise determination trees, neural mesh network, rule elicitation, machine learning and graphic visualization. N. M. A. Al-Yazeed., 2015 proposed a hybrid web page rating mannequin headquartered on internet usage mining process with the aid of exploiting session data of users, to enhance the ideas of the following candidate web page to be accessed.

P. Lakar et al.,2015 proposed a novel method of enabling web engines with device shrewd FW which will extract the contextual person's behavior patterns in actual time through device delicate sensors, WoT Sensors and searching/net-apps usage logs. P. Suppa and E. Zimeo.,2016 proposed a recommender that presents, in addition to the traditional search tactics, some help for context- mindful implicit search of services, centered on social expertise. S. Aggarwal and V. Mangat.,2015 offered a holistic view as to what clickstream data analysis used to be, how mining strategies are applied on such information to generate useful knowledge and what kind of purposes take advantage of it to get useful know-how. T. Gopalakrishnan et al.,2014 recognized the countless semantic family members that exist between two given phrases, authors proposed a novel pattern extraction algorithm and a sample clustering algorithm. The most efficient combination of page counts-founded co-incidence measures and lexical pattern clusters used to be realized utilizing support vector machines.

Journal of Theoretical and Applied Information Technology
30th June 2018. Vol.96. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. FEATURE SELECTION USING MODIFIED ANT COLONY OPTIMIZATION APPROACH (FS-MACO)

FS-MACO is based on utilizing faith formation in optimization. FS-MACO considers constant faith progression on a populace of M ant as a way to deal with optimization problem in online shopping. To take care of the issue, we planned to characterize the faith vector of ant i as $Y_i = (Y_i^1, Y_i^2, \ldots, Y_i^C), Y_i^c \in [-1, +1]$ where C is the quantity of measurements of the faith vector, i.e., the issue estimate. For every operator i there exists a relating competitor arrangement $Q_i = (Q_i^1, Q_i^2, \ldots, Q_i^C), Q_i^c \in \{0,1\}$ to the issue that is figured as an element of ant's faith vector as:

$$Q_i = W(Y_i) = \frac{1 + sign(Y_i)}{2} \quad (1)$$

where the indication of each measurement of the faith vector decides the ant colony incentive for the comparing measurement in the arrangement, and its total esteem speaks to the assurance (certainty) under which the operator holds the answer for the issue.

The target work $f(Q_i): \{0,1\}^C \rightarrow [0,1]$ of the ant colony optimization issue decides the decency for faith vectors as answers for the issue. Here it's necessary think about the standard form of an unconstrained ant colony optimization issue, and thus the objective is to limit the cost work. As characterized in the past segment, M_i is the arrangement of neighboring ants to operator i that is dictated by the basic association organize structure. At each time of step w, each individual i in the populace chooses the best individual j in M_i (as per their figured cost esteems) as its conversationalist. At that point, ant i produces another faith vector, meant by the trial faith vector ̂, based on its present faith vector Y_i affected by its conversationalist operator j through a procedure called cooperation. The present faith vector of an ant may be substituted with the created trial faith vector of a similar importance, in specific situations. At every cycle w of the calculation, all ant produce their own particular trial faith vector as indicated by the communication calculation characterized as beneath:

$$\begin{cases} j = \arg_{j \in M_i}\left(f\left(Q_j(w)\right)\right) \\ \hat{Y}_i(w) = Y_i(w) + \mu_i\left(Y_j(w) - Y_i(w)\right) \qquad i = \end{cases} \quad (2)$$

where $\mu \in [0,1]$ is known as the merging parameter and it controls the speed of meeting [29]. $Y(w)$ and $Q(w)$ are the faith arrangement vectors, separately, at time step w. Union parameter can be tuned to control the development speed of faiths toward agreement; the nearer the incentive to 1 the faiths will advance quicker toward accord.

