Asia Pacific Journal of Information Systems
Vol. 32 No. 2 (June 2022), 226-248

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2022.32.2.226

Aspect-based Sentiment Analysis of Product Reviews using Multi-agent Deep Reinforcement Learning

M. Sivakumar[a], Srinivasulu Reddy Uyyala[b,*]

[a] *Assistant Professor, Epartment of computer science and engineering, K. Ramakrishnan College of Technology, India*
[b] *Assistant Professor, National Institute of Technology (NIT), Tiruchirappalli, India.*

ABSTRACT

The existing model for sentiment analysis of product reviews learned from past data and new data was labeled based on training. But new data was never used by the existing system for making a decision. The proposed Aspect-based multi-agent Deep Reinforcement learning Sentiment Analysis (ADRSA) model learned from its very first data without the help of any training dataset and labeled a sentence with aspect category and sentiment polarity. It keeps on learning from the new data and updates its knowledge for improving its intelligence. The decision of the proposed system changed over time based on the new data. So, the accuracy of the sentiment analysis using deep reinforcement learning was improved over supervised learning and unsupervised learning methods. Hence, the sentiments of premium customers on a particular site can be explored to other customers effectively. A dynamic environment with a strong knowledge base can help the system to remember the sentences and usage State Action Reward State Action (SARSA) algorithm with Bidirectional Encoder Representations from Transformers (BERT) model improved the performance of the proposed system in terms of accuracy when compared to the state of art methods.

Keywords: Sentiment Analysis, Deep Reinforcement Learning, Product Review, Artificial Intelligence, Machine Learning

Ⅰ. Introduction

Sentiment analysis is one of the tasks of machine learning systems that take text, videos, images, and voice data for detecting sentiments (positive, negative, or neutral) expressed in those data. Sentiment analysis can be done at a document level, sentence level, and word level. Generally, the sentiment analysis task gives the overall polarity of the sentence, paragraph, and document. But detecting sentiments expressed in each

*Corresponding Author. E-mail: usreddy@nitt.edu

sentence with respect (for example, 'display' of a camera) to a particular aspect which is called aspect-based sentiment analysis was a challenging task. In the digital data era, enormous amounts of text data as opinions with respect to various domains such as healthcare, retail, service, and education are populated over social media by human beings every second. To create an aspect-based sentiment analysis model, three different flavors of machine learning algorithms such as supervised, unsupervised, and reinforcement learning can be used. Aspect-based sentiment analysis now-a-days is widely performed by creating a model using supervised learning approaches and training the model with past data, whereas the Reinforcement Learning (RL) methods are rarely used for creating aspect-based sentiment analysis models.

The supervised and unsupervised learning methods are struggling to work with a dynamic unknown environment. Hence, the research focus is now on reinforcement learning for handling dynamic data to perform various Natural Language Processing (NLP) tasks such as text analytics. Instead of making models and giving more training with massive past datasets, making RL-based models that learn on their own to do NLP tasks would be a better alternative to the supervised learning models. In RL-based models, the agents mimic the human beings by observing the environment like humans through returns (or rewards) received from the environment and changing its state by an action based on the rewards. So, the RL approaches handle unknown environments effectively for the learning process. Constructing an aspect-based sentiment analysis with reinforcement learning has a lot of challenges such as designing rewards, configuring actions, managing time, and performing multiple tasks in parallel. But, Deep Reinforcement Learning (DRL) (Mnih et al., 2013) overcomes these challenges by performing all aforementioned tasks on its own. DRL approaches (Chu et al., 2020) were good at performing multiple tasks in parallel using multiple agents to detect aspect-wise sentiments, since detecting polarity, negation, and aspect need to be performed in parallel.

DRL is a hierarchical learning algorithm that combines the power of deep learning, reinforcement learning. DRL have been widely used in various fields such as gaming (Goldwaser and Thielscher, 2020), healthcare (Jonsson, 2019), retail (Boute et al., 2021), smart grid (Yu et al., 2020), NLP (Wang et al., 2018) and robotics (Liu et al., 2021) that are sequential decision making, time-related, and accept delayed feedback.

An intelligent bot was created to suggest products to the consumer by tracking their past online behavior (Shahmanzari and Ozkan, 2014). The proposed bots also increased retailer sales. The proposed model was constraint-based and explicit rules dependent. The intelligent bots generated recommendations of interesting products to the consumers. Hence, designing such intelligent bots using DRL agents that process customer live review text will track the online behavior of consumers. The proposed model suggests interesting products which are based on the policy adopted by DRL.

Bank chat bot was created by combining NLP, classification algorithms and vectorization for serving the customer (Kulkarni et al., 2017). It removed human efforts and gave 24/7 customer support for increasing productivity. Bag-Of-Words (BOW) method was used to convert text data into vector format. Using predefined possible set of questions with answers, the system responded to all possible set of customer questions. Since, the system was developed using classification algorithms, the system will be able to respond only to the available questions. DRL approach will improve the system to understand all possible live questions raised by the customers.

Social bots are software programs that were used in election campaigns by creating personalized preferences without internet users' interaction (Keller et al., 2019). Social bots were used in various online platforms such as Twitter and GitHub for generating and analyzing the contents in the platforms. Creating these social bots using DRL algorithms will help the system to read and analyze contents from the social media effectively.

The usage of DRL in the field of NLP was limited due to the difficulties in understanding text without training. Also, aspect-based sentiment analysis on customer reviews was not performed using DRL. The performance of the existing aspect-based sentiment analysis using RL algorithm (Chen et al., 2019) was not satisfactory in terms of accuracy. The results of the aspect-based sentiment analysis model developed by the DRL algorithm using live datasets were better than the aspect-based sentiment analysis system developed using supervised learning algorithms.

Based on the observations from the previous works using reinforcement learning and DRL, it was found that the deep reinforcement learning algorithm can be used for aspect-based sentiment analysis of live review text data and generate better results in terms of accuracy. Also, it was found that the deep reinforcement learning algorithm was not used yet for aspect-based sentiment analysis. So, a new framework named Aspect-based multi-agent Deep Reinforcement learning Sentiment Analysis (ADRSA) was proposed for detecting sentiments based on particular aspects expressed in the customer product reviews in the online shopping portal. The proposed system added [aspect, sentiment] label pairs on each input review sentence by three agents simultaneously in which one agent detected aspect and another two detected sentiment polarity. In the proposed framework, the policy network was designed using State Action Reward State Action (SARSA) algorithm for observing the environment and the Bidirectional Encoder Representations from Transformers (BERT) model for detecting aspect-wise sentiments.

The RL agents read a sentence word by word from the environment and each word should be verified with the help of its inbuilt sentiment word dictionary and gets a reward or return from the environment. Based on the reward, it then decides whether the sentence can be considered for further process or not. Finally, it labels the given input review sentence with aspect and polarity. So, the proposed framework for aspect-based sentiment analysis of customer reviews uses deep reinforcement learning algorithm with three agents to interact with the environment in parallel for detecting sentiment word, negation term, and aspect term in parallel and assigned two labels such as sentiment polarity and aspect category to each input review sentence. The proposed framework was the first of its kind for aspect-based sentiment analysis of product reviews using deep reinforcement learning that processes the dynamic datasets without the help of any training datasets. The proposed model can perform aspect-based sentiment analysis on customer reviews of all sectors like education, healthcare, agriculture, and hotel reviews effectively. The contributions of the proposed aspect-based sentiment analysis of product reviews using multi-agent deep reinforcement learning are framework as follows:

- SARSA algorithm for constructing the policy network of the RL agent to decide the action
- BERT model for aspect word, negation word and sentiment word detection by DRL agent
- Environment setup based using the dynamic customer text review dataset
- Multiple DRL agents to perform aspect detection, negation, and sentiment word detection

by interacting with the environment
- Knowledge building as an index of word corpus for aspects, negation and sentiment words

The remaining part of the paper was organized in the following order. In section II, similar works carried out using RL were described and in section III various deep reinforcement learning strategies were elaborated. In section IV, the proposed system methodology was presented and in section V, the experimental results were discussed.

