ADAPTIVE WILD GEESE ALGORITHM WITH HIERARCHICAL ATTENTION NETWORK FOR SENTIMENT ANALYSIS USING PRODUCT REVIEW

Dr. S.Selvarani[1] , R. Catherin Ida Shylu[2]

[1]Assistant Professor, Department Of Computer Science, Alagappa Government Arts College, Karaikudi - 630 003 mail id: samyselvaa@gmail.com

[2]Research Scholar Department of Computer and Information Science,Annamalai University, Chidambaram, Tamilnadu, India. Catherinmca.me@gmail.com

Abstract

The elevated accessibility of reviews produced online needs a pertinent solution for drawing the chronological insights from the streaming reviews. The majority of customers depend on the reviews for making decisions to buy products. However, the occurrence of fake reviews is termed as a major problem which is still not addressed by the former techniques. Hence this paper designs a model for sentiment analysis using product reviews. The aim is to develop an Adaptive Wild Geese algorithm (AWGA) algorithm with Hierarchical Attention Networks for sentiment analysis using product review. Here, the Natural language processing (NLP) features along with Term Frequency-Inverse Document Frequency (TF-IDF) are obtained. After feature extraction, the sentiment analysis is executed considering Hierarchical Attention Network (HAN) and trained with AWGA. The AWGA is obtained by combination of Adaptive Concept (AC) and wild geese algorithm (WGA) and the output is considered for sentiment analysis. The proposed AWGA-HAN outperformed with augmented precision of 90.9%, recall of 93.6% and F1-score of 92.2%.

Keywords: *Sentiment analysis, HAN, product review, NLP, Opinion mining.*

1. Introduction

NLP represents a major domain concerning computational and mathematical modeling of several factors of language and the design of huge number of models. The significance of NLP helps in progressing the computer science and telecommunications and can never be underestimated. Sentiment analysis indicates a huge domain in NLP, text mining and computational linguistics [10]. The usage of this method led to extraction of opinion concerning a product. The mining of opinion describes opinion as negative or positive and assessment of sentiment describes value of polarity considering the opinion of customer for that specific product. The present techniques of sentiment assessment are lexicon-based methods, machine learning algorithms, and hybrid models [14]. The discovery of sentiment is imperative for several business institutions. The identification of public views with goods, institution, policies are of huge advantage to decision making models. The Sentiment analysis refers to a process which can be beneficial in detecting the polarity. It is essential to resolve the issues for elevating the efficacy of data mining [13].

Deep learning assisted methods are emerging due to its significant usage in several applications. When analyzing the efficiency of single technique on a single database in a particular domain, then the outcomes suggested that recurrent neural network (RNN) and convolutional neural network (CNN) provided improved accuracy. The majority of the studies exposed that deep models precisely determine the sentiment. The researchers are concentrating on designing stronger techniques for handling big data complications, and elevating sentiment analysis to huge range of applications like marketing strategies and other domains [15]. Sentiment analysis through the classification seems to solve the above problems by mining specific data offered from the provided texts like reviews of consumer. The categorization of online reviews with deep model defines the complete semanticization of user reviews with their classes that include positive and negative sentiments [3].

The aim is to provide optimized HAN model to perform sentiment analysis with product review. Initially, the review of product is considered and from that features are obtained. Here, the obtained features are all-caps, hashtag, emoticon, elongated units, along with bag of units, punctuation and TF-IDF. After the feature extraction, the analysis of sentiment is implemented with HAN and trained with AWGA. Here, the AWGA is obtained by blending adaptive idea in WGA and the output is considered for sentiment analysis.

The noteworthy paper contributions are:

- **Designed AWGA-HAN for sentiment analysis**: The AWGA is adapted for tuning the weights of HAN for analyzing the sentiments. Here, the HAN undergoes training with AWGA, which is developed by inducing Adaptive Concept in WGA.

The remaining sections are arranged as: Section 2 defines literary works of sentiment evaluation. Section 3 offers proposed model for analyzing the sentiment. Section 4 describes the discussion of outcomes and section 5 generates conclusion.