In the proposed faith formation concept, the social power is presented by choosing the best ant in the area set, before performing any interaction. Social power is the capacity to impact others in a populace. In this way, an ant holding a superior faith will probably be chosen as intermediate ant and thus more inclined to impact different ants in the populace. Counteractive action of being caught in nearby minima is essential for optimization methods. With a specific end goal to do as such, we utilize change procedure by thinking about a likelihood, i.e. transformation rate, for each quality of the faith vector to change autonomously of others. Allocating a change rate $\rho(t)$ for the with cycle of the calculation, the indication of each property in an faith vector is autonomously exchanged with likelihood $\rho(t)$. In the wake of performing the transformation trial faith vector, the cost estimation of the relating arrangement vector is processed and contrasted and that of the present faith vector. At that point, the faith with the lower cost esteem is additionally prepared through a procedure, called assurance update process, and the result of this procedure will be the new faith vector hold by the operator. The conviction update process changes just the confidence values, i.e., the outright values, in the faith vector and leaves the signs unaltered. This procedure considers both the difference in the cost with an incentive in the trial vector as for the present vector and furthermore changes in the faith characteristics. The proposed update process can make the ants increasingly sure in a property of the faith vector by changing the assurance incentive to its greatest conceivable esteem which is 1. An ant may be made indeterminate in a characteristic when it communicates with a questioner having the faith with inverse sign or the one with a similar sign,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

however with less assurance. Diminishing the sureness represents the higher likelihood of low sound flip while expanding the assurance will make it more outlandish. Update system recognizes a subset of characteristics for which the sign has changed through the association or transformation, from those for which the sign stayed unaltered in the trial faith vector. Characterizing an arrangement of traits as:

$$A_i(w) = \left\{a \in \{1,2,\ldots,C\} \middle| \mathrm{sign}\left(\tilde{Y}_i^a(w)\right) \neq \mathrm{sign} \right. \quad (3)$$

The proposed mechanism of conviction update delivers the new faith vector $Y_i(w+1)$ as below:

$$Y_i^c(w+1) = \begin{cases} \mathrm{sign}(Y_i^c(w)), & f\left(\tilde{Q}_i(w)\right) \\ \mathrm{sign}\left(\tilde{Y}_i^c(w)\right), & f\left(\tilde{Q}_i(w)\right. \end{cases} \quad (4)$$

where $\tilde{Q}_i = T(Y_i)$ and alternate properties in the new faith vector, where sign does not contrast in current and trial vector, will be constantly taken from the trial faith vector as beneath:

$$Y_i^c(w+1) = Y_i^c(w), \quad \forall c \in \{1,2,\ldots,C\} AND\ c \quad (5)$$

The objective of the above update process is to screen a talented faith sign not changing. This is accomplished by expanding the sureness in the relating measurement of the faith vector. Keeping in mind the end goal to tackle the feature selection issue, the proposed optimization procedure considers every applicant arrangement Q_i as an encoded ant colony vector that decides selection or end of features in a dataset with ant colony esteems 1 or 0, separately. A classifier with a k-fold cross-approval system is then connected to the dataset, barring every one of the features relating to 0 esteems in the competitor arrangement vector, and the received rate of misclassification is then utilized as the cost an incentive for the arrangement, i.e. $f(Q_i)$. Giving the chance to indicate this optimization method by FS-MACO, pseudo-code is as per the following:

For every ant $i = 1, \ldots, M$
  Produce random vector $Y_i = (Y_i^1, Y_i^2, \ldots, Y_i^C), Y_i^c \in [-1, +1]$ as starting faith
  Calculate $f(Q_i(W))$ that is fitness for the solution by comparing faith of operator i
End for
For every cycle $w = 1, \ldots, w_{max}$
  For every ant $i = 1, \ldots, M$
    1. Decide the ant M_i, for ant i as possibility to perform communication
    2. Select the best ant in M_i as conversationalist j, which is the ant with which ant i will connect
    3. Perform association and produce trial faith vector
    4. Perform transformation on trial vector and calculate the fitness value by comparing solution $f(\tilde{Q}_i(w))$
    5. Perform the process of update and generate the new faith vector as per 1, 2 and 3.
  End for
End for

Five Layered Artificial Neural Network (FLANN) Classifier

The classification undertaking is performed by FLANN. FL-ANN is a Five Layered Outspread Premise Function (RBF) based classifier neural network that makes use of influenced use of descent approach and regression based classification. It enhances smoothing parameter of RBF part through descent approach. It enhances smoothing parameter of RBF part through slant plummet approach. It comprises of five layers named as info, pattern (or design), summation, institutionalization and yield and is depicted the Fig. 1.