Ⅱ. Related Works

The contributions of reinforcement learning algorithms in the sentiment analysis task were not very impressive. Recently, the need for reinforcement learning algorithms is gaining importance on sentiment analysis tasks in various domains that need to handle live datasets.

2.1. Sentiment Analysis Using RL Algorithms

The RL algorithm was used to develop a system that mimics human cognition for identifying sentiments expressed on the given text sentences (Yang et al., 2019). Using RL policy grade procedure for acquiring the representation of the context was presented in the input review text sentences. But this system failed to extract long-distance context words from the input sentences. A system for Chinese sentiment analysis was developed by learning phonetic information from both pinyin token corpus and audio clips using the RL algorithm (Peng et al., 2021). Though RL was used for sentiment analysis, the proposed system did not perform aspect-wise sentiment analysis. A novel framework for natural language was presented to detect interpreting negation using the RL algorithm without manual labeling for each word (Pröllochs et al., 2020). The proposed system experimented on movie reviews and attained 50.20% accuracy on detecting negation terms. The proposed model performed negation detection only at the document level. A description-based text classification method using a RL algorithm was proposed for generating descriptions automatically (Chai et al., 2020). The proposed work required explicit guidelines for what to classify which was a time-consuming task. A model using RL algorithm was developed for generating sentiment-based reward from state representation (Deshpande and Fleisig, 2020). The proposed model used sentiments of text descriptions for generating rewards and did not concentrate on the aspects. A novel framework that was designed using the RL and Long Term Short Memory (LSTM) was built for word-level sentiment analysis (Chen et al., 2019). In the proposed framework, the sentiment polarities such as positive, neutral, and negative were considered as actions to be performed by the agent. The proposed model detects sentiment from each review sentence, but the aspect detection on each sentence was not performed.

2.2. Sentiment Analysis Using DRL Algorithms

A framework using a hierarchical RL method was developed for document-level aspect-based sentiment classification (Wang et al., 2019). The proposed model experimented on the trip user, trip advisor, and BeerAdvocate public datasets with the accuracy of 62.97%, 52.71%, and 43.41% respectively. The proposed system did not concentrate on the negation detection and aspect-opinion co-extraction. A framework for aspect-based sentiment analysis using hierarchical RL was presented with three-level subtasks such as sentiment polarity extraction, opinion, and aspect term extraction (Yu et al., 2021). The proposed framework performed the subtask as a sequential process that was a time-consuming task.

2.3. Sentiment Analysis Using Hierarchical RL Algorithms

Hierarchical RL algorithms are based on hierarchies of policies. A novel approach with a clause-level structure using DRL was developed for sentiment analysis (Zhang et al., 2019). In the proposed approach, feedback of sentiment classifier and trial-and-error-search method was used to discover the clause-level structures in the utterance in the multi-model fashion. They structured the system using LSTM for training the policy and performed sentiment classification with positive and negative categories. The proposed model was compared with the state-of-art methods such as Deep Averaging Network (DAN), Hierarchically Structured LSTM (HS-LSTM), Temporally Selective Attention Model (TSAM), and Hybrid Deep Multimodal Structure (H-DMS). The accuracy of the proposed model was not satisfactory with the benchmark datasets. A framework based on DRL and sentiment analysis was presented for stock market trading (Zhang, 2019). Since the proposed framework used a proximal policy optimization algorithm as RL algorithm for policy network and Convolutional Neural Network (CNN) model for sentiment classification, the framework required more training. A model using DRL was constructed for portfolio allocation by utilizing market sentiment dynamically from the real world (Koratamaddi et al., 2021). Since the model was developed using a single agent, it was able to address only a single domain.

2.4. Comparative Analysis of Related Works

Based on the comparative analysis of related works as presented in <Table 1>, the aspect-based sentiment analysis was performed on Laptop, restaurant, Twitter, Trip user, Trip advisor and BeerAdvocate datasets with the highest accuracy of 83.50%. But the proposed model achieved better accuracy of 86.62%, 97.94% and 97.80% on Movie Review (MR), Amazon Mobile (AM) review and Amazon Food (AF) review datasets respectively. Various models (Wang et al., 2019; Yang et al., 2019) were developed for aspect based sentiment analysis using reinforcement learning with policy gradient and delayed reward function, but they failed to process long distance sentences and negation sentences. The proposed model overcomes these issues by having multiple agents that interacted with the environment in parallel and implementing DRL agents using the BERT model.

Ⅲ. System Methodology

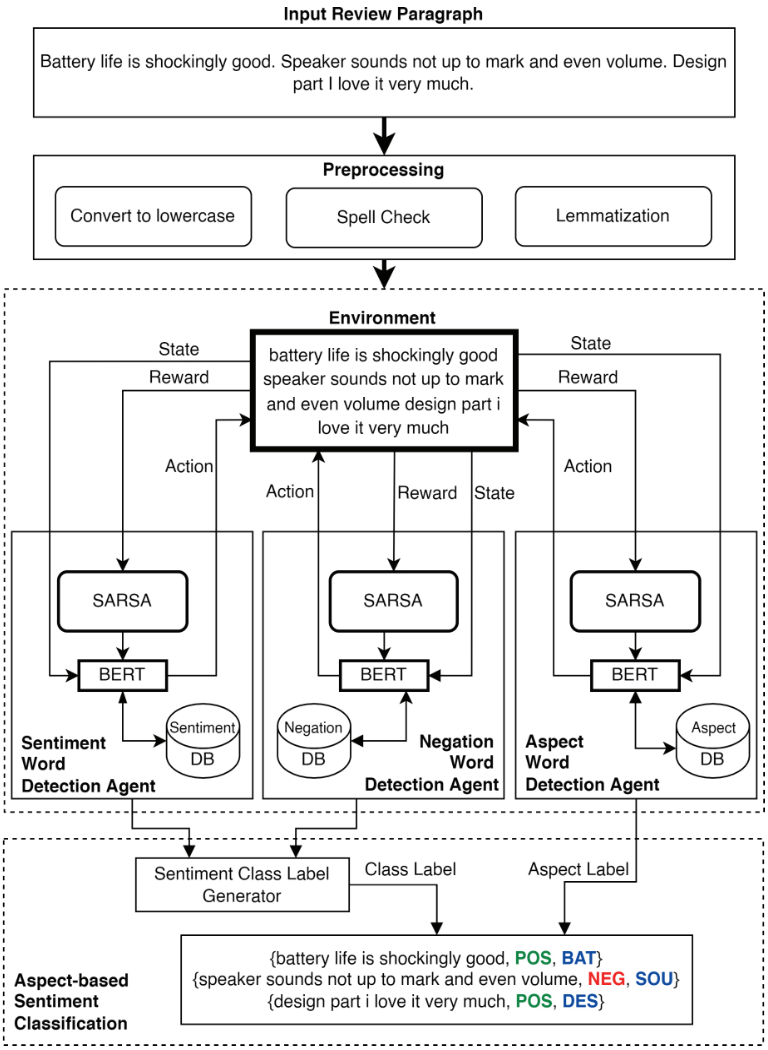
The proposed ADRSA framework, shown in <Figure 1>, was made up of an environment and various agents such as aspect detection, sentiment word detection, negation detection, and class label generator. Each agent of the proposed model generates output as action based on state value and gets delayed reward from the environment. Sentiment word detector, Negation word detector, and Aspect word detector agents process the input product review sentence from the environment simultaneously. The output of the Sentiment Word Detector agent and Negation Word Detector agent were given to the Class Label Generator that performed the XOR operation to detect the correct polarity of the sentiment word. Finally, the given input sentence was labeled with aspect and polarity. Based on this value, the state of the input review sentence was updated.