2. Motivations

The majority of schemes are devised to analyze sentiment using product reviews. However, the most of techniques suffered from the ignorance of word rankings, grammar rules and most of these techniques suffered from elevated dimensionality and sparsity of feature vectors. Thus, motive is to devise an optimum model for sentiment analysis using product reviews.

2.1. Literature survey

Alzahrani, M.E.,et al.[1] designed a long short term memory (LSTM) and CNN based model for analyzing the sentiment using product reviews. The method used preprocessing steps for cleaning the data and finally CNN-LSTM was applied for analyzing the sentiment to obtain output as either positive or negative. However, this technique faced the occurrence of fake reviews that led users to choose undesired products. Onan, A. [2] developed deep learning-assisted technique for analyzing the sentiment based on product reviews. The method used TF-IDF weighted Glove word embedding with CNN-LSTM for analyzing sentiments. However, this method suffered from overfitting issue due to huge document length. Iqbal, A.,et al.[3] utilized LSTM for analyzing the sentiment using product reviews. The method used preprocessing and classification steps to interpret the data. However, the method did not use suitable encoding of data to improve system performance. AL-Sharuee, M.T.,et al. [4] devises automatic contextual analysis and ensemble clustering (ACAEC) algorithm for analyzing the sentiment using product review. The method used contextual analysis and ensemble learning for attaining enhanced performance. It also utilized additional weight scheme to enhance the window sequential clustering outcome. However, the method produced low accuracy rate.

2.1. Challenges

The drawbacks tackled by sentiment analysis schemes considering the product review are stated below,

- In [3], LSTM is devised for analyzing the sentiments using product reviews. The method used pre-processing to examine specific data type. However, it can be complex for classifier to generate precise findings on provided data if preprocessing step is executed appropriately.

- In [4], two machine learning techniques are developed to examine sentiment using product review. However, this technique did not adapt automatic window setting to make a decision regarding the window length using data density.
- The issue of sentiment assessment using product reviews of Chinese e-commerce relies in mapping dimension, disambiguation amid sentiment words.

3. Designed AWGA-based HAN for sentiment evaluation using product review

The analysis of sentiment is a main process in processing the natural language wherein the thoughts, attitudes, and opinions over a specific subject is obtained. Web indicates a rich information source through which several text data with opinions are obtained. Hence, the discovery of sentiment is useful for making decision. The aim is to develop an AWGA algorithm with HAN for sentiment analysis using product review. For that, initially, the product review is taken as the input and then the features are extracted. Here, the features extracted [12] includes all-caps, emoticon, hashtag, elongated units, bag of units, punctuation and TF-IDF. After the feature extraction, the sentiment analysis is done using HAN [9], which is trained by AWGA. The proposed AWGA is obtained by the combination of Adaptive Concept and WGA [16] and the required output are noted. Figure 1 reveals the outlook of sentiment analysis model using AWGA-HAN with product reviews.