Connected info vector x is transmitted to design layer through information layer. Design layer incorporates each preparation datum with RBF part. Squared Euclidean separation between input vector x and preparing information vector t is ascertained

Journal of Theoretical and Applied Information Technology
30th June 2018. Vol.96. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as in (6) where p means aggregate of preparing information at design layer.

$$dist(j)=\|x-t_j\|^2, 1\le j\le p \quad (6)$$

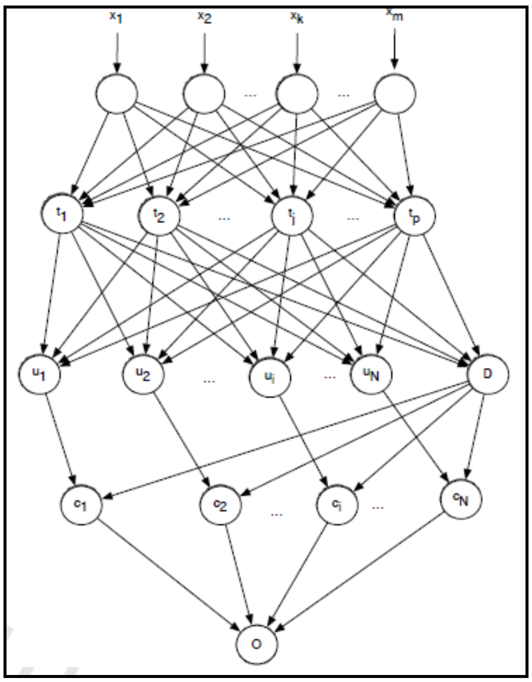

Fig. 1. Five Layered Artificial Neural Network

Computed squared Euclidean separations are utilized as a piece of RBF divide fill in as in (5) where $r(j)$ indicates yield of jth prepared information and speaks to straightening parameter. Yields of RBF piece work are the yield estimations of example layer neurons. What's more, this layer incorporates N target estimations of each preparation datum controlled by contrasting class.

$$r(j)=e^{\left(-1*\frac{dist(j)}{2\sigma^2}\right)}, 1\le j\le p \quad (7)$$

When a preparation datum belongs to i^{th} class then its i^{th} value will be 0.9 and others will be 0.1, as given in (8).

$$y(j,i)=\begin{cases}0.9 \; t_j \; belongs\; to\; i^{th}\; class & 1\le i\le \\ 0.1\; else & 1\le j\le \end{cases} \quad (8)$$

$N+1$ neurons are placed at summation layer where N is the aggregate number of classes and extra one term to N is for one neuron to acquire denominator. PL-ANN uses diverge effect term at summation layer to increase the distances among classes. Diverge effect term value is calculated as in (9) where $d(j,i)$ denotes diverge effect term of j^thtraining information and i^{th} class. y_{max} is initialized to 0.9 which denotes the greatest value of $y(j,i)$. y_{max} value is updated with the most extreme value of yield layer after each iteration of optimization. Diverge effect term is calculated by N neurons of summation layer. This count includes exponential form of $y(j,i)-y_{max}$ to increase the effect of y (j,i).

$$d(j,i)=e^{(y(j,i)-y_{\max})}*y(j,i) \quad (9)$$

Diverge effect term is used in figuring nominator values at summation layer as in (10). Moreover, denominator value is additionally calculated at this layer as in (11).

$$u_i=\sum_{j=1}^{p} d(j,i)*r(j), 1\le i\le N \quad (10)$$

When N neurons, represented with u_i, calculate nominator values by summing spot result of diverging effect terms and pattern layer yields, another neuron calculates denominator value the same as PL-ANN represented by D.

$$D=\sum_{j=1}^{p} r(j) \quad (11)$$

Each class is represented by a neuron at standardization layer. These neurons divide corresponding nominator value by denominator value calculated at summation layer, as per (12) where c_i denotes the normalized yield of i^{th} class.

$$c_i=\frac{u_i}{D}, 1\le i\le N \quad (12)$$

Class of information vector is determined at yield layer through the champ decision mechanism as given in (13) where c is the yield vector of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

standardization layer, c_{id} and id denote champ neuron value and indices of the class, respectively.

$$[c_{id}, id] = \max(c) \qquad (13)$$

Gradient descent based interactive learning is utilized in PL-ANN for getting optimized flattening parameter value. Each preparation datum at pattern layer is sequentially applied to neural network and three steps are executed until the point when greatest iteration confine exceeds. Firstly, squared error e is calculated for each contribution, as in (14) where $y(z, id)$ represents the value of z^{th} preparing input information for id^{th} class and c_{id} is the value of champ class.

$$e = (y(z, id) - c_{id})^2 \qquad (14)$$

4. ABOUT IMPLEMENTATION TOOL AND DATASET

Scilab is one of the fine-free and open source software for numerical computation providing a powerful computing environment for engineering and scientific applications. It has hundreds of mathematical functions inbuilt. It has a high level programming language allowing access to advanced data structures, 2-D and 3-D graphical functions.