<Table 1> Comparative Analysis of Proposed Model with Relevant Models

Author & Year	Datasets	Accuracy (%)	F1 (%)	Policy Network	Reward	Merits	Limitation
Yang et al. (2019)	Laptop Restaurant Twitter	77.50 83.50 72.40	76.40 74.60 69.80	Policy Gradient algorithm	Delayed Reward	Aspect-level sentiment analysis with external sentiment resources	Failed to extract long-distance context words from the input sentences.
Peng et al. (2019)	Weibo It168 Chn2000 Review-4 Review-5	75.75 86.12 85.45 90.42 90.03	–	Policy Gradient algorithm	Delayed Reward	Learn phonetic information out of pinyin (both from audio clips and pinyin token corpus)	Did not perform aspect-wise sentiment analysis
Pröllochs et al. (2020)	IMDB Movie reviews	43.39	-	Optimal Policy	Simple Reward	Negation detection at the document level	No sentence level negation detection
Chai et al. (2020)	AGNews 20news DBPedia Yahoo Yelp IMDB	95.00 84.60 99.39 78.20 98.00 94.50	-	Policy Gradient algorithm	Normal Reward	Explicit guidelines for what to classify	Time-consuming task
Deshpande and Fleisig (2020)	TextWorld output ClubFloyd banter	-	89.80 75.19	Simple Policy algorithm	Dense Rewards	Generates rewards using sentiments of text descriptions	Did not concentrate on the aspects detection
Chen et al. (2019)	MR AF review AM review	78.90 95.10 93.50	-	Policy Gradient algorithm	Delayed Reward	Detects sentiment from each review sentence	No aspect detection
Wang et al. (2019)	Trip user Trip advisor BeerAdvocate	62.97 52.71 42.41	-	High-level policy and Low-level policy	Simple Reward and Delayed Reward	Aspect-opinion co-extraction from the input sentences	Did not concentrate on the negation detection
Yu et al. (2021)	14Lap 14Rest 15Rest 16Rest	-	59.50 69.61 69.72 68.41	Stochastic policy and Policy Gradient algorithm	Simple Reward	Performed the subtask as a sequential process	Time-consuming task
Zhang et al. (2019)	CMU-MOSI CMU-MOSEI	73.60 71.20	73.50 71.10	Policy Gradient algorithm	Delayed Reward	Feature Extraction - GloVe word embedding Clause structures detection - RL Sentiment classification - LSTM	Less accuracy on the benchmark datasets.

3.1. Preprocessing

The sentences from the review were preprocessed to make it ready to feed the sentence to the proposed ADRSA framework. First, the noise removal task was performed on input review paragraphs for removing punctuations, special characters, and stop words. Then, all the words of the input product review sentence were converted into lower case letters, and the sentences in the input product review paragraph were separated into different sentences based on the dot operator. Spell check was carried out on every input product review sentence using a standard dictionary and the misspelled words were converted into proper words. Then, tokenization was performed for splitting the words of the given sentences into individual tokens. Finally, a lemmatization task converts tokens into standard word forms.



<Figure 1> Proposed ADRSA Framework

3.2. Environment

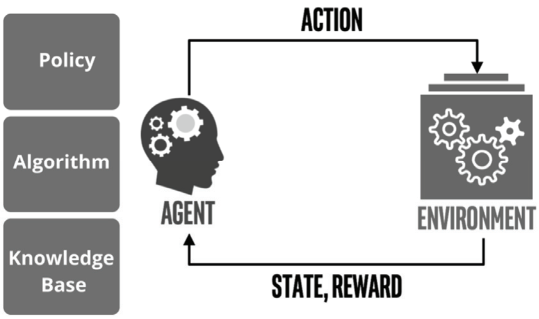
In the proposed model, the environment was constructed as a dynamic environment (<Figure 2>) using an input product review paragraph. Each word from the input review sentences was placed in each row of a matrix. The size of the matrix was dynamic based on the number of words in the review sentence. All the agents interacted with the environment using action and collected state and reward values.

3.3. Agent

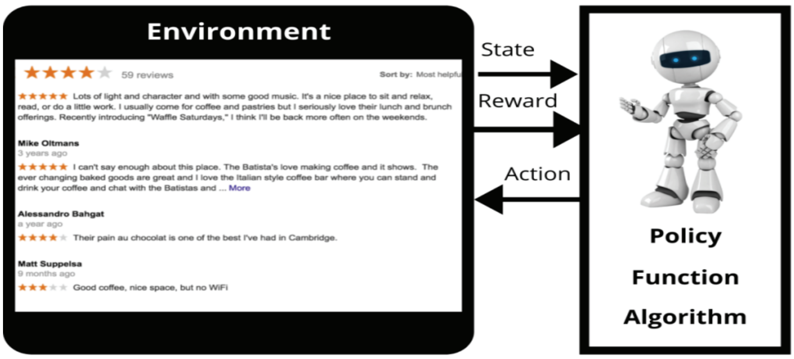
Each agent (<Figure 3>) of the proposed ADRSA model was made up of an algorithm, memory base, filters, neural networks, and policy. Each agent interacted with the environment by its algorithm. Each agent built its knowledge base from the experience by reading review sentences from the environment using state, action values, and received rewards from the environment. The agent used a filter for completing its tasks and savingtime. The agent was implemented using the BERT (Devlin et al., 2019) base model. Input embedding in the BERT model was performed using position embedding, segment embedding, and token embedding as follows:

- Position Embedding: The position of the words in the input product review sentence was expressed by the position embedding.
- Segment Embedding: The pair of sentences in the input product review was compared to find the relationship among them using segment embedding.
- Token Embedding: The specific token from the input product review sentence was learned by BERT using token embedding.

The proposed model was working based on the various parameters shows in <Table 2>.



<Figure 3> Components of an Agent in Reinforcement Learning System



<Figure 2> Dynamic Environment of the Proposed RL Model

<Table 2> Reinforcement Learning System Parameters

Name of the Parameter	Description
S	State
A	Action
s_t	State at time t
a_t	Action at time t
r	Reward at time t
α	Learning rate
γ	Discount factor
Q	Q-value
max_a	Maximum reward

3.4. Reinforcement Learning algorithms

There are two variations of RL algorithms such as on-policy and off-policy are being used. The on-policy RL approach learns the values using the current action value, whereas the off-policy RL approach learns from the action values of other policies. Q-learning (Elavarasan and Vincent, 2020) and Deep Q-Network (DQN) (Hasselt et al., 2016) algorithms are the off-policy RL algorithms, whereas SARSA (Zhao et al., 2016) is an on-policy RL algorithm.

3.4.1. SARSA (State Action Reward State Action)

In the on-policy SARSA algorithm, the agent learns the value according to the current action which is created from the present policy. It was proved in the proposed work that the SARSA algorithm worked better in the NLP problem. Among all RL algorithms, SARSA results were very satisfactory. It is represented with quintuple of parameters (s, a, r, s´, a´), where s represents the current state, a represents the current action taken, r represents the reward after taking action(a), s´ represents the next state and a´ represents the new action. Since it updates the values based on present action (a) and policy, the next possible action (a´) would be correctly chosen by the agent. Agent updates its knowledge frequently and uses its knowledge for further action selection. It updates its Q value based on the following mathematics equation (1) as follows:

$$Q(s, a) = Q(s, a) + \alpha(r+\gamma Q(s', a') - Q(s, a)) \quad (1)$$

The process of SARSA method was explained in the following algorithm as follows:

Algorithm SARSA

Initialize Q(s, a) arbitrarily for all s, a
Repeat (for each episode):
  Initialize s
  Choose **a** from **s** using policy derived from Q
  Repeat (for each step of episode):
    Take action a, observe r, s'
    Choose a' from s' using policy derived from Q
    $y=r + \gamma Q(s', a')$
    $Q(s, a) \leftarrow Q(s, a) + \alpha [y - Q(s, a)]$
    $s \leftarrow s'; a \leftarrow a'$
  until s is terminal
endfor

In SARSA, when an agent is in state s, takes action a, receives a reward r, and moves to the next state s´ based on a policy. The policy maps each state to a particular action. From the new state s´, the agent takes one of the state-action pair values from the current Q-table. In this algorithm, the target policy is the same as the behavior policy. Let $x_1, x_2, \cdots, x_n$ be the sequence of texts in an input review sentence, the agent gets the reward after reading(a) each text from the sentence and moves to the next text(s′) according to the policy. Here the policy says that if the input text is a polarity word (Example: good, bad, worthy, etc.) and there is an aspect word present before the end of the sentence, then the agent gets the reward of 1 for that aspect. Otherwise, the agent gets the reward of 0. If there is a polarity word without an aspect word, then the agent gets the reward for the overall category.