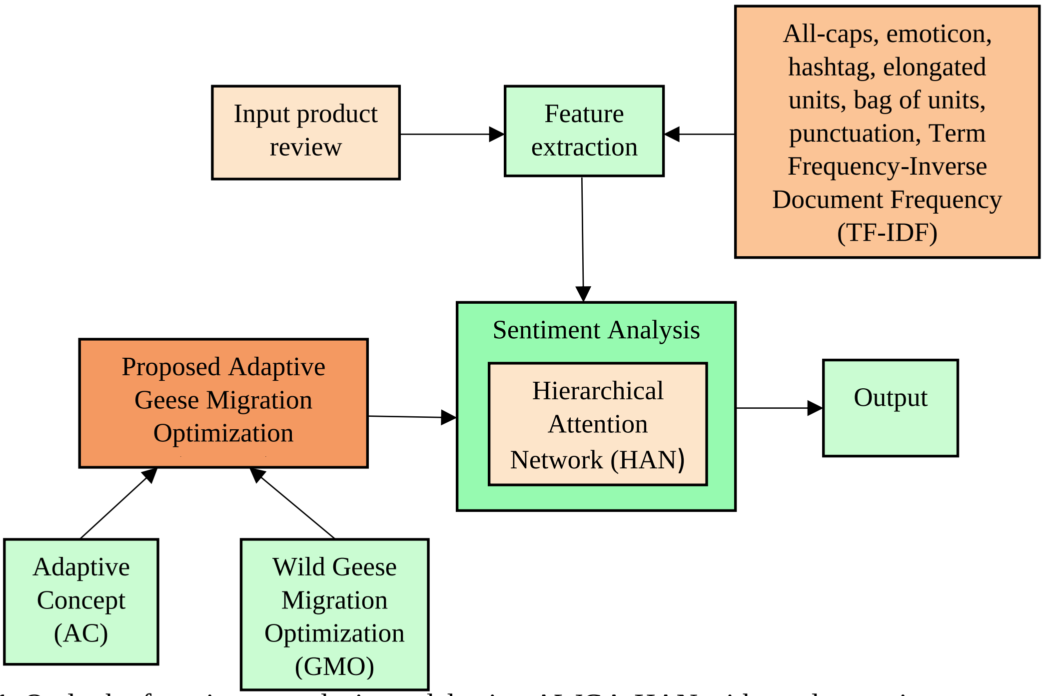


Figure 1. Outlook of sentiment analysis model using AWGA-HAN with product review

3.1. Collect review data

Assume a product review database F with total count of samples and is notified as,

$$F = \{F_1, F_2, \ldots, F_o, \ldots, F_n\} \quad (1)$$

Hence, n depicts total product reviews, and F_o specifies o^{th} product review.

3.2. Extraction of features

Here, the product review F_o is provided to feature mining stage in which mining of specific features is done.

a) All caps

It indicates count of words that poses all characters in terms of upper case and it is notated as C_1.

b) Emoticon

It aids to express feelings of person and can be positive or negative and is stated by,

$$C_2 = \sum_{i=1}^{j} \alpha_q^i \qquad (2)$$

Hence, the entire emoticon number of i^{th} data is notated as α_q^i and is assigned with value "1" if emotion exists and "0" else. It is articulated as C_2.

c) Hashtag

It is used to draw attentiveness, and promotion. The hashtag occurrence considering a data is stated as,

$$C_3 = \sum_{i=1}^{j} \alpha_\ell^i \qquad (3)$$

Hence, α_ℓ^i depicts number of hashtag in i^{th} data. It is signified as C_3.

d) Elongated units

It expresses character which is repeating twice in a product review data. Thus elongated words count is stated as,

$$C_4 = \sum_{t=1}^{u} \chi_t^\nu \qquad (4)$$

Hence, χ_t^ν states total hash tags in ν^{th} data where 0 depicts presence of elongated words, and 1 is non-existence of elongated words and signified by C_4.

e) Bag of units

It indicates texts modeling that describe the occurrence of words from a data. Each word contains a data which is provided using score that adapts Bag-of-words method and is notated as C_5.

f) Punctuations

The punctuation marks C_6 are considered as exclamation mark, apostrophe, or dot present in a review.

$$C_6 = \sum_{c=1}^{d} \varepsilon_l^c \qquad (5)$$

Here, ε_l^c depicts complete punctuations present in c^{th} review. Thus, ε_l is offered 1 considering punctuations that happened in a review, else 0.

g) TF-IDF

The TF-IDF [5] expresses an arithmetic signal that indicates how important a word is to its data in a set of data. Here, the TF is stated as,

$$R_b = \frac{e}{f} \tag{6}$$

Hence, e reveals total word appearance contained in a data and f depicts total words contained in a data.

The IDF is notated by,

$$A_b = \log \frac{g}{a} \tag{7}$$

Hence, g indicates total data and a depicts data frequency and TF-IDF is stated by,

$$C_7 = R_b \times A_b \tag{8}$$

The outcome produced with TF-IDF is provided by C_7. The feature vector modelled is provided by,

$$C = \{C_1, C_2, \cdots, C_7\} \tag{9}$$

3.3. Sentiment analysis using AWGA-HAN

The AWGA-HAN is employed to analyze sentiments using reviews with a feature vector C. Thus, the AWGA-HAN is generated by unifying AWGA in HAN [9]. The HAN outlook and its training using AWGA are provided below.

a) HAN outlook

HAN [9] focus on data to build memories and concentrate on specific process. It is reliable and simple for usage. The network contains the potential for handing several types of data and concentrates on precious components. Consider a review contains G sentences H_r and each sentence comprises N_r words, O_{rw} with $N \in [1, N]$ indicating words in r^{th} sentence.