The dataset from John Blitzer et al.,2007 was used for experiments. It contains a collection of item reviews from Amazon.com. This dataset contains three types of files positive, negative and unlabelled in XML format. These files were extracted using XML file splitter and reviews were converted into the text file. The dataset contains 1000 positive files and 1000 negative files for each domain. The reviews are about four item domains: Books (B), DVDs (D), Electronics (E) and Kitchen appliances (K) and are written in English language. For the experiment, labeled dataset of 1000 positive and 1000 negative files was used. An instance in each domain is recorded in Table 1. From this dataset, 12 cross-domain sentiment classification errands were constructed: B → D; B → E; B → K; D → B; D → E; D → K; E → B; E → D; E → K; K → B; K → D; K → E, where the word before arrow corresponds to the source domain and the word after an arrow corresponds to the target domain.

Table 1: Negative and positive instances for multi-domain dataset.

Domain Name	Negative Instances	Positive Instances
Book	73500	72794
DVD	66126	76759
Electronics	43806	44321
Kitchen Appliances	36106	36733

5. ABOUT PERFORMANCE METRICS

This research work uses the below-mentioned performance metrics for the performance measure:

- ✓ Accuracy is used as an evaluation measure. Accuracy is the extent of correctly classified examples to the aggregate number of examples; then again, error rate refers to incorrectly classified examples to correctly classified examples. F-measure or precision and recall can be used as evaluation measures.
- ✓ F-measure is just defined in terms of true positive (TP), false positive (FP) and false negative (FN), while true negative (TN) isn't considered. Accuracy and F-measure are compared for a proposed approach which demonstrates that, in general, F-measure is like accuracy.

6. RESULTS AND DISCUSSION

In this research work, the proposed work is compared with the existing methods such as and Supervised word clustering (SWC) [Min Xiao et al.,2013], Spectral feature alignment (SFA) [Min Xiao et al.,2015], Feature Ensemble plus Sample selection (SS-FE) [S.Pan and Q.Yang,2009], Entropy based classifier [Pan SinnoJialin et al.,2010]. The proposed work is termed as FS-MACO-FLANN. When compared with the existing works, the proposed work attains better performance in terms of accuracy. The accuracy attained at the minimum of 88.53% and at the maximum of 91.35%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and is presented in Table-2. The obtained results are portrayed in Fig.2 and Fig.3.

From Table-2, it is evident that Book and DVD, if considered as a source domain, achieve a good compatibility with electronics and kitchen domain, which is considered as target domain. Besides, electronic and kitchen are compatible domains.