3.4.2. Q-Learning Algorithm

The proposed ADRSA framework was also experimented with using a Q-learning algorithm. Q-learning is the value-oriented RL algorithm for finding optimal policy using Q-function. It is also called an off-policy algorithm which maximizes the Q-value. This algorithm first constructed the Q-table with all values set with 0. Then it chose an action randomly based on the epsilon greedy strategy. Based on the particular action, the values for state and reward were generated. Based on state and reward values, the Q-values of the Q-table were updated using bellman equation (2) as follows:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{(t+1)} + \gamma \max_{\alpha} Q(s_{(t+1)}, a) - Q(s_t, a_t)) \quad (2)$$

The Q-value parameters were updated using a different approach. This states that future rewards are more important than immediate rewards. It was always set less than 0 to converge the algorithm. The maximum reward (max_{α}) helps to choose the optimal action.

3.4.3. Deep Q-Learning Algorithm

Deep Q-Learning algorithm estimates the Q-value function using deep neural network structure which generates all possible actions to be performed by RL agent. It uses memory to store all past experiences. So that using stored experience can find the best solution for the RL agent. It overcomes the Q-learning algorithm generality problem by neural network characteristics which can generate unseen states. DQN works with two techniques namely practice replay and the separate target network. The experience replay solves the sample distribution problem by selecting samples from transition pool storage of transitions for updating the knowledge. A separate target network resolves the issues on fluctuations and stable training by resetting the target network which estimates the value for the Q function. DQN represent value function by deep Q-network with weight (w) which reduces Mean Square Error (MSE) loss based on equations (3), (4), and (5) as follows:

$$Q(s_t, a_t, w) \approx Q'(s_t, a_t) \quad (3)$$

Generally the output is derived from the following equation

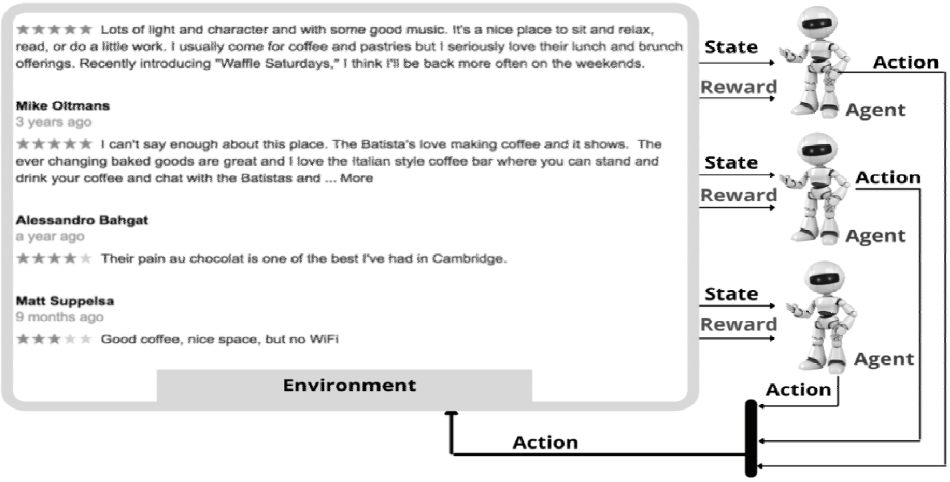
$$Y(S, A, R, S') = R + \gamma \max Q_{\theta}(S', A^1) \quad (4)$$

The loss function is calculated as follows:

$$Loss(\theta) = E_{Y\text{-}U(D)}[(Y-Q)^2] \quad (5)$$

3.5. Multi-Agent RL System

The proposed ADRSA system was designed with a dynamic environment that interacted with more than one agent in which the environment adjusted its matrix size based on the review text length. These systems are called multi-agent RL systems (<Figure 4>). A single-agent RL system was designed to generate single output, whereas the multi-agent system



<Figure 4> Multi-Agent Deep Reinforcement Learning System

was needed where there was more than one output to be generated. Each agent in the multi-agent system functioned autonomously and interacted with the environment in parallel.

For example,

Input review: "battery life is shockingly good. Speaker sounds not up to mark and even volume. Design part I love it very much."

Sentence-1: battery life is shockingly good.

Agent-1 detects a sentiment word "good"
Agent-2 detects negation term "shockingly"
Agent-3 detects an aspect "battery life"

Sentence-2: speaker sounds not up to mark and even volume.

Agent-1 detects a sentiment word "up to mark"
Agent-2 detects negation "not"
Agent-3 detects an aspect term "speaker"

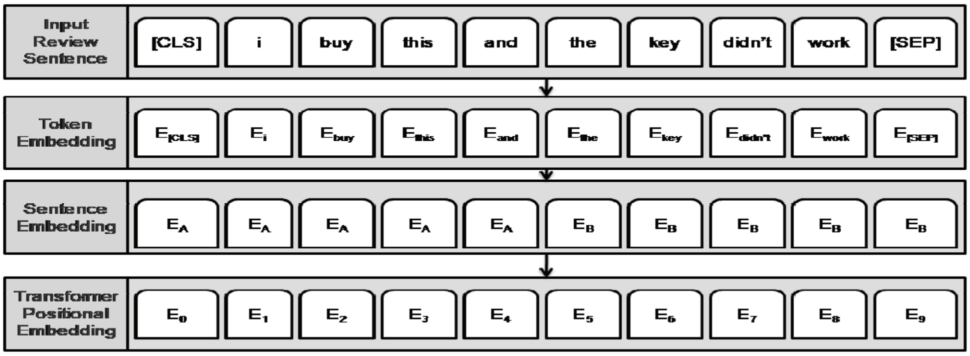
Sentence-3: design part i love it very much.

Agent-1 detects a sentiment term "love"
Agent-2 detects nothing because sentence doesn't have negation term
Agent-3 detects an aspect term "design"

3.6. Aspect Word Detection and Labeling

Aspect word detection agent was implemented using the BERT base model (<Figure 5>) where it finds the semantic relationship between words present in the input review sentence and sub-words with the help of a transformer. The transformer consists of an encoder and decoder, where the encoder reads the input text and the decoder produces the output aspect label. Each word in the review sentence was processed with aspect corpus to detect specific aspects. It identifies the contextual relationship between the words in the given input review text with respect to the set of specific aspects such as battery life, display, camera, sound, performance, storage, value-for-money, and design. When it finds similar words of a given aspect, it adds the label to the sentence and gets a positive reward from the environment.

The input review text sequence was added with the 'CLS' and 'SEP' tokens at the beginning and end. Then, the embedding of each token was created and sentence embedding was created. Transformer positional embedding was made and the aspect was



<Figure 5> BERT Model for Aspect Detection

detected. The aspect categories its label, and its seed words of mobile phone, food, and the movie were presented in <Table 3> as follows:

3.7. Sentiment Word Detection

The proposed model used the BERT model, a bidirectional context-based language model, shown in

<Table 3> Seed Words and Label for Aspect Categories in Mobile, Food and Movie Domain

Domain	Aspect	Label	Seed Words
Mobile	battery	BAT	battery, life, charge, last, power
	display	DIS	display, screen, led, monitor, resolution
	camera	CAM	camera, video, photo
	storage	STO	storage, memory, ram, backup
	sound	SOU	sound, speaker, volume, audio
	design	DES	design, look, appearance
	value-for-money	VFM	price, cost, money, value
Food	food	FOO	food, drink, meal, tiffin, dinner, lunch
	quality	QLT	quality, taste
	price	PRI	price, money, cost
	quantity	QTY	Quantity
Movie	overall	OVA	movie, film
	cast	CAS	act, acting, actress, actor, role, portray, character, villain, performance, performed, played, casting, cast
	director	DIR	direct, direction, directing, director, filmed, filming, filmmaking, filmmaker, cinematic, edition, cinematography
	story	STO	storyline, story, tale, romance, dialog, script, storyteller, ending, storytelling, revenge, betrayal, plot, writing, twist, drama
	scene	SCE	scene, scenery, animation, violence, screenplay, action, special effect, stunt, shot, visual, props, camera, graphic
	music	MUS	lyric, sound, music, audio, musical, title track, sound effect, sound track

<Figure 6> for sentiment word detection from the input review sentence. BERT can read a whole sentence on both sides and understand the context of the sentence. SentiWordNet 3.0 (Baccianella et al., 2010) is a public lexical resource and it was used in this sentiment word detection agent for detecting the sentiment word from each given review sentence. A deep neural network using the BERT system performed extensively to trace the sentiment words on more than 100 epochs. Since it is a bidirectional system using transformers it can overcome the drawbacks of the LSTM system. This task was quite easy, but it affects the results of sentiment analysis.