(i)Word Encoder:

Provided a sentence using words $O_{r,w}, w \in [0, N]$, initially words embedding towards vectors is carried out using embedding matrix $\varpi_o, \rho_{r,\kappa} = \varpi_o O_{r,\kappa}$. The bidirectional GRU contains forward GRU $\overline{\eta}$ which reads sentence H_r using O_{r1} to O_{rN} and backward GRU $\bar{\eta}$ that read with O_{rN} to O_{r1} :

$$\rho_{rw} = \varpi_o O_{rw}, w \in [1, N], \tag{10}$$

$$\bar{g}_{rw} = \overrightarrow{GRU}(\rho_{rw}), w \in [1, N], \tag{11}$$

$$\vec{g}_{rw} = \overleftarrow{GRU}(\rho_{rw}), w \in [N,1], \qquad (12)$$

This unit generated word annotation for a provided word O_{rw} by unifying forward hidden state $\vec{g}_{rw}$ and backward hidden state $\bar{g}_{rw}$ in which $g_{rw} = [\vec{g}_{rw}, \bar{g}_{rw}]$ that summarize data through whole sentence centered amid g_{rw}.

(ii)Word Attention:

It is used to mine words which are essential to define a sentence and accumulate design of words to form sentence vector.

$$Q_{rw} = \tanh(\varpi_O g_{rw} + \mu_O) \qquad (13)$$

$$\beta_{rw} = \frac{\exp(Q_{rw} u_w)}{\sum_w \exp(u_{rw} u_w)} \qquad (14)$$

$$H_r = \sum_w \beta_{r,w} g_{r,w} \qquad (15)$$

Thus, the word annotation $g_{r,w}$ is fed into one-layer MLP to acquire $Q_{r,w}$ as hidden depiction of $g_{r,w}$, and compute significance of word $Q_{r,w}$ using word level context vector Q_{ϖ} and attain normalized implication weight β_{rw} using softmax function. Thus, the word context vector Q_{ϖ} is arbitrarily initialized and learned jointly in the process of training.

(iii) Sentence Encoder

Given sentence vector H_r, a review vector is generated wherein bidirectional GRU is employed for encoding sentences and is stated by,

$$\vec{g}_r = \overrightarrow{GRU}(H_r), r \in [1, G], \qquad (16)$$

$$\bar{g}_r = \overleftarrow{GRU}(H_r), r \in [G,1], \qquad (17)$$

The unification of $\vec{g}_r$ and $\bar{g}_r$ to generate sentence annotation r that is $g_r = [\vec{g}_r, \bar{g}_r]$, and g_r exposes neighbour sentence around sentence r and focuses on sentence r.

(iv) Sentence Attention:

The sentence level context vector Q_s is stated by,

$$Q_r = \tanh(\varpi_s g_r + \mu_s), \qquad (18)$$

Here, $\vec{g}_r$ depicts annotation of sentence, ϖ_s denote sentence attention matrix, and μ_s indicate bias.

$$\beta_r = \frac{\exp(Q_r^T, Q_s)}{\sum_i \exp(Q_r^T, Q_s)}, \qquad (19)$$

Hence, Q_s attained normalized significance weight and Q_r^T depicts transpose of hidden depiction.

$$y = \sum_r \beta_r g_r, \qquad (20)$$

Hence, y depicts review vector.