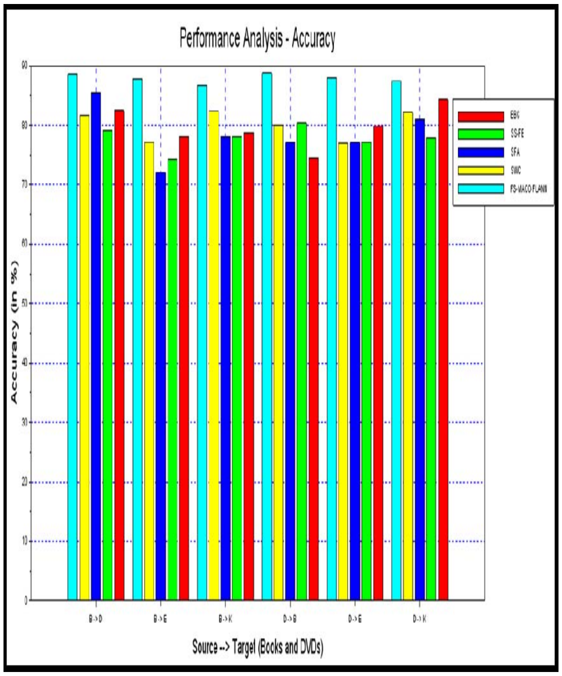

Fig.2. Accuracy for Books and DVDs in Cross Domain Opinion Mining

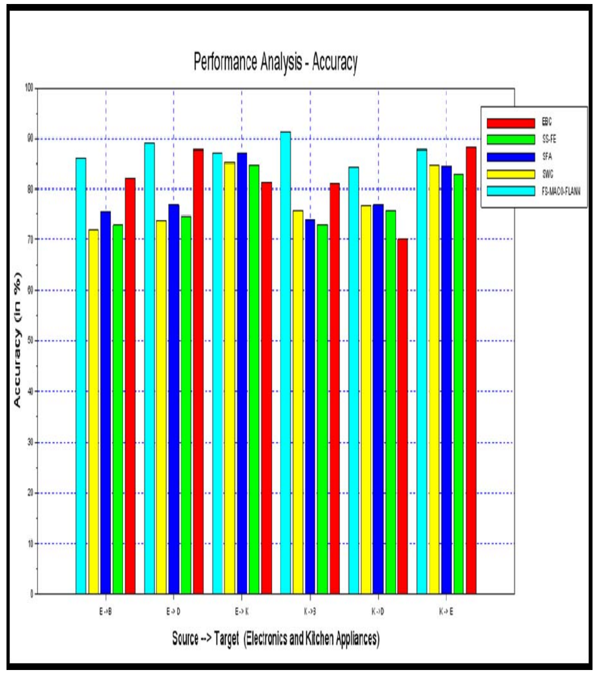

Fig.3. Accuracy for Electronics and Kitchen Appliances in Cross Domain Opinion Mining

Journal of Theoretical and Applied Information Technology
30th June 2018. Vol.96. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table-2: Accuracy

Source – Target	EBC [23]	SS-FE [22]	SFA [21]	SWC [20]	FS-MACO-FLANN(Proposed)
B → D	82.45	79.10	85.55	81.66	88.53
B→ E	78.00	74.24	72.00	77.04	87.86
B→ K	78.65	78.07	78.00	82.26	86.66
D → B	74.35	80.38	77.00	79.95	88.75
D → E	79.78	77.07	77.00	76.98	88.01
D → K	84.21	77.82	81.00	82.13	87.49
E →B	82.15	72.86	75.50	72.11	86.15
E → D	87.8	74.60	77.00	73.81	89.06
E → K	81.44	84.87	87.10	85.33	87.24
K → B	81.05	72.94	74.00	75.78	91.35
K → D	70.00	75.70	77.00	76.88	84.35
K → E	88.35	82.93	84.6	84.78	87.85

Classified words are used to find domain-independent and domain-specific words from the respective domains. Domain-independent and domain-specific words are compared to the SentiWordNet [Baccianella Stefano et al.,2010] and FS-MACO-FLANN, in order to find out how many words match with them (Tables 3–6).

Table – 3: Comparison Of Domain-Specific And Domain-Independent Words Against Sentiwordnet And FS-MACO-FLANN Considering Book (B) As A Source Domain

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet (Baccianella Stefano et al.,2010)		No. of words matching in FS-MACO-FLANN (Proposed)	
			Domain-specific words	Domain-independent words	Domain-words	Domain-independent words
B → D	11503	9744	8934	8972	10028	9364
B → E	5250	4796	4077	4395	4583	4578
B → K	4200	4325	3262	3979	3761	4092

Table – 4: Comparison of domain-specific and domain-independent words against SentiWordNet and FS-MACO-FLANN considering DVD (D) as a source domain

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet (Baccianella Stefano et al.,2010)		No. specific of words matching in FS-MACO-FLANN (Proposed)	
			Domain-specific words	Domain-independent words	Domain-specific words	Domain-independent words
D → B	11130	9744	8644	9009	9278	9348
D → E	5238	4781	4068	4418	4527	4603
D → K	4205	4320	3266	3980	3726	4183

Table – 5: Comparison of domain-specific and domain-independent words against SentiWordNet and FS-MACO-FLANN considering Electronics (E) as a source domain

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet (Baccianella Stefano et al.,2010)		No. of words matching in FS-MACO-FLANN (Proposed)	
			Domain-specific words	Domain-independent words	Domain-specific words	Domain-independent words
E → B	16105	4769	12508	4430	14529	4605
E → D	16466	4781	12789	4462	14358	4598
E → K	4802	3723	3729	3454	4299	3542