3.8. Negation Detection

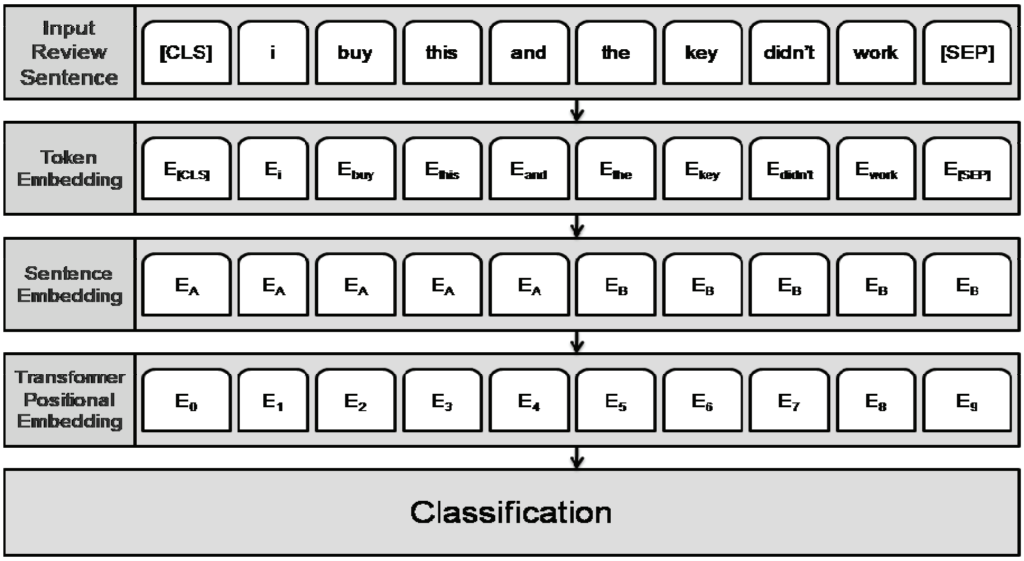
Negation detection helps to classify the sentence correctly. Negation words affect the classification results. So, the presence of the negation words needs to be detected in each review sentence. Some of the words like 'not', 'never', 'no', 'nothing', 'neither', 'nowhere', etc. change the polarity of the word positive to negative and negative to positive. This agent was implemented using the BoW model.

3.9. Sentiment Label Generation

The sentiment label generator received inputs from the sentiment word detection agent and negation detection agent. It generated a sentiment label ($L_{polarity}$)for an input review sentence after performing the 'XOR' operation between detected negation words and detected sentiment word as per equation (6). It worked with the n-gram model to understand the given sentence polarity exactly. When the negation word is followed by the sentiment word, it generates its output. Otherwise, it skipped the process and proceeded with the next sentence.

$$L_{polarity} = L_{sentiment} \oplus L_{negation} \tag{6}$$

where, $L_{polarity}$ refers to final sentiment label, $L_{sentiment}$ refers to label generated by sentiment detector agent and $L_{negation}$ refers to label generated by negation



<Figure 6> BERT Model for Sentiment Word Detection

detector agent.

Ⅳ. Experimental Results and Discussion

4.1. Description of Datasets

The proposed ADRSA model was evaluated using the following three publicly available benchmark datasets namely Movie Review (MR) dataset (Pang and Lee, 2005), Amazon Mobile (AM) review dataset (Meenakshi et al., 2020) and Amazon Food (AF) review dataset (McAuley and Leskovec, 2013). MR dataset contains 5331 positive reviews and 5331 negative reviews. The AM dataset contains 30,000 mobile phone reviews of different brands like Apple, OnePlus and Redmi. AF dataset contains 5,68,454 fine food reviews from Amazon.

4.2. Experimental Setup

TextWorld (Cote et al., 2019) python framework was used in the proposed framework to experiment with the datasets for aspect-based sentiment analysis by building text-based games and for training and testing the RL agents. The agents send the text commands to interact with the environment through a game interpreter. The environment changes its state based on valid commands and returns a delayed reward as sentiment. This text-based game understands various aspects of the product as objects and follows a set of rules as the policy to generate class labels. The policy network was implemented using the deep neural network which consists of 300 neurons, assuming an initial input length of 300, with three hidden layers along with a learning rate of 0.01. Since no training was given to the system, the system was tested with a live input dataset that contains 10,000 reviews. 0.99 was used as a discount factor in this system. SentiWordNet 3.0 was used for finding sentiment words in the input review sentences.

4.3. Hyper Parameters of the Proposed Model

The fine-tuning of the proposed model was done using the following hyper parameters.

- Learning rate: The learning rate (α) value is set between 0 and 1. When a value is 0, there is no learning occuring and when it is 0.8 for example, learning happens quickly
- *Discount factor:* The discount factor value affects the future rewards of the agents. The discount factor (ɣ) is also set between 0 and 1. It was used to make future rewards worth it for the agents
- *Optimizer:* Adam optimizer was used in the proposed model on the BERT network for minimizing prediction loss
- *BERT network parameters:* Two forms of BERT neural network was experimented in the proposed model with the following parameters shown in <Table 4> as follows:

<Table 4> BERT Base and BERT Large Parameters

Parameter	BERT Base	BERT Large
Transformer Blocks	12 Layers	24 Layers
Hidden Layers	768	1024
Attention Heads	12	16
Parameters	110 million	340 million

4.4. Performance of the Proposed Model

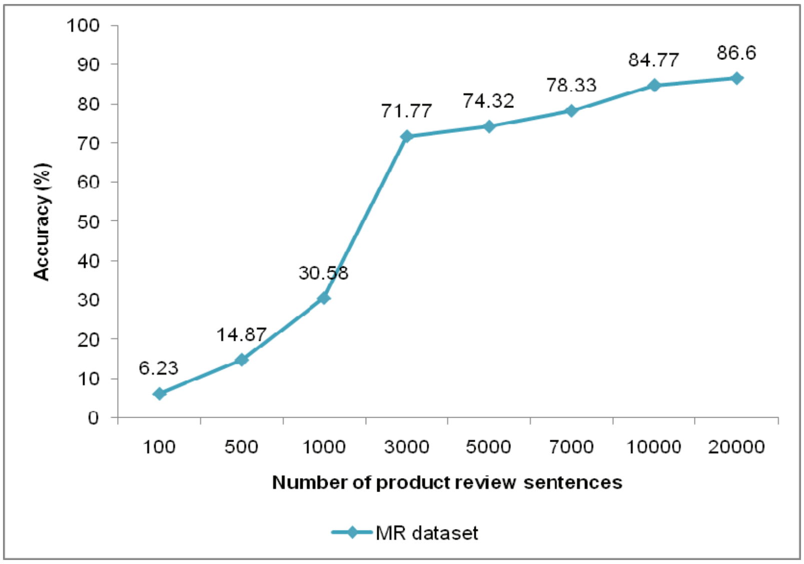
The proposed ADRSA model was experimented on three benchmark datasets using Q-learning, DQN, and SARSA RL algorithms as shown in <Table 5>. Since the Q-learning RL algorithm is a model-free algorithm for learning a value of an action, it handled the dataset with stochastic transitions and received rewards without adaptations. So, the accuracy of the proposed model using Q-learning algorithms was 11.26%, 9.30%, and 16.16% less than the proposed model accuracy with SARSA algorithm on MR, AM,

<Table 5> Comparison of Proposed ADRSA Model Accuracy Implemented using SARSA, DQN and Q-Learning RL Algorithms

