Emotional Intelligence Attention Unsupervised Learning Using Lexicon Analysis for Irony-based