Table – 6: Comparison of domain-specific and domain-independent words against SentiWordNet and FS-MACO-FLANN considering Kitchen Appliances (K) as a source domain

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet (Baccianella Stefano et al.,2010)		No. of words matching in FS-MACO-FLANN (Proposed)	
			Domain-specific words	Domain-independent words	Domain-specific words	Domain-independent words
K → B	16549	4325	12853	3980	14508	4108
K → D	16927	4320	13147	3987	14952	4182
K → E	6296	3723	4890	3449	5583	3601

F-Measure is computed and is presented in Table -7. It is evident that F-Measure performance is better than that of accuracy and is portrayed in the Fig. 4. But only single class is considered in F-measure as positive class. On the other hand, when calculating accuracy equal weight is given to both the classes.

Table – 7. Fs-Maco-Flann Accuracy Vs F-Measure

Accuracy	**88.5**	**87.9**	**86.7**	**88.7**	**88.0**	**87.5**	**86.2**	**89.1**	**87.2**	**91.3**	**84.3**	**87.9**
F-Measure	89.6	90.9	89.7	89.7	89.7	92.1	91.2	90.0	92.3	89.7	90.0	90.2

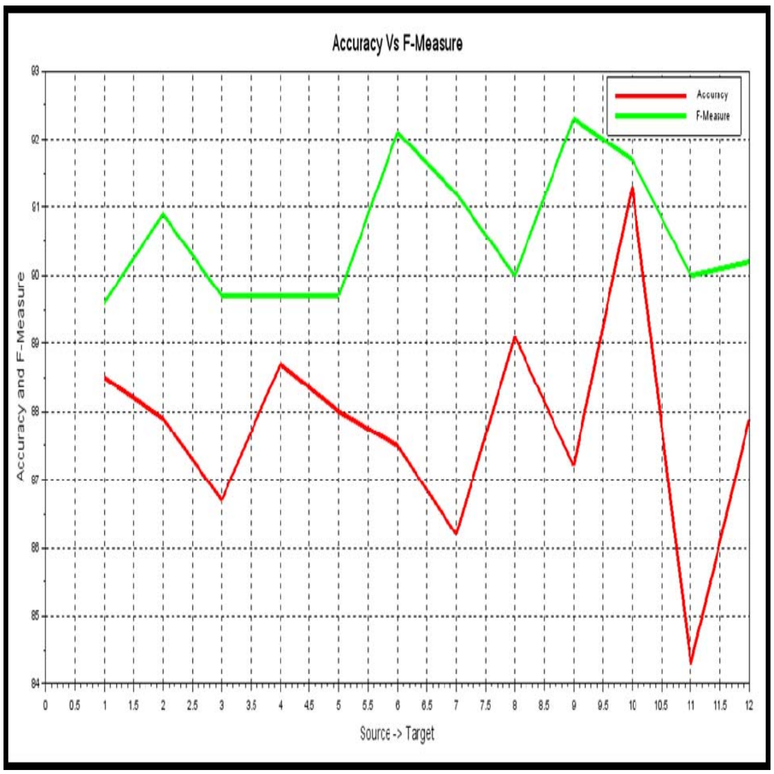

Fig.4. Fs-Maco-Flann - Accuracy Vs F-Measure

7. CONCLUSIONS

This research article aims to propose Feature Selection using Modified Ant Colony Optimization Approach based Five Layered Artificial Neural Network (FS-MACO-FLANN) for Cross Domain Opinion Mining. The dataset contains a collection of product reviews from Amazon.com that has three types of files positive, negative and unlabelled in XML format. These files were extracted using XML file splitter and reviews were converted into text file. The dataset contains 1000 positive files and 1000 negative files for each domain. The reviews are about four product domains: Books (B), DVDs (D), Electronics (E) and Kitchen appliances (K) and are written in English language. For experiment, labeled dataset of 1000 positive and 1000 negative files was used. FS-MACO-FLANN is compared with the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

existing works such as supervised word clustering (SWC), spectral feature alignment (SFA), Feature Ensemble plus Sample selection (SS-FE) and Entropy based classifier. When compared with the existing works, the proposed work attains better performance in terms of accuracy and F-measure.