Algorithm	MR Dataset (%)	AM Dataset (%)	AF Dataset (%)
ADRSA + Q-Learning	75.34	76.22	74.54
ADRSA + DQN	82.15	83.11	85.12
ADRSA + SARSA	**86.62**	**97.94**	**97.80**

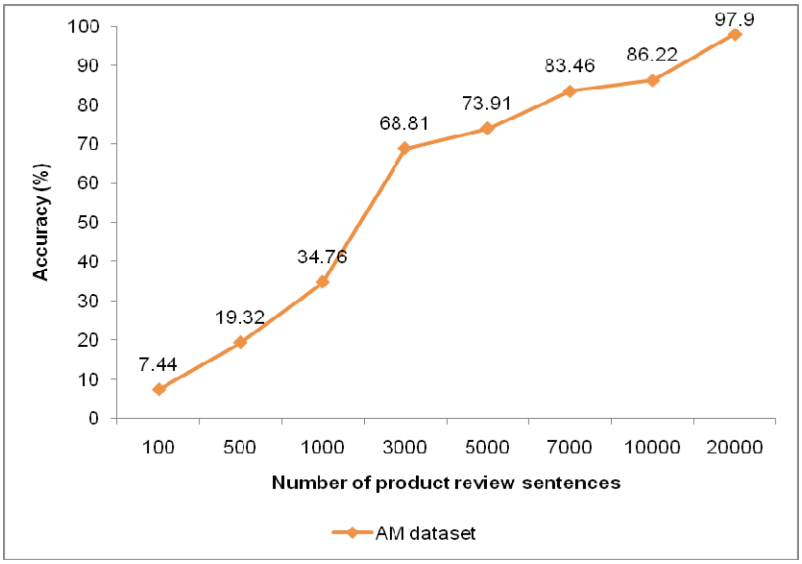
<Table 6> Comparing Accuracy of Proposed ADRSA Model using BERT Network with CNN, LSTM, RCNN and WS-LSTM Network on Benchmark Datasets

Method	MR Dataset (%)	AM Dataset (%)	AF Dataset (%)
CNN(Chen et al.,2019)	73.83	90.20	94.33
LSTM(Chen et al.,2019)	74.51	96.74	92.24
RCNN(Chen et al.,2019)	76.13	95.22	94.70
WS-LSTM(Chen et al.,2019)	78.61	93.50	95.13
ADRSA with BERT	**86.62**	**97.94**	**97.80**

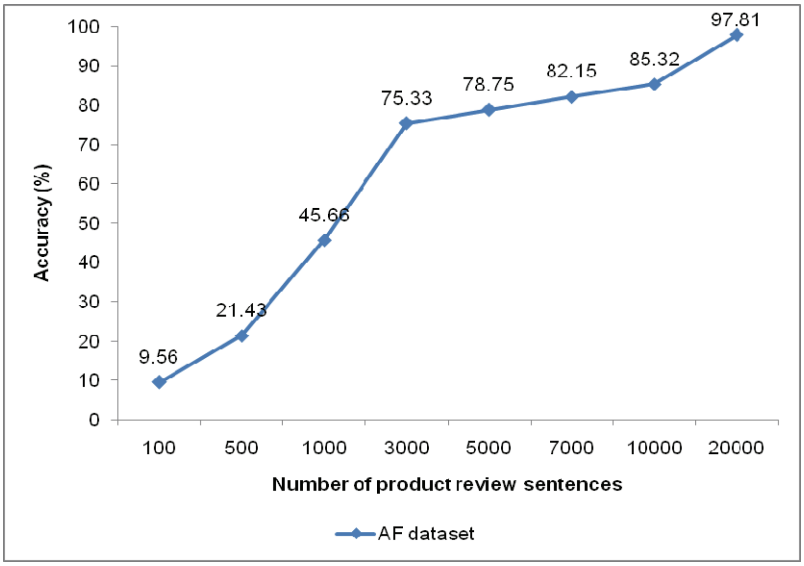


<Figure 7> Accuracy of Sentiment Classiction of Proposed ADRSA Model on MR Dataset

and AF datasets respectively. Since the DQN is a value-based, model-free, and off-policy RL algorithm, the agent was trained with discrete action space and received rewards with action specifications. So, the accuracy of the proposed model using DQN algorithms was 4.45%, 2.39%, and 5.58% less than the proposed model accuracy with SARSA algorithm on MR, AM, and AF datasets respectively.



<Figure 8> Accuracy of Sentiment Classification of Proposed ADRSA Model on AM Dataset



<Figure 9> Accuracy of Sentiment Classification of Proposed ADRSA Model on AF Dataset

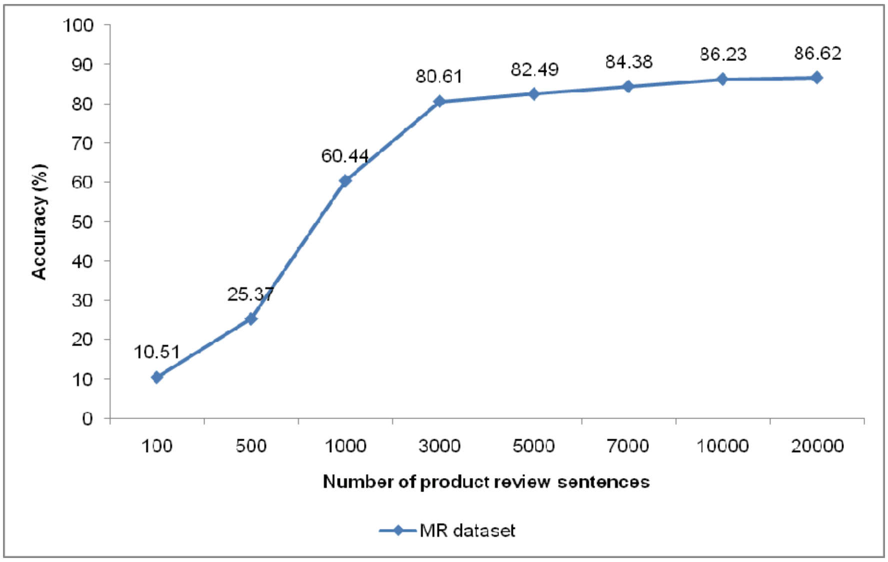
Aspect-based sentiment classification performance at sentence level was presented in <Table 6>. The proposed ADRSA model was compared with well-known neural network algorithms for text sentiment classification such as CNN, LSTM, Recurrent CNN (RCNN), and Weight Sharing LSTM (WS-LSTM). The proposed model outperformed all other existing models with an 8.0% accuracy difference on the MR dataset, 1.2% improvement accuracy on AM dataset, and 2.7% more accuracy on the AF dataset.

The performance of the proposed ADRSA model using three agents for aspect-based sentiment analysis was monitored for every sentence of an input product review. The performance of the sentiment classification was presented in <Figure 7-9>.

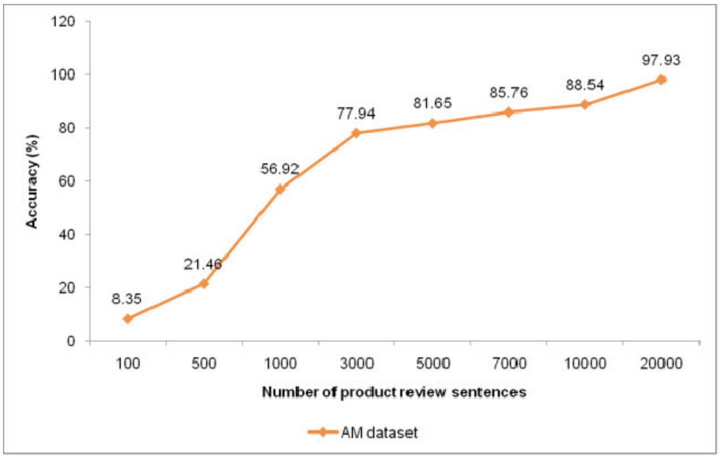
The accuracy of the sentiment classification was gradually increased from 9.56% after reading 100 sentences to 94.33% after reading 20,000 sentences from the MR dataset. The accuracy of the sentiment classification was increased from 7.44% after reading 100 sentences to 97.9% after reading 20,000 sentences from the AM dataset. Similarly, the accuracy of the sentiment classification was increased from 9.56% after reading 100 sentences to 97.81% after reading 20,000 sentences from the AF dataset.