Thus, review vector v express high level depiction of review and is depicted by

$$\vartheta = soft\max(\varpi_c y + \mu_c) \qquad (21)$$

The negative log likelihood is used considering right labels, and is depicted as,

$$G = -\sum_{\partial} \log \vartheta_{\partial\kappa} \qquad (22)$$

Here, κ depicts label considering review ∂. The output produced through HAN is denoted by N_{out} and is known for classification of sentiment.

b) Train HAN with AWGA

WGA [16] is motivated form the wild geese discovered in the nature. This algorithm is utilized for optimizing high-dimensional problems. It is easy and fundamental technique for dealing with huge-scale optimization and are utilized for several addressing real-world optimization issues. The adaptive idea is incorporated to make the processing of tasks separately. Adaptive technique alters setting of control attribute and incorporates to each genome. Thus, adaptive method is unified with WGA for acquiring improved efficiency. The AWGA steps are defined as,

Step 1) Initialization

The first step is to create the wild geese population and hence the position vector of s^{th} wild goose is similar to y_s. The optimum position h_s and migration velocity k_s are discovered.

Step 2) Find error

The fitness is modelled as,

$$\Im_{err} = \frac{1}{n}\sum_{o=1}^{n}\left[\mathrm{N}_o^* - \mathrm{N}_{out}\right]^2 \qquad (23)$$

Thus, N articulates total instances acquired, N_{out} depicts generated HAN output, and N_o^* stated predicted outcome.

Step3) Derive Ordered migration of group

Wild geese migration indicate a set which tends to be ordered and coordinated based on attaining upfront and neighboring individuals in a grouped population. The displacement and velocity expressions based on the coordinated geese velocity provided as,

$$k_{s,p}^{m+1} = \left(z_{1,p} \times k_{s,p}^{m} + z_{2,p} \times \left(k_{s+1,p}^{m} - k_{s-1,p}^{m}\right)\right) + z_{3,p} \times \left(h_{s,p}^{m} - y_{s-1,p}^{m}\right) + z_{4,p} \times \left(h_{s+1,p}^{m} - y_{s,p}^{m}\right)$$
$$+ z_{5,p} \times \left(h_{s+2,p}^{m} - y_{s+1,p}^{m}\right) - z_{6,p} \times \left(h_{s-1,p}^{m} - y_{s+2,p}^{m}\right)$$
(24)

where, $y_{s,p}, h_{s,p}$ and $k_{s,p}$ denote p^{th} dimension of present position, best position and present velocity of p^{th} wild geese and $z_{k,p}$ is uniformly distributed random numbers.

Here, the parameter $z_{k,p}$ is made adaptive which is provided as,

$$z_{k,p} = 1 - \left(\frac{m}{m_{\max}}\right) * \mathrm{M}_o \quad (25)$$

The change in position is implemented in an ordered format and synchronized using the members for modeling the member movement as an arranged sequence and is provided as,

$$y_{s,p}^{k} = h_{s,p}^{k} + z_{7,p} \times z_{8,p} \times \left(\left(x_{p}^{m} + h_{s+1,p}^{m} - 2 \times h_{s,p}^{m}\right) + k_{s,p}^{m+1}\right) \quad (26)$$

where, x_p indicate global best location amidst all members.

Step 4) Derive Walking and search for food by wild geese

It is observed that s^{th} wild goose transforms to its upfront member. Hence, the expression of walking and wild goose searching is modelled as,

$$y_{s,p}^{\tau} = h_{s,p}^{m} + z_{9,p} \times z_{10,p} \times \left(h_{s+1,p}^{m} - h_{s,p}^{m}\right) \quad (27)$$

Step 5) Derive Reproduction and evolution of wild geese

Other wild geese stage is evolution and reproduction. Here, the unification amid migration equation and walking and search of food is utilized. The equation is given as,

$$y_{s,p}^{m+1} = \begin{cases} y_{s,p}^{k} & ; If \ z_{11,p} \le \vartheta \\ y_{s,p}^{\tau} ; Otherwise \end{cases} \quad (28)$$

Step 6) Derive Death, migration and ordered evolution

It is adapted to balance efficiency for all test functions and it is formulated as,

$$\Re = round\left(\begin{array}{l}\Re^{initial} \\ -\left(\left(\Re^{initial} - \Re^{final}\right) * \left(\frac{\Im}{\Im_{\max}}\right)\right)\end{array}\right) \tag{29}$$

where, $\Im$ and $\Im_{\max}$ depicts functional evaluations cost and its maximal value.

Step 7) Re-caluclate error to find better solution

The solution is re-calculated with error and solution having optimum value of fitness is known as better solution.

Step 8) Termination

The steps are repeated until highest iteration count is accomplished.

4. Results and Discussion

Potential of AWGA-based HAN is examined using various kinds of evaluation criterions.

4.1. Experimental set-up

AWGA-based HAN is scripted in Python.

4.2. Dataset description

The analysis of techniques and algorithms is executed with Amazon reviews for sentiment analysis dataset [11]. This dataset contains Amazon NLP for performing sentiment analysis. Moreover, it provides ratings of ranking product and reviews on Amazon. In addition, it has Amazon product data which involves product classes and several metadata.

4.3. Evaluation measures

Aptitude computation graphs of AWGA-based HAN are produced by comparing proposed with previous methods and is attained by employing definite metrics that are illustrated below.

a) Precision

It measures count of positive class prediction that fundamentally fit into a positive class and is notified by,

$$\psi = \frac{\zeta_\mu}{\zeta_\mu + \gamma_\mu} \tag{30}$$

Hence, ζ_μ delineate true positive, γ_μ depicts false positive.

b) Recall

It offers count of positive class prediction made with positive samples in a database and is stated as,

$$\delta = \frac{\zeta_{\mu}}{\zeta_{\mu} + \gamma_{\eta}} \tag{31}$$

Hence, γ_{η} states false negative.

c) F1-score

It gives a specific score which make better balances amid the issues of recall and precision in specific number and is stated as,

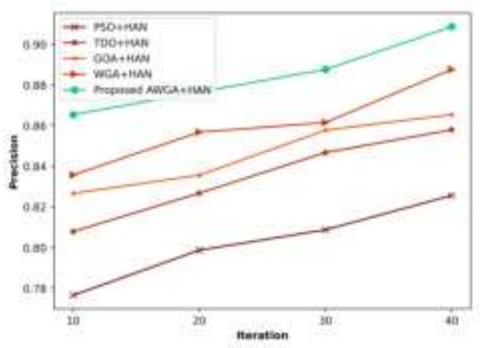
$$M = 2 * \frac{\psi * \delta}{\psi + \delta} \tag{32}$$

Thus, ψ and δ states precision and recall.

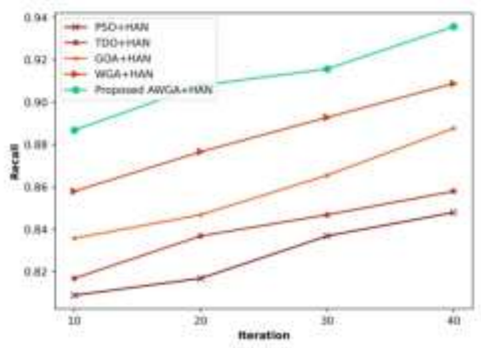
4.4. Algorithmic methods

The algorithm included for the purpose of efficiency computation is PSO+HAN [6][9], TDO+HAN [7][9], GOA+HAN [8][9], WGA+HAN [16][9] and AWGA+HAN.

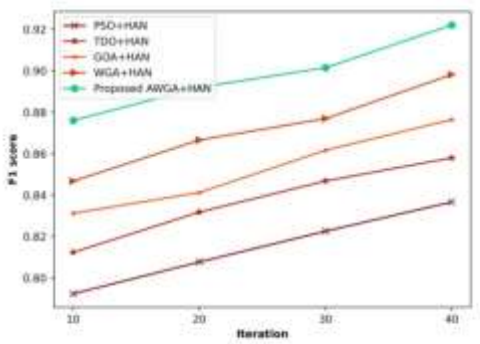
4.5. Algorithm estimation



(a)



(b)



(c)

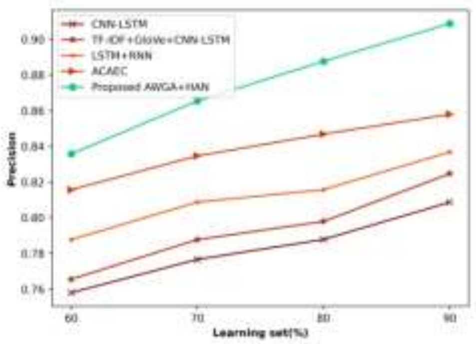
Figure 2. Algorithm efficiency estimate using a) Precision b) Recall c) F1-score

Figure 2 deliberates estimation of algorithm efficiency. The evaluation of efficiency based on precision is delineated in figure 2a). When iteration tends to be 40, the corresponding precision measured by PSO+HAN is 0.826, TDO+HAN is 0.858, GOA+HAN is 0.865, WGA+HAN is 0.888, and AWGA+HAN is 0.909. The graphical analysis considering recall is stated in figure 2b). When iteration =40, the elevated recall of 0.936 is calculated by AWGA+HAN while recall of PSO+HAN, TDO+HAN, GOA+HAN, WGA+HAN are 0.848, 0.858, 0.888, 0.909. The F1-score evaluation graph is articulated in figure 2c). Considering iteration as 40, the equivalent F1-score calculated is 0.837 for PSO+HAN, 0.858 for TDO+HAN, 0.876 for GOA+HAN, 0.898 for WGA+HAN and 0.922 for AWGA+HAN.

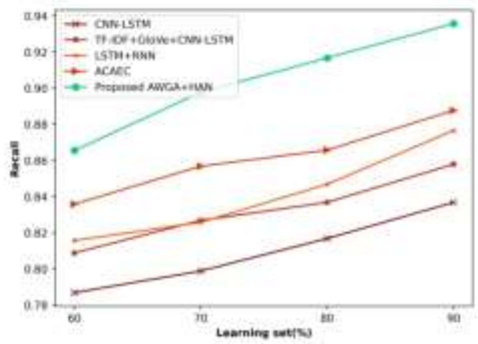
4.6. Comparative methods

The techniques involved with the aim of efficiency evaluation are CNN-LSTM [1], TF-IDF+GloVe+CNN-LSTM [2], LSTM+RNN [3], ACAEC [4], and AWGA+HAN.

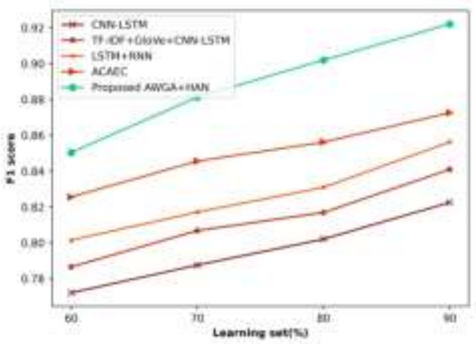
4.7. Comparative estimation



(a)



(b)



(c)

Figure 3. Technique efficiency estimate by means of a) Precision b) Recall c) F-measure

Figure 3 displays method efficiency evaluation graphs. The precision efficiency graph is stated in figure 3a). When learning set depicts 90%, the corresponding precision calculated by CNN-LSTM is 0.809, TF-IDF+GloVe+CNN-LSTM is 0.825, LSTM+RNN is 0.837, ACAEC is 0.858, and AWGA+HAN is 0.909. The recall evaluation graph is articulated in figure 3b). Assuming learning set = 90%, the augmented recall of 0.936 is noted by AWGA+HAN whereas recall of CNN-LSTM, TF-IDF+GloVe+CNN-LSTM, LSTM+RNN, ACAEC are 0.837, 0.858, 0.877, 0.888. The evaluation of efficiency with F-measure is delineated in figure 3c). Utilizing learning set as 90%, the F-measure obtained is 0.822 for CNN-LSTM, 0.841 for TF-IDF+GloVe+CNN-LSTM, 0.856 for LSTM+RNN, 0.873 for ACAEC, and 0.922 for AWGA+HAN.

5. Conclusion

The analysis of sentiments indicates a huge area in NLP, text mining and computational linguistics. The usage of this method led to mining and assessment of opinion using provided product. The extraction of opinion describes an opinion into negative and positive and sentiment analysis describes the value of polarity to obtain customer opinion on certain product. The aim is to develop an AWGA algorithm with HAN for sentiment analysis using product review. Here, the product review is considered as input and then the features are extracted. Here, the extracted features include NLP-assisted features and TF-IDF. After feature mining, the evaluation of sentiment is implemented with HAN and trained using AWGA. The AWGA is produced by the unifying adaptive idea and WGA and the output is considered for sentiment analysis. The AWGA-HAN gives augmented efficacy depicting precision of 90.9%, recall of 93.6% and F1-score of 92.2%.

References

[1]Alzahrani, M.E., Aldhyani, T.H., Alsubari, S.N., Althobaiti, M.M. and Fahad, A., “Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews”, Computational Intelligence and Neuroscience, vol.2022, 2022.

[2]Onan, A., “Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks”, Concurrency and Computation: Practice and Experience, vol.33, no.23, pp.e5909, 2021.

[3]Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A. and Hussain, M., “Sentiment Analysis of Consumer Reviews Using Deep Learning”,Sustainability, vol.14, no.17, pp.10844, 2022.

[4]AL-Sharuee, M.T., Liu, F. and Pratama, M., “Sentiment analysis: dynamic and temporal clustering of product reviews”, Applied Intelligence, vol.51, pp.51-70, 2021.

[5] Christian, H., Agus, M.P. and Suhartono, D, "Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF)," ComTech: Computer, Mathematics and Engineering Applications, vol.7, no.4, pp.285-294, 2016.

[6] Kennedy, J. and Eberhart, R, "Particle swarm optimization," In Proceedings of ICNN'95-international conference on neural networks, vol.4, pp. 1942-1948, 1995.

[7] Dehghani, M., Hubálovský, Š. and Trojovský, P, "Tasmanian devil optimization: a new bio-inspired optimization algorithm for solving optimization algorithm," IEEE Access, vol.10, pp.19599-19620, 2022.

[8] Pan, J.S., Zhang, L.G., Wang, R.B., Snášel, V. and Chu, S.C, "Gannet optimization algorithm: A new metaheuristic algorithm for solving engineering optimization problems," Mathematics and Computers in Simulation, vol.202, pp.343-373, 2022.

[9] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. and Hovy, E., “Hierarchical attention networks for document classification”, In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp. 1480-1489, June 2016.

[10]Jagdale, O., Harmalkar, V., Chavan, S. and Sharma, N., “Twitter mining using R”, Int. J. Eng. Res. Adv. Tech, vol.3, pp.252-256, 2017.

[11] The Amazon Reviews for sentiment analysis data set will be taken from, https://www.kaggle.com/datasets/tarkkaanko/amazonaccessed on March 2023.

[12]Tang, D., Qin, B., Wei, F., Dong, L., Liu, T. and Zhou, M., “A joint segmentation and classification framework for sentence level sentiment classification”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.23, no.11, pp.1750-1761, 2015.

[13]Akhtar, M.J., Ahmad, Z., Amin, R., Almotiri, S.H., Al Ghamdi, M.A. and Aldabbas, H., “An efficient mechanism for product data extraction from e-commerce websites”, Comput. Mater. Contin, vol.65, no.3, pp.2639-2663, 2020.

[14]Prabowo, R. and Thelwall, M., “Sentiment analysis: A combined approach”, Journal ofInformetrics, vol.3, no.2, pp.143-157, 2009.

[15]Roy, S.S., Biba, M., Kumar, R., Kumar, R. and Samui, P., “A new SVM method for recognizing polarity of sentiments in twitter”, In proceedings of Handbook of Research on Soft Computing and Nature-Inspired Algorithms, IGI Global pp. 281-291, 2017.

[16] Ghasemi, M., Rahimnejad, A., Hemmati, R., Akbari, E. and Gadsden, S.A, "Wild Geese Algorithm: A novel algorithm for large scale optimization based on the natural life and death of wild geese," Array, vol.11, pp.100074, 2021.