REFERENCES:

[1] Baccianella Stefano, Esuli Andrea, Sebastiani Fabrizio, SentiWordNet 3.0: An Enhance Lexical Resource for Sentiment Analysis and Opinion Mining, in: Proceedings of the 7th Language Resources and Evaluation Conference (LREC 2010), Valletta, Malta, May 17–23, 2010, pp. 2200–2204.

[2] John Blitzer, Mark Dredze, Fernando Pereira, Biographies, Bollywood, boomboxes and blenders: domain adaptation for sentiment classification, in: Association of Proceedings of the 45th Annual Meeting of the Computational Linguistics (ACL), Prague, Czech Republic, June 2007, pp. 440–447.

[3] Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan, Ashraf Ullah, Mining opinion components from unstructured reviews: A review, Journal of King Saud University - Computer and Information Sciences, Volume 26, Issue 3, September 2014, Pages 258-275.

[4] M. Kolahkaj and M. Khalilian, "A recommender system by using classification based on frequent pattern mining and J48 algorithm," 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, 2015, pp. 780-786.

[5] M. Roccetti, A. Casari and G. Marfia, "Inside chronic autoimmune disease communities: A social networks perspective to Crohn's patient behavior and medical information," 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, 2015, pp. 1089-1096.

[6] Marios Belk, Efi Papatheocharous, Panagiotis Germanakos, George Samaras, Modeling users on the World Wide Web based on cognitive factors, navigation behavior and clustering techniques, Journal of Systems and Software, Volume 86, Issue 12, December 2013, Pages 2995-3012.

[7] Min Xiao, Feipeng Zhao, Yuhong Guo, Learning latent word representations for domain adaptation using supervised word clustering, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 152–162, Seattle, Washington, USA, 18–21 October, 2013.

[8] Min Xiao, Yuhong Guo, Feature space independent semi-supervised domain adaptation via kernel matching, IEEE Trans. Pattern Anal. Mach. Intelligence 37 (1) (2015) 52–66.

[9] Mohamed M. Mostafa, Ahmed A. El-Masry, Citizens as consumers: Profiling e-government services' users in Egypt via data mining techniques, International Journal of Information Management, Volume 33, Issue 4, August 2013, Pages 627-641.

[10] N.M.A.Al-Yazeed, A.M.Gadallah, H.A.Hefny, "A hybrid recommendation model for web navigation," 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, 2015, pp. 552-560.

[11] P. Lakar, S. P. Samal, S. R. Muthupandi and N. B. Patil, "Smart web miner - extending web browser with mining framework based on user behavior & web-of-thing patterns for web personalization," Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on, Noida, 2015, pp. 522-527.

[12] P. Suppa and E. Zimeo, "A Context-Aware Mashup Recommender Based on Social Networks Data Mining and User Activities," 2016 IEEE International Conference on Smart Computing (SMARTCOMP), St. Louis, MO, 2016, pp. 1-6.

[13] Pan SinnoJialin, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, Zheng Chen, Crossdomain sentiment classification via spectral feature alignment, in: Proceedings of the 19th International World Wide Web Conference, ACM, Raleigh, USA, April 26–30, 2010.

[14] Pei-Ling Hsu, Hsiao-Shan Hsieh, Jheng-He Liang, Yi-Shin Chen, Mining various semantic relationships from unstructured user-generated web data, Web Semantics: Science, Services

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and Agents on the World Wide Web, Volume 31, March 2015, Pages 27-38.

[15] S. Aggarwal and V. Mangat, "Application Areas of Web Usage Mining," 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, 2015, pp. 208-211.

[16] S.Pan, Q.Yang, A survey on transfer learning, IEEE Trans. Knowledge Eng. 22 (10) (2009) 1345–1359.

[17] Sumaiya Kabir, Shamim Ripon, Mamunur Rahman, Tanjim Rahman, Knowledge-based Data Mining Using Semantic Web, IERI Procedia, Volume 7, 2014, Pages 113-119.

[18] T. Gopalakrishnan, P. Segottuvelan and J. Sathyamoorthy, "User Profile Discovery for Web Search," Intelligent Computing Applications (ICICA), 2014 International Conference on, Coimbatore, 2014, pp. 377-381.

[19] X. Gong, X. Guo, R. Zhang, X. He and A. Zhou, "Search Behavior Based Latent Semantic User Segmentation for Advertising Targeting," 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, 2013, pp. 211-220.