The performance of the aspect classification was presented in <Figure 10-12>. The accuracy of the aspect classification was gradually increased from 10.51% after reading 100 sentences to 90.56% after reading 20,000 sentences from the MR dataset. The accuracy of the aspect classification was increased from 8.35% after reading 100 sentences to 92.67% after reading 20,000 sentences from the AM dataset. Similarly, the accuracy of the aspect classification was increased from 9.23% after reading 100 sentences to 96.56% after reading 20,000 sentences from the AF dataset.

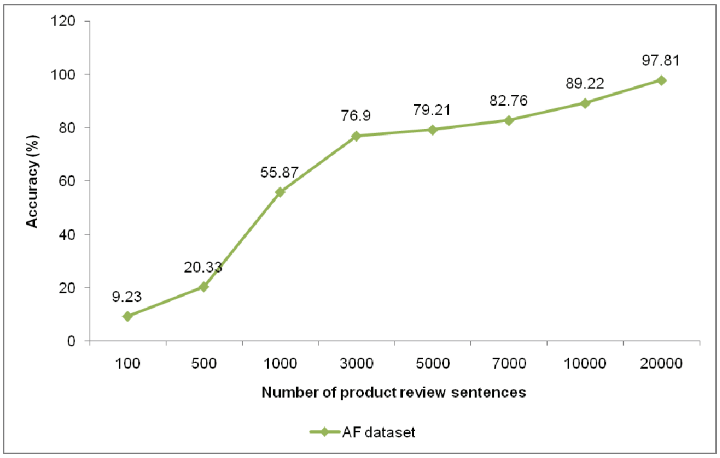
The knowledge level of the agents was improved by constructing a knowledge base on each agent. The proposed system was trained and tested using three datasets with more than 1 million reviews. It was converted into an intelligent system with very good



<Figure 10> Accuracy of Aspect Classification of Proposed ADRSA Model on MR Dataset



<Figure 11> Accuracy of Aspect Classification of Proposed ADRSA Model on AM Dataset



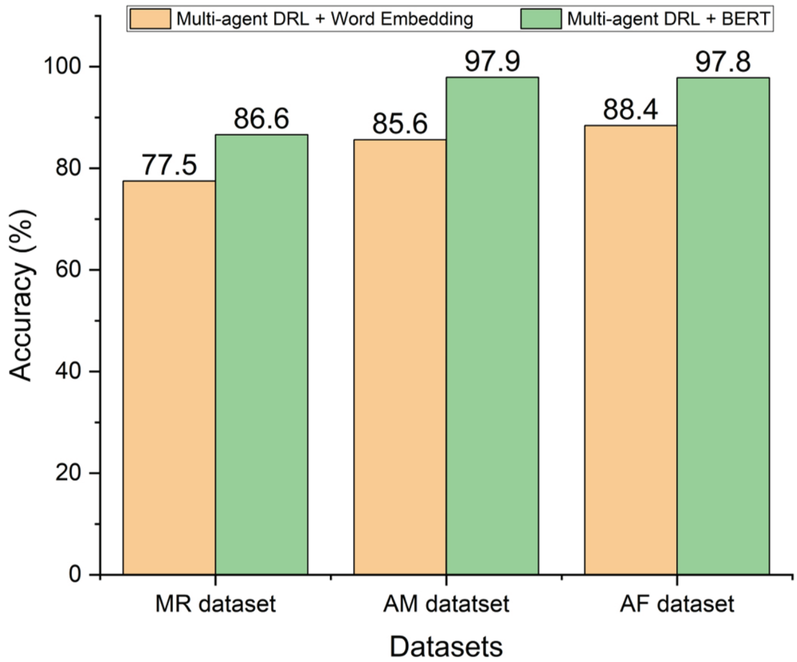
<Figure 12> Accuracy of Aspect Classification of Proposed ADRSA Model on AF Dataset

performance. From this, we can conclude that this system can be used on any new dataset of any domain for understanding the sentiment expressed in the sentences. The execution of the experiment was stopped after 20,000 sentences and results converged after that. It was also observed that the system was able to build an index of aspects for the particular domain with very good vocabulary.

The proposed model was also experimented with using CNN, LSTM, RCNN, and WS-LSTM networks, and comparison was presented in <Table 7>. This proposed BERT network model performed well with the accuracy of 86.6%, 97.9%, and 97.8% on MR, AM, and AF datasets respectively. The accuracy of the proposed model was better than the CNN network with 15.8%, 11.4%, and 14.2% on MR, AM, and AF datasets respectively. The performance of the proposed model was better than the LSTM network with higher accuracy of 13.3%, 11.8%, and 8.0% on MR, AM, and AF datasets. The proposed model performed

<Table 7> Comparison the Proposed Model Accuracy Implemented using State-of-Art Models for Aspect-Based Sentiment Analysis

Model	MR Dataset (%)	AM Dataset (%)	AF Dataset (%)
CNN	70.82	86.52	83.61
LSTM	73.50	88.11	89.82
RCNN	72.12	90.30	92.40
WS-LSTM	75.64	90.83	91.91
Multi-agent DRL + Word Embedding	77.51	85.61	88.42
Multi-agent DRL + BERT	**86.62**	**97.94**	**97.80**



<Figure 13> Accuracy of the Proposed Multi-Agent DRL Model with BERT and Word Embedding

better than the RCNN network with the higher accuracy of 14.5%, 7.6%, and 5.4% on MR, AM, and AF datasets. The proposed system outperforms with the higher accuracy of 11.0%, 7.1%, and 5.9% on MR, AM, and AF datasets than the system using the WS-LSTM network. So, the usage of the BERT network in the proposed model for aspect-based sentiment analysis improved the performance of the system with good accuracy.

The proposed ADRSA model was also experimented with and without using the BERT network for aspect detection and sentiment word detection on three benchmark datasets. It was presented in <Figure 13> and the proposed model using BERT performance was extremely good with an accuracy difference of 9.1% on the MR dataset, 12.3% on AM dataset, and 9.4% on the AF dataset.

4.5. The Proposed Model as a Business Problem

At present, customer feedback is collected through star rating and optionally customers are expected to give feedback in the form of text. The present system is showing overall feedback about the product on its page. But, the customer will be happy, if they are able to visualize aspect-wise feedback about a product. Customers spend more time searching aspect-wise feedback about the product. So, the proposed ADRSA framework performs aspect-wise sentiment analysis on live customer review and visualizes aspect-wise sentiment immediately to the other customers. After processing more customer reviews, the framework becomes an intelligent system to categorize sentiments expressed on various aspects that may reduce the customer time on reading more reviews to understand the pros and cons of a product. This may also help the product manufacturers to improve their business by knowing product drawbacks and customers' expectations.

Ⅴ. Conclusion and Future Work

Aspect-based sentiment analysis was experimented with using supervised and unsupervised learning methods in past decades with static datasets. But the proposed ADRSA model performed the aspect-based sentiment analysis using deep reinforcement learning with SARSA RL algorithm on the dynamic data sets from online web resources for helping people to make a better decision before buying any product online on-demand. Instead of word embedding, the proposed model used the BERT network for feature extraction and aspect extraction from input product reviews at sentence level without losing the meaning of the sentences and the relationship among the sentences. This first step will move on further for all sectors for processing dynamic data without past datasets help for training the system. The accuracy of the proposed ADRSA model was higher than the state of art methods for aspect-based sentiment analysis. This work can also be extended by using a model-based reinforcement learning system to improve further on understanding all languages.

References

[1] Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), 26-38.

[2] Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10),* 2200-2204.

[3] Boute, R. N., Gijsbrechts, J., van Jaarsveld, W., and Vanvuchelen, N. (2021). Deep reinforcement learning for inventory control: A roadmap. *European Journal of Operational Research*, *298*(2), 401-412. https://doi.org/10.1016/j.ejor.2021.07.016.

[4] Chai, D., Wu, W., Han, Q., Wu, F., and Li, J. (2020). Description based text classification with reinforcement learning. *International Conference on Machine Learning,* 1371-1382.

[5] Chen, R., Zhou, Y., Zhang, L., and Duan, X. (2019). Word-level sentiment analysis with reinforcement learning. *IOP Conference Series: Materials Science and Engineering, 490,* 062063.

[6] Chu, T., Wang, J., Codecà, L., and Li, Z. (2020). Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems, 21*, 1086-1095.

[7] Côté, M., Kádár, Á., Yuan, X., Kybartas, B. A., Barnes, T., Fine, E., Moore, J., Hausknecht, M. J., Asri, L.

E., Adada, M., Tay, W., and Trischler, A. (2018). TextWorld a learning environment for text-based games. *Computer Games Workshop at ICML/IJCAI*. arXiv:1806.11532.

[8] Deshpande, A., and Fleisig, E. (2020). Sentiment Analysis for Reinforcement Learning. arXiv:2010.02316.

[9] Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT),* 4171-4186.

[10] Diallo, E. A. O., Sugiyama, A., and Sugawara, T. (2020). Coordinated behavior of cooperative agents using deep reinforcement learning. *Neuro computing*, *396*, 230-240.

[11] Elavarasan, D., and Vincent, P. M. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE Access, 8*, 86886-86901.

[12] Goh, O. S., Fung, C. C., and Depickere, A. (2008). Domain knowledge query conversation bots in instant messaging (IM). *Knowledge-Based Systems*, *21*(7), 681-691.

[13] Goldwaser, A., and Thielscher, M. (2020). Deep reinforcement learning for general game playing. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(2), 1701-1708.

[14] Hasselt, H. V., Guez, A., and Silver, D. (2016). Deep reinforcement learning with double Q-learning. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, 2094-2100.

[15] Jonsson, A. (2019), Deep reinforcement learning in medicine. *Kidney Dis*, *5*, 18-22.

[16] Keller, T. R., and Klinger, U. (2019). Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, *36*(1), 171-189.

[17] Koratamaddi, P., Wadhwani, K., Gupta, M., and Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal*, *24*(4), 848-859.

[18] Kulkarni, C. S., Bhavsar, A. U., Pingale, S. R., and Kumbhar, S. S. (2017). BANK CHAT BOT-an intelligent assistant system using NLP and machine learning. *International Research Journal of Engineering and Technology*, *4*(5), 2374-2377.

[19] Liu, R., Nageotte, F., Zanne, P., Mathelin, M. D., and Dresp, B. (2021). Deep reinforcement learning for the control of robotic manipulation: A focussed mini-review. *Robotics, 10*(22), 1-13.

[20] Meenakshi, Banerjee, A., Intwala, N., and Sawant, V. (2020). Sentiment analysis of amazon mobile reviews. In M. Tuba, S. Akashe, and A. Joshi (Eds.), *ICT Systems and Sustainability. Advances in Intelligent Systems and Computing*, *1077*. Singapore. https://
doi.org/10.1007/978-981-15-0936-0_4

[21] McAuley, J., and Leskovec, J. (2013). From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. *Proceedings of the 22nd international conference on World Wide Web*.

[22] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

[23] Pang, B., and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Association for Computational Linguistics*, 115-124.

[24] Peng, H., Ma, Y., Poria, S., Li, Y., and Cambria, E. (2021). Phonetic-enriched text representation for chinese sentiment analysis with reinforcement learning. *Information Fusion,70*, 88-99.

[25] Pröllochs, N., Feuerriegel, S., Lutz, B., and Neumann, D. (2020). Negation scope detection for sentiment analysis a reinforcement learning framework for replicating human interpretations. *Information Sciences*, 536, 205-221.

[26] Rowley, J. (2000). Product searching with shopping bots. *Internet Research*, *10*(3), 203-214.

[27] Shahmanzari, M., and Ozkan, S. (2014). Accessing

the effect of e-commerce intelligent bots on online consumers' post-adoption behavior for future use. *American Academic & Scholarly Research Journal*, *6*(4), 163.

[28] Wang, J., Sun, C., Li, S., Wang, J., Si, L., Zhang, M., Liu, X., and Zhou, G. (2019). Human-like decision making: Document-level aspect sentiment classification via hierarchical reinforcement learning. *EMNLP*, 5581-5590.

[29] Wang, W. Y., Li, J., and He, X. (2018). Deep reinforcement learning for NLP. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, 19-21.

[30] Yang, M., Jiang, Q., Shen, Y., Wu, Q., Zhao, Z., and Zhou, W. (2019). Hierarchical human-like strategy for aspect-level sentiment classification with sentiment linguistic knowledge and reinforcement learning. *Neural Networks: The Official Journal of the International Neural Network Society, 117*, 240-248.

[31] Yu, L., Gao, Z., Qin, S., Zhang, M., Shen, C., Guan, X., and Yue, D. (2020). Deep Reinforcement Learning for Smart Grid Protection Against Coordinated Multistage Transmission Line Attacks. *arXiv:2011. 14526*.

[32] Yu, S., Nayak, T., Majumder, N., and Poria, S. (2021). Aspect sentiment triplet extraction using reinforcement learning. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management,* 3603-3607.

[33] Zhang, C. (2019). Sentiment Analysis and Deep Reinforcement Learning for Algorithmic Trading. Retrieved from http://ink-ron.usc.edu/xiangren/ml4know19spring/public/midterm/Chi_Zhang_Report.pdf

[34] Zhang, D., Li, S., Zhu, Q., and Zhou, G. (2019). Modeling the clause-level structure to multimodal sentiment analysis via reinforcement learning. *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 730-735.

[35] Zhao, D., Wang, H., Shao, K., and Zhu, Y. (2016). Deep reinforcement learning with experience replay based on SARSA. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1-6.

◆ About the Authors ◆

M.Sivakumar

M.Sivakumar is a Ph.D. candidate in the Department of Computer Applications at National Institute of Technology in Tiruchirappali, India. He is currently working as an assistant professor in the department of computer science and engineering at K.Ramakrishnan College of Technology, Tiruchirappalli, Tamilnadu, India. He is a life time member of Indian Society of Technical Education (ISTE). He is a certified Microsoft Innovative Educator. He has a master's degree in Computer Science and Engineering from Anna University in Chennai, India. His research interests include Sentiment analysis of customer reviews using deep learning. He has published more than 10 journal articles & conferences and 1 book chapter.

Srinivasulu Reddy Uyyala

Srinivasulu Reddy Uyyala received his Ph.D. in 2013, Department of Computer Applications, National Institute of Technology (NIT), Tiruchirappalli, Tamilnadu, India. He started his research career as a full-time research scholar in the department of computer applications, NIT, Tiruchirappalli from 2006. He joined NIT, Tiruchirappalli as an Assistant Professor in 2008. His research interest lies in applying various artifcial intelligent approaches to extract useful patterns from different facets of societal challenges of high demand with economic benefit. The main aim of his work is to develop machine learning and deep learning models for various domains like agriculture, bioinformatics, banking, retail, telecom, and healthcare. With this motivation, he has established "Machine Learning and Data Analytics Lab" in 2015 and he is a core team member of the Centre of Excellence (CoE) in Artifcial Intelligence established in 2020. His research team of 12 doctoral students, 10 post-graduation students are highly motivated and enthusiastic in working out to achieve better outcomes. For his credit, he has produced 7 doctoral till date and 10 post-graduation students annually. He has published more than 50 journal articles & conferences and 2 book chapters. He received "Best Researcher Award" in 2018, "Best Performer Award" in 2019, and "Best Performer Award" in 2020 from NIT-Tiruchirappalli. Two "Best Paper Awards" in International Conferences. He received a full educational scholarship under "Residential Care Program" from Kindernothilfe (KNH), Germany from schooling to Master's.

Submitted: August 26, 2021; 1st Revision: December 7, 2021; 2nd Revision: February 16, 2022;
Accepted: March 4, 2022