Multi-Source Domain Adaptation in Sentiment Analysis using Optimized Neural Network and Cross-Domain Semantic Library

Dipak Patel[1]*   Kiran Amin[2]

[1]*Information Technology Department, Vishwakarma Government Engineering College, India*
[2]*Executive Dean, Faculty of Engineering & Technology, Ganpat University, India*
* Corresponding author's Email: dipak.eng@gmail.com

Abstract: Nowadays, the rapid growth of internet has led to the way for easiest data generation. Sentiment Analysis (SA) is necessary to scrutinize the user-generated data for classifying the sentiment polarity into anyone of the two classifications (negative and positive). SA is implemented by machine learning techniques and lexicon oriented techniques. However, the traditional SA models could not deal with missing labelled data and internet slang data. This paper aims to propose a new multi-source sentiment analysis model that includes six stages to perform the task of domain adaptation. For checking similarity between statements, higher order statistics based on text features is exploited. The higher order statistics based features extraction includes use of modified cross entropy measure. Remaining features are then given to proposed classifier that predicts the polarity of target domain in a precise way. For classification purpose Neural Network (NN) is exploited. Particularly, the weights of NN are tuned in an optimal manner using Improved Grey Wolf Optimization (IGWO) algorithm, which is the enhanced version of GWO algorithm. Finally, the performance of the proposed model is compared over several conventional methods of domain adaptation and gained an improvement of 28% to 6% in accuracy.

Keywords: Sentiment classification, Enhanced cross entropy, Neural network, Improved GWO algorithm.

Nomenclature

Abbreviation	Description
NLP	Natural Language Processing
SA	Sentiment Analysis
CDSC	Cross-Domain Sentiment Analysis
NN	Neural Network
DNN	Deep Neural Network
IGWO	Improved Grey Wolf Optimization
CNN	Convolutional Neural Network
RNN	Recurrent Neural Networks
K-NN	K-Nearest Neighbour
CN	Capsule Network
CITK	Caps Net with Identifying Transferable Knowledge
BERT	Bidirectional Encoder Representations from Transformers
HANP	Hierarchical Attention Network with Prior knowledge information
CTN	Cloze Task Network
CE	Cross Entropy
CHAN	Convolutional Hierarchical Attention Networks
GA	Genetic Algorithm
SVM	Support Vector Machine
NPV	Negative Predictive Value
ECE	Enhanced Cross Entropy

1. Introduction

Sentiment classification is a significant task in NLP and it is necessary to be aware of views of users in social networks regarding the product reviews [1, 2]. This task intends to forecast the sentiment polarity (which may be negative or positive) of a data. Generally, individuals make remarks on blogs, product review sites, societal media platforms, etc [3, 4]. On the other hand, polarity of a word to express a view varies in diverse domains. For instance, a phrase like "take too long is positive in the electronic domain, but negative in the restaurant domain" [5, 6]. Due to the rising count of fields, annotation of data turns out

540

to be a "labour-intensive activity". Therefore, a "cross-domain sentiment analysis technology" is required for predicting the dataset of the unlabelled target field accurately through learning and training in the labelled source field [7-9].

For the previous decades, CDSC has grown-up from an intellectual tool to a crucial analysis tool [10-11]. The most general and significant issue related with the "cross-domain sentiment classification is domain adaptation". It intends in mapping the data features of diverse domains to the similar feature space [12] and therefore the data from other domains could be exploited for enhancing the training in the target domain. Here, the major aspect relies in minimizing the distribution variation among domains and accordingly the variation of data distribution among target domain and source domain should be solved effectively [13]. This paper aims to provide domain adaptation by matching features of available source domains to the features of target domain and solves the problem of data labelling.

Existing techniques are separated into two groups depending on the information and knowledge they employ: corpus- oriented techniques and lexicon-oriented techniques. The latter is based on the number of positive and negative words in the document and the former is a machine learning technique that is used for training a classifier to forecast the unlabelled target field and it has gained a more attraction nowadays due to its accuracy [14-17].

The major contribution of this research work is:

- Introduces a new "cross domain adaptation based sentiment classification" with proposed Cross-domain similarity and combination of NN and IGWO algorithms.
- Optimization assisted NN is introduced in this work, where the training will be carried out by a new improved optimization algorithm via tuning the optimal weights.
- Introduces a novel IGWO algorithm, which is the modified version of grey wolf optimization algorithm with new neighbourhood field based updating evaluation.

The rest of the paper is organized as: Section 2 portrays the reviews done under this topic. Section 3 depicts the brief description of the proposed sentiment domain adaptation model. Further, Section 4 addresses the keyword extraction: formation of frequency matrix and distribution table. Construction of the sentiment knowledge graph is depicted in Section 5. Section 6 describes the sentiment classification using proposed model. The proposed improved GWO algorithm is discussed in Section 7 and results and comparative analysis is briefed in Section 8.

2. Literature review

In this section, the existing domain adaptation based sentiment analysis techniques are described and their drawbacks are discussed.

In 2019, Zhao *et al.* [18] have established a new approach with "multi-source domain adaptation for CDSC tasks". This model exploited CNN and "bi-directional gated recurrent units" for soft parameter sharing and deep feature extraction for transferring information among tasks. Finally, the investigational results have pointed out that the aspect-level sentiment classification has been improved by the adopted model.

In 2019, Miguel *et al.* [19] have introduced a novel method for optimizing a set of existing Sentiment Analysis (SA) techniques in an ensemble classifier based on the input text domain. Here, an exact group of varied SA techniques has lessened the domain adaptation issue. As a final point, the outcomes have revealed the effectiveness of the adopted technique by showing enhancement on the performance of sentiment analysis.

In 2019, Xing *et al.* [20] have introduced a novel method for training word polarities and vanilla sentiment classifier to the target field simultaneously. In particular, the erroneously predicted sentences were tracked sequentially and exploited as the supervision. Investigational outcomes on numerous renowned datasets demonstrated the betterment of the adopted model in terms of sentiment classification for several domains using sentimental lexicons.

In 2019, Yin *et al.* [21] have established a novel "CITK method for cross-domain sentiment classification". The adopted scheme has exploited capsule network for encoding the domain invariant knowledge that bridged the information gap among the target and source domains. Moreover, BERT was exploited for converting sentences to equivalent length and it was termed as pre-training, which attained more comprehensive semantic embedded model. In the end, investigational outcomes have demonstrated the betterment of the adopted model over the traditional models.

In 2016, Bollegala *et al.* [22] have modelled "embedding learning", and constructed three operations that captured: (a) pivots properties of (i.e., general characteristics that emerge in both target and source domains), (b) label parameters in the documents of source, and (c) geometric features in both target and source domains. Accordingly, the adopted "joint optimisation" technique has

discovered embeddings, which were susceptible to SA classification. Finally, the outcomes of the presented approach have established the improved performance of the presented model in terms of SA classification.

In 2019, Manshu and Bing [23] have established a HANP framework for carrying out CDSC task. The presented approach has obtained both domain specific features and domain independent features by considering the previous knowledge. Moreover, the HANP integrated an attention based technique and therefore important sentences and words associated with the sentiments can be captured. At the end, experimentations on the "Amazon review datasets" have demonstrated that the adopted HANP considerably performed better than the existing schemes.

In 2019, Manshu and Xuemin [24] have proposed an end to end model. This approach included two divisions; one was a "CTN, and another was a CHAN". The latter one captured significant sentences and words regarding sentiments based on the attention oriented model. As a final point, analysis was performed using "Amazon review datasets" for validating the enhancement of the adopted scheme.

In 2019, Wei *et al.* [25] have established a general "domain specific feature transfer framework" that linked up diverse domains by means of general features. The presented work also offered minimal domain variances. Particularly, the translations among domain specific features and general features were discovered and they were cross-used for transferring the domain based features of one domain to another. At last, wide-ranging experimentations confirmed the efficiency of the implemented technique.

ConvNets method was introduced in [18], which maximizes the accuracy and it also offers reduced error. However, multi modal transfer has to be focused more and need to have some labelled data in target domain. GA was exploited in [19] that offer optimal F-measure and it also provides improved precision, but it has to focus more on the unbalanced datasets and their model uses pre trained models on some domains for DA task. It may not generalize on source domains. SVM model was used in [20] that offer high accuracy and there is no negative learning. However, it is not considered for concept level adaptation of sentiments. In addition, CNN model was implemented in [21] that offers reduced loss and it has got better accuracy but it have to be executed for complex sentence patterns to find negative words and multi-source inputs was not considered. K-NN model was presented in [22] that offer improved

accuracy with reduced complexity, but, it needs consideration on different rules. Moreover, HANP method was implemented in [23] that provide enhanced accuracy along with reduced noise. However, it has to discriminate the various meanings of dis-pivots and the model was very complex. In addition, CCHAN method was suggested in [24] which offer high accuracy and it can be applied easily for other domains. However, it requires analysis on time consumption and for multi-source domains, it becomes more complex. SVM was used in [25] which provide F-measure and it also offers better sensitivity. However, it is more complex. Our multi-source domains adaptation model is very simple and it improves the generalization capability of classification task on target domain and achieves a significant improvement in accuracy compare to other models discussed here.

3. A brief description on proposed sentiment domain adaptation model

3.1 Problem definition

In multi-source domain adaptation problem, there are multiple source domains with labelled data and only one target domain with unlabelled data. The task is to classify the target domain data using machine trained on source domains. This task is formulated as below:

Consider X_s is a set of source domains. Let $X_s = \{D_{s_1}, D_{s_2}, ... D_{s_n}\}$ and each source domain D_{s_i} is consist of domain specific features as well as some common features and corresponding labels which is represented as $D_{s_i} = \{X_{s_i}, Y_{s_{i,j}}\}$, Where $X_{s_i} \in R^{ns_i \times (C_i + S_i)}$ and $\{Y_{s_{i,j}}\}_{j=1}^{ns_i} \in \{0,1\}$. Here ns_i denotes the number of rows (which contains review statements) in source domain i, S_i is the source domain i specific features and C_i represents the common features available in domain i. $Y_{s_{i,j}}$ is the label of jth row of domain i. $D_T = \{X_t\}$ Where $X_t \in R^{n_t \times (C+T)}$. Here n_t denotes the number of rows (which contains review statements) in target domain and T is the target domain specific features and C represents the common features available in target domain. The task is to learn the classifier trained on source domain set X_s and predict the sentiment polarity of X_t.

3.2 Proposed architecture

The presented work intends to introduce a new framework for tagging sentiments for a target domain by labelled data from a source domain. This includes major phases like (i) pre-processing (ii) Keyword
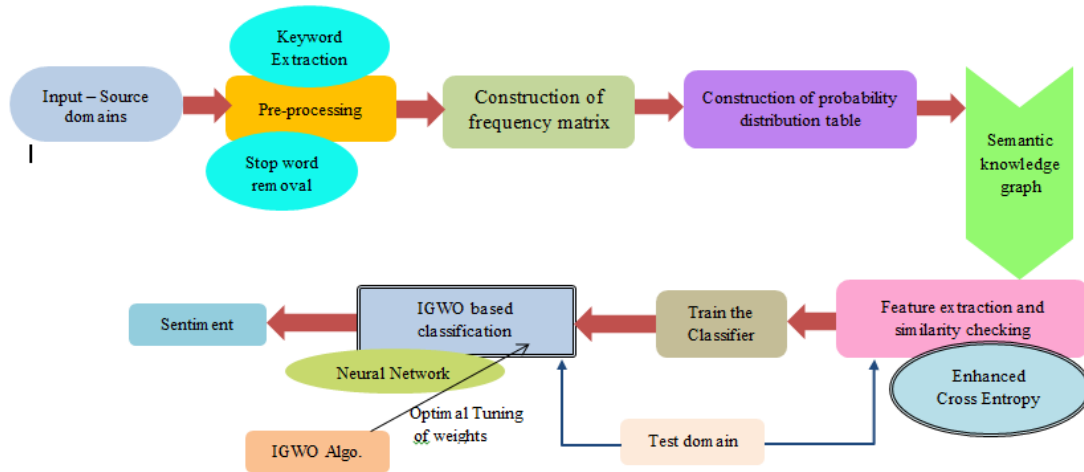
Figure. 1 Block diagram of proposed sentiment domain adaptation framework

Extraction (iii) Frequency matrix construction (iv) Distribution table formation (v) Feature Extraction: semantic knowledge graph construction (vi) Similarity checking using modified cross entropy between target domain and source domains based on probability distribution (vii) Sentiment classification, illustrated in Fig 1. In this context, we explain the proposed sentiment classification process on considering four domains such as; book, DVD, electronics review and kitchen appliances review.

Pre-processing is the primary step, where stop word removal takes place and the key words are extracted. Subsequent to the keyword extraction, frequency matrix and probability based distribution table are constructed. Accordingly, feature extraction from target domain and similarity checking of extracted features with the source domains is performed that exploits modified cross entropy measure. The most similar features of target domain to source domains are extracted and given to classifier in addition to source domains data for the purpose of enhancing the training features. The rest of target domain data is given to classifier for the purpose of classification. For carrying out accurate classification, NN is deployed with IGWO algorithm. In order to enhance the classification accuracy, the NN model is trained using a new IGWO algorithm via selecting the optimal weights.

4. Keyword extraction: formation of frequency matrix and distribution table

4.1 Pre-processing

This is the preliminary step; where the keywords are extracted from each domain that is carried out using stop word removal. Let d_1, d_2 and d_3 be the three considered source domains namely, book review, electronics review and kitchen appliances

Table 1. Extraction of keywords from each domain

Domain	Keywords
d_1	w_1 w_2 w_3 w_4 w_5 w_6
d_2	w_2 w_5 w_6 w_7
d_3	w_1 w_3 w_4 w_5 w_7

respectively. From each domain, the keywords denoted by w are extracted.

Stop word removal and Keyword Extraction: Usually, the stop words are articles and pronouns that do not offer meaning to the sentence. When the stop words are recognized to be malicious, they have to be eliminated. Based on Stop word, classification decision cannot be taken as they misguide the classifiers. This removal minimizes the term space dimensionality. Thereby, the key words are extracted. For example, consider Table 1 is formed as a result of keyword extraction from domains.

4.2 Frequency matrix construction

From the extracted key words, the frequency matrix is formed. The frequency matrix shows the number of occurrences of each word in each domain. It also includes the total number of words present in each statement of the domain as shown in Table 2.

When a new target domain d_4 arrives into the network, the frequency matrix will be formed for the newly arrived domain with relevant particulars along with the source domains.

Table 2. Frequency matrix construction for source domains

Domain	w_1	w_2	w_3	w_4	w_5	w_6	w_7	S
d_1	5	2	4	5	7	2	0	25
d_2	0	3	0	0	5	2	3	13
d_3	7	0	5	2	3	0	5	22

Table 3. Frequency matrix construction for source domains as well as target domains

D	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	S
d_1	5	2	4	5	7	2	0	0	25
d_2	0	3	0	0	5	2	3	0	13
d_3	7	0	5	2	3	0	5	0	22
d_4	0	2	0	3	0	0	7	5	17

Table 4. Distribution Table construction for source domains as well as target domains

D	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
d_1	0.20	0.08	0.16	0.20	0.28	0.08	0	0
d_2	0	0.23	0	0	0.38	0.15	0.23	0
d_3	0.32	0.00	0.23	0.09	0.14	0	0.23	0
d_4	0	0.12	0	0.18	0	0	0.41	0.29

For example, on assuming the extracted key words of d_4 as w_2, w_4, w_7 and w_8 and if the frequency of occurrence is 2, 3, 7 and 5 then the frequency matrix is formed as shown in Table 3. Here in Table 3, Domain is indicated by D and sum of frequency of words is indicated by S.

4.3 Constructing the distribution table

Following the formation of frequency matrix, the probability distribution table is constructed. The distribution table is formed by dividing the elements' frequency of each keyword in a domain with the sum of all words in that particular statement (here only one statement shown in domain). Table 4 shows the exemplary representation of distribution table with respect to the frequency matrix constructed in Table 3.

5. Constructing sentiment knowledge graph: determining the similarity using enhanced cross entropy measure

5.1 Cross entropy and enhanced cross entropy

The original formula of CE between review statements of source domain S and Target domain T is shown in Eq. (1), where $P(x_i)$ denotes the probability distribution of words in one review statement of target domain d_4 and similarly $Q(x_i)$ denotes the probability distribution of words in one review statement of source domains (d_1, d_2 and d_3).

$$CE(S,T) = -\sum_{P(x_i)) \in S,\ Q(x_i)) \in T} P(x_i) \log Q(x_i) \quad (1)$$

The Formula for enhanced cross entropy is given in following Eq. (2):

$$ECE(S,T) = -(1 - |P(x_i) - Q(x_i)|) \cdot \sum_{P(x_i) \in S, Q(x_i \in T)} (1 - P(x_i)) \log(Q(x_i))$$
$$(2)$$

Where $P(x_i)$ and $Q(x_i)$ are non-zero values.

Here multiplication factor in enhanced cross entropy formula boosts the score of vector with similar probability distribution and with this enhancement we want to discourage the score of words having high probability as that can be stop word. Using above Eq. (2) of enhanced cross entropy, similarity between two probability distribution of source domain and target statement is calculated.

On considering the target domain d_4 and source domain d_1, enhanced cross entropy between $P(x_i)$ and $Q(x_i)$ can be evaluated as shown in Table 5, where the probability of keywords is taken from distribution table (Table 4).

5.2 Sentiment knowledge graph

In the proposed model, the similarity between the reviews of the target domain and source domains are determined from the enhanced cross entropy (ECE). The semantic knowledge graph determines the probability among the target domain and each source domains. This probability value determined among target domain and each source domain is utilized for computing the ECE. For example, the semantic knowledge graph between source domain d_1 and target domain d_4 is constructed as per Table 5. In Table 5, word vector of domain d_1 is (0.2, 0.08, 0.16, 0.2, 0.28, 0.08, 0, 0) and similarly for domain d_4 is (0, 0.12, 0, 0.18, 0, 0, 0.41, 0.29). Computation of ECE is shown in Table 5 and final score is also computed as 1.5.

In the same way, the probability of words in each statement of target domain d_4 and each statement of source domains d_1, d_2 and d_3 should be computed for

Table 5. Construction of sentiment knowledge graph between source domain and target domain

D	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
d_1	.2	.08	.16	.2	.28	.08	0	0
d_4	0	.12	0	.18	0	0	.41	.29
ECE	0	.9	0	.6	0	0	0	0
Final Score	1.5							

544

determining the ECE. The review statements from target domain having high similarity with the statements in source domains are assigned the polarity similar to source domain statement and now features from these review statements having known polarity helps to extend the training dataset in later stage of IGWO based NN classification.

6. Sentiment classification using optimized neural network

In our proposed work, we have exploited optimized neural network for improvement in classification accuracy. In optimized neural network, weights are tuned using improved grey wolf optimizer.

6.1 Optimized neural network

Optimized NN is used for sentiment classification. As the work concerns on three source domains, the keywords of those domains are trained using NN model. Here, the extracted keywords (words) of domains denoted by w are subjected to NN classification. NN [26] considers the words w as input specified by Eq. (3), where nu signifies the total count of words.

$$w = \{w_1, w_2, ..., w_{nu}\} \qquad (3)$$

The model includes input, output, and hidden layers. The output of the hidden layer $e^{(H)}$ is defined as in Eq. (4), where F refers to the "activation function", $\hat{i}$ and j refers to the neurons of hidden and input layers correspondingly, $W_{(B\hat{i})}^{(H)}$ denotes bias weight to $\hat{i}^{th}$ hidden neuron, n_i symbolizes count of input neurons and $W_{(J\hat{i})}^{(H)}$ denotes the weight from j^{th} input neuron to $\hat{i}^{th}$ hidden neuron. The output of the network $\hat{G}_0$ is determined as in Eq. (5), where $\hat{o}$ refers to the output neurons, n_h indicates the number of hidden neurons, $W_{(B\hat{o})}^{(G)}$ denotes output bias weight to the $\hat{o}^{th}$ output layer, and $W_{(\hat{i}\hat{o})}^{(G)}$ specifies the weight from $\hat{i}^{th}$ hidden layer to $\hat{o}^{th}$ output layer. Consequently, the error amongst the predicted and actual values is computed as per Eq. (6) that should be reduced. In Eq. (6), n_G symbolizes the output neuron count, $G_{\hat{o}}$ and $\hat{G}_{\hat{o}}$ refers to the actual and predicted output respectively.

$$e^{(H)} = F\left(W_{(B\hat{i})}^{(H)} + \sum_{j=1}^{n_i} W_{(J\hat{i})}^{(H)} w\right) \qquad (4)$$

$$\hat{G}_{\hat{o}} = F\left(W_{(B\hat{o})}^{(G)} + \sum_{j=1}^{n_h} W_{(\hat{i}\hat{o})}^{(G)} e^{(H)}\right) \qquad (5)$$

$$Er^* = argmin \sum_{\{W_{(B\hat{i})}^{(H)}, W_{(J\hat{i})}^{(H)}, W_{(B\hat{o})}^{(G)}, W_{(\hat{i}\hat{o})}^{(G)}\}=1}^{n_G} |G_{\hat{o}} - \hat{G}_{\hat{o}}| \qquad (6)$$

As mentioned above, the training of NN model is carried out using a new IGWO algorithm via optimizing the weights $W = W_{(B\hat{i})}^{(H)}$, $W_{(J\hat{i})}^{(H)}$, $W_{(B\hat{o})}^{(G)}$ and $W_{(\hat{i}\hat{o})}^{(G)}$. Thus, the sentiments of words are attained as outputs. The objective function OF of the presented work is defined in Eq. (7)

$$OF = Min\ (Er^*) \qquad (7)$$

7. Improved optimization framework for NN training

7.1 Proposed IGWO algorithm

Though conventional GWO approach includes various advantages; it is encountered with certain drawbacks such as, local optima, slow convergence and so on. Hence, to prevail over the limitations of the existing model, certain enhancements are made to introduce a new improved model named as IGWO. The procedure of improved GWO model [27] is as follows: The wolves α, β, and γ are said to be the major wolves that focus on the process of hunting. Among these wolves, α is considered as the leader that makes decisions relating to hunting process, sleeping location, time to awake,etc. whereas, β, and γ holds the 2nd and 3rd level that helps α in taking decisions. In addition, the final level of wolves is concerned as ζ, which concerns on eating. The encircling characteristics are modeled as per Eqs. (8) and (9), where G and L denotes coefficient vectors, J_p indicates prey's position vectors, J denotes position vectors of grey wolves and it specifies current iteration. Eqs. (10) and (11) denotes the model for G and L, where $\hat{a}$ is a parameter which is minimized steadily from 2 to 0 in entire iterations. Here, ra_1 and ra_2 specifies the random vectors that lie among [0, 1] and it_{max} denotes the maximum iteration.

$$Z = |L.J_p(it) - J(it)| \qquad (8)$$

$$J(it+1) = J_p\ (it) - G.Z \qquad (9)$$

$$G = 2.\hat{a}.ra_1 - \hat{a} \qquad (10)$$

$$L = 2.ra_2 \qquad (11)$$

The arithmetical formula for describing the hunting character of wolves is given from Eqs. (12) to (17). Here, the proposed hybrid contribution takes place during the final position updating formulation of GWO. Conventionally, the position update of GWO takes place based on the positions of α, β, and γ. However, in the proposed logic, the evaluation of $J(it+1)$ also takes account of the neighbourhood field (J_k) [28] as shown in Eq. (18). Accordingly, J_k is computed as per Eq. (19), where J^+ and J^- denote the immediate best and worst solutions of J. With this modification, it reaches to solution in less number of iterations and increases the accuracy value as discussed in result section.

$$Z_\alpha = |\, L_1 \,.\, J_\alpha - J \,| \tag{12}$$

$$Z_\beta = |\, L_2 \,.\, J_\beta - J \,| \tag{13}$$

$$Z_\gamma = |\, L_3 \,.\, J_\gamma - J \,| \tag{14}$$

$$J_1 = J_\alpha - G_1 .\, (Z_\alpha) \tag{15}$$

$$J_2 = J_\beta - G_2 .\, (Z_\beta) \tag{16}$$

$$J_3 = J_\gamma - G_3 .\, (Z_\gamma) \tag{17}$$

$$J\,(it+1) = \frac{J_1 + J_2 + J_3 + J_K}{4} \tag{18}$$

$$J_K = J + ra_1\,(J^+ - J^-) + ra_2\,(J^+ - J^-) \tag{19}$$

The pseudo code of the proposed IGWO algorithm is presented below.

Proposed IGWO Algorithm

Initialization

Evaluate the fitness of entire search agents

Set J_α as best search agent

Set J_β as 2nd best search agent

Set J_δ as 3rd best search agent

While $(it < it_{max})$

   For every wolf

     Update position as per Eq. (18)

   End for

   Update $\hat{a}$, G and L

   Evaluate fitness for entire search agents

   Update J_α, J_β and J_δ

   $it = it + 1$

End while

Return J_α

---

8. Results and discussion

8.1 Experimental setup

The proposed Sentiment Domain Adaptation model was implemented in Python and the results were observed. Here, evaluation was done using dataset available at "https://github.com/hsqmlzno1/ HATN" [Access date: 2021-01-11]. It is a dataset containing review on various domains like Book, DVD, Electronics, Kitchen appliances and Video. We haven't considered the video domain for our experiment. Accordingly, the betterment of the proposed IGWO model for domain adaptation was compared over other traditional models of domain adaptation as KNN: term similarity [22], KNN: geometric + term similarity [22], HANP [23], CCHAN [24], GWO+NN [26-27]. Here, the performance of proposed work was compared over other traditional models of domain adaptation with respect to measures like accuracy, specificity, precision, recall, F-measure, and NPV. These evaluation measures are explained in next section.

8.2 Evaluation measures

Here in this section, explanation of evaluation measures is given. Consider the confusion matrix which is constructed based on actual data labels and predicted data labels as shown in Table 6.

Precision is a ration of correctly identified positive labels to all predicted positive labels. So Precision= (TP/(TP+FP)). Recall is a ratio of correctly identified positive labels to actual positive labels that is Recall=(TP/(TP+FN). F-measure is a harmonic mean of precision and recall. Which is formulated as F-measure= 2 x (Precision x Recall)/(Precision + Recall).

Specificity is a ration of correctly identified negative labels to actual negative labels present in dataset that is Specificity = (TN/(TP+FP)). Accuracy is a portion of correctly predicted labels to all predicted labels which is given as Accuracy = (TP/(TP+TN+FP+FN)). Negative Predicted Value (NPV) is a probability that how negative test prediction is accurate which is given by NPV=(TN/(TN+FN)).

Table 6. Confusion matrix

		Actual labels	
		Positive	Negative
Predicated	Positive	TP	FP
	Negative	FN	TN

8.3 Performance analysis

The performance of proposed model over the conventional models of domain adaptation with respect to various measures is represented in Fig. 2 and the some of the measures have been tabulated in Table 7. The analysis was carried out for four domains namely, book, DVD, kitchen and electronics. We kept one domain as a test domain and rest of the domains as source domains. For each test domain, result is shown in Fig. 2. On noticing the outcomes, the presented IGWO+NN model has accomplished

better performance when compared over the existing models for all four domains. Here, from Fig. 2(a), the presented method has achieved an accuracy of 93.00, 92.20, 92.40 and 92.60 for test domain as Book, DVD, Kitchen and Electronics respectively over other conventional models. Which is 27.96%, 21.29%, 10.97%, 07.96% and 5.81% higher to KNN-term similarity, KNN- geometry + term similarity, CCHAN, HANP and GWO + NN models when Book domain is the selected as a test domain. From Fig. 2(b), the specificity of the proposed method for
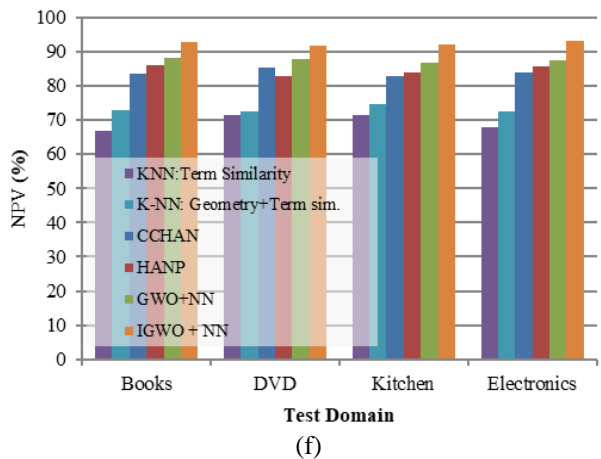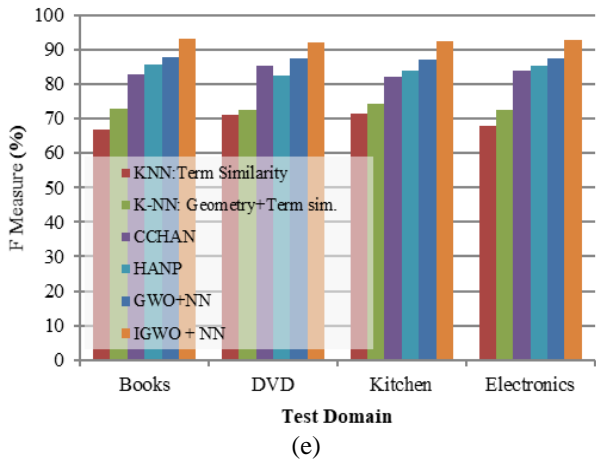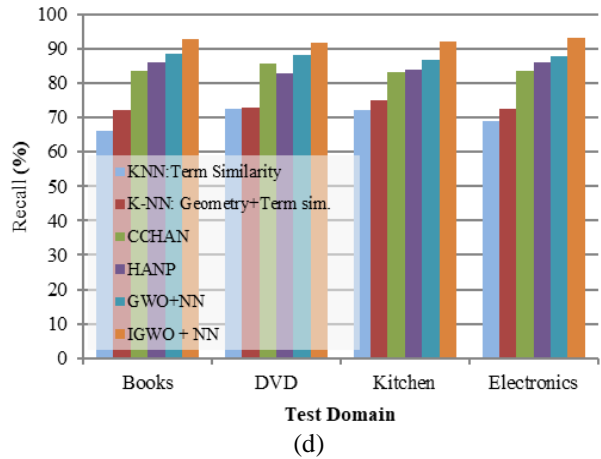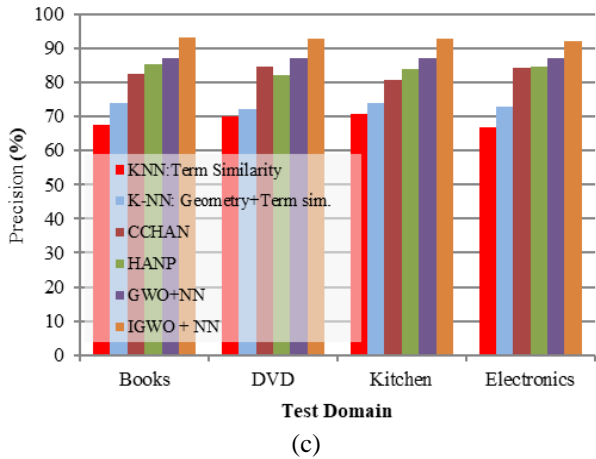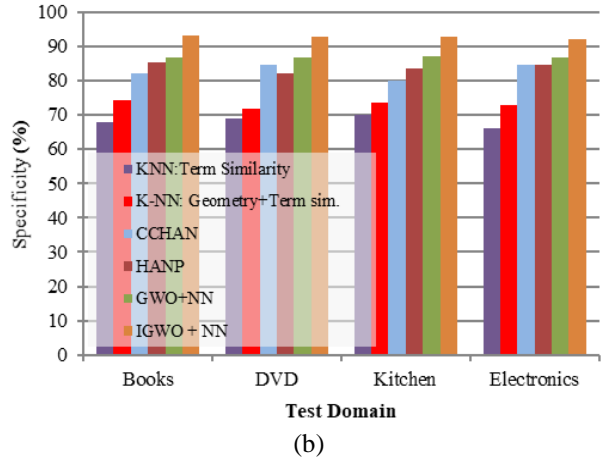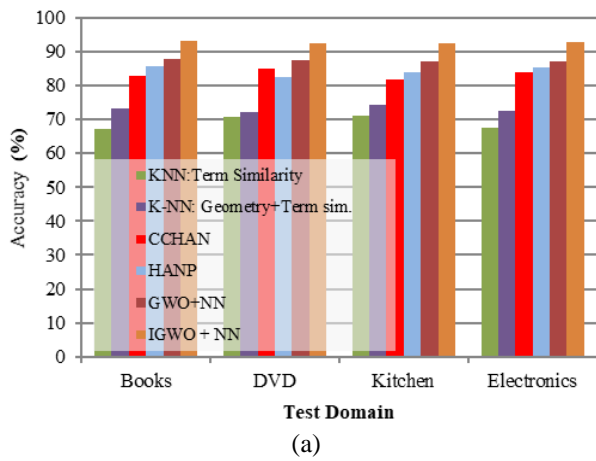

Figure. 2 Performance analysis of presented approach over traditional schemes in terms measures such as: (a) Accuracy, (b) Specificity, (c) Precision, (d) Recall, (e) F-measure, and (f) Negative Predictive Value (NPV)

547

Table 7. Performance analysis of the presented work over the existing work

Accuracy in %						
Test Domain	KNN:Term Similarity	K-NN: Geometry+ Term sim.	CCHAN	HANP	GWO+NN	IGWO + NN
Books	67.00	73.20	82.80	85.60	87.60	93.00
DVD	70.60	72.20	85.00	82.40	87.40	92.20
Kitchen	71.00	74.20	81.60	83.80	87.00	92.40
Electronics	67.40	72.60	84.00	85.20	87.20	92.60
F-Measure in %						
Test Domain	KNN:Term Similarity	K-NN: Geometry+ Term sim.	CCHAN	HANP	GWO+NN	IGWO + NN
Books	66.67	72.87	82.94	85.66	87.70	92.99
DVD	71.12	72.37	85.09	82.47	87.48	92.15
Kitchen	71.29	74.35	81.89	83.83	86.97	92.37
Electronics	67.85	72.55	83.94	85.32	87.25	92.64

Book domain is 27.04%, 20.17%, 12.02%, 08.58% and 06.87% superior to traditional KNN-term similarity, KNN-geometry + term similarity, CCHAN, HANP and GWO + NN models. Similarly, the precision and recall of adopted model is higher than existing models. The F-measure of proposed model for Book domain is 28.30%, 21.63%, 10.81%, 7.88%, and 5.69% higher than traditional models as discussed earlier. Similar higher results are achieved for all other domains as a target domain.

From Fig. 2(f), NPV of the adopted scheme for Book domain is 92.83% whereas, the existing models namely KNN-term similarity, KNN-geometry + term similarity, CCHAN, HANP and GWO + NN attain a comparatively lower NPV values of 66.67%, 72.66%, 83.33%, 85.89% and 88.21% respectively. This indicates that NPV measure is 28.18%, 21.73%, 10.23%, 7.48%, and 4.98% higher than other traditional models. All comparative charts are shown in Fig. 2 and corresponding values of some important measures are also shown in Table 7. Results show that the proposed model of domain adaptation with the enhanced cross entropy and IGWO has gained improvement of 6% to 28% in accuracy and similar improvement is observed in other evaluation measures compare to traditional models of domain adaptation.

9. Conclusion and future work

This paper has developed a new multi-source domain adaptation technique to solve the problem of data annotation that includes enhance cross entropy measure and IGWO technique. In proposed model, pre-processing was the primary step, where stop word removal takes place and the key words were extracted. Following the keyword extraction, frequency matrix and distribution table were constructed. Accordingly, feature extraction was carried out, which exploits enhanced cross entropy measure. These extracted features from target domain enhance the training dataset for classification of other data from target domain. For carrying out classification, IGWO based NN was exploited. More particularly, the weights of NN were tuned in an optimal manner using IGWO algorithm. Finally, simulation was carried out to validate the enhancement of the presented scheme. On observing the outcomes, the presented IGWO based NN model has accomplished better accuracy and f-measure when compared over the existing models for all four domains. Specifically, the presented method has achieved an accuracy of 28% to 6% better than KNN-term similarity, KNN-geometry + term similarity, CCHAN, HANP and GWO + NN models for all four domains. Similarly, F-measure is also 28% to 6% better and that of specificity is 27% to 7% higher than other models.

The model can be extended for applying domain adaptation on multiple targets that is many to many mappings. The combination of IGWO with other optimization based model can also be considered for further studies.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contribution

The paper conceptualization, original draft preparation, validation, deciding methodology and software and also the formal analysis was carried

out by 1st author. The supervision and project administration were done by 2nd author.

References

[1] T. Kincl, M. Novák, and J. Přibil, "Improving sentiment analysis performance on morphologically rich languages: Language and domain independent approach", *Computer Speech & Language*, Vol. 56, pp. 36-51, 2019.

[2] M. Atzeni and D. R. Recupero, "Multi-domain sentiment analysis with mimicked and polarized word embeddings for human-robot interaction", *Future Generation Computer Systems*, Vol. 110, pp. 984-999, 2020.

[3] Z. Yuan, S. Wu, F. Wu, J. Liu, and Y. Huang, "Domain attention model for multi-domain sentiment classification", *Knowledge-Based Systems*, Vol. 155, pp. 1-10, 2018.

[4] F. Wu, Y. Huang, and Z. Yuan, "Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources", *Information Fusion*, Vol. 35, pp. 26-37, 2017.

[5] O. Araque, G. Zhu, and C. A. Iglesias, "A semantic similarity-based perspective of affect lexicons for sentiment analysis", *Knowledge-Based Systems*, Vol. 165, pp. 346-359, 2019.

[6] M. Ghiassi and S. Lee, "A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach", *Expert Systems with Applications*, Vol. 106, pp. 197-216, 2018.

[7] J. Liang, K. Zhang, X. Zhou, Y. Hu, J. Tan, and S. Bai, "Leveraging Latent Sentiment Constraint in Probabilistic Matrix Factorization for Cross-domain Sentiment Classification", *Procedia Computer Science*, Vol. 80, pp. 366-375, 2016.

[8] C. Zhao, S. Wang, and D. Li, "Exploiting social and local contexts propagation for inducing Chinese microblog-specific sentiment lexicons", *Computer Speech & Language*, Vol. 55, pp. 57-81, 2019.

[9] D. Das and C. S. G. Lee, "Sample-to-sample correspondence for unsupervised domain adaptation", *Engineering Applications of Artificial Intelligence*, Vol. 73, pp. 80-91, 2018.

[10] A. Kumar, K. Srinivasan, W. H. Cheng, and A. Y. Zomaya, "Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data", *Information Processing & Management*, Vol. 57, No. 1, pp. 102141-102165, 2020.

[11] M. Araújo, A. Pereira, and F. Benevenuto, "A comparative study of machine translation for multilingual sentence-level sentiment analysis", *Information Sciences*, Vol. 512, pp. 1078-1102, 2020.

[12] Q. Sun, J. Niu, Z. Yao, and H. Yan, "Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level", *Engineering Applications of Artificial Intelligence*, Vol. 81, pp. 68-78, 2019.

[13] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, "Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion", *Information Processing & Management*, Vol. 56, No. 4, pp. 1245-1259, 2019.

[14] G. Zhou, Y. Zhou, X. Guo, X. Tu, and T. He, "Cross-domain sentiment classification via topical correspondence transfer", *Neurocomputing*, Vol. 159, pp. 298-305, 2015.

[15] H. Cho, S. Kim, J. Lee, and J. S. Lee, "Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews", *Knowledge-Based Systems*, Vol. 71, pp. 61-71, 2014.

[16] R. K. Amplayo, S. Lee, and M. Song, "Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis", *Information Sciences*, Vol. 454-455, pp. 200-215, 2018.

[17] A. Balahur and J. M. P. Ortega, "Sentiment analysis system adaptation for multilingual processing: The case of tweets", *Information Processing & Management*, Vol. 51, No. 4, pp. 547-556, 2015.

[18] C. Zhao, S. Wang, and D. Li, "Multi-source domain adaptation with joint learning for cross-domain sentiment classification", *Knowledge-Based Systems*, Vol. 191, pp. 105254-105269, 2020.

[19] M. López, A. Valdivia, E. M. Cámara, M. V. Luzón, and F. Herrera, "E2SAM: EVolutionary ensemble of sentiment analysis methods for domain adaptation", *Information Sciences*, Vol. 480, pp. 273-286, 2019.

[20] F. Z. Xing, F. Pallucchini, and E. Cambria, "Cognitive-inspired domain adaptation of sentiment lexicons", *Information Processing & Management*, Vol. 56, No. 3, pp. 554-564, 2019.

[21] H. Yin, P. Liu, Z. Zhu, W. Li, and Q. Wang, "Capsule Network With Identifying Transferable Knowledge for Cross-Domain Sentiment Classification", *IEEE Access*, Vol. 7, pp. 153171-153182, 2019.

[22] D. Bollegala, T. Mu, and J. Y. Goulermas, "Cross-Domain Sentiment Classification Using Sentiment Sensitive Embeddings", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 2, pp. 398-410, 2016.

[23] T. Manshu and W. Bing, "Adding Prior Knowledge in Hierarchical Attention Neural Network for Cross Domain Sentiment Classification", *IEEE Access*, Vol. 7, pp. 32578-32588, 2019.

[24] T. Manshu and Z. Xuemin, "CCHAN: An End to End Model for Cross Domain Sentiment Classification", *IEEE Access*, Vol. 7, pp. 50232-50239, 2019.

[25] P. Wei, Y. Ke, and C. K. Goh, "A General Domain Specific Feature Transfer Framework for Hybrid Domain Adaptation", *IEEE Trans. Knowl. Data Eng.*, Vol. 31, No. 8, pp. 1440-1451, 2019.

[26] Y. Mohan, S. S. Chee, D. K. P. Xin, and L. P. Foong, "Artificial neural network for classification of depressive and normal in EEG", In: *Proc. of 2016 IEEE EMBS Conf. on Biomedical Engineering and Sciences (IECBES)*, pp. 286-290, 2016.

[27] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer", *Advances in Engineering Software*, Vol. 69, pp. 46-61, 2014.

[28] Z. Wu and T. W. S. Chow, "Neighborhood field for cooperative optimization", *Soft Comput*, Vol. 17, No. 5, pp. 819-834, 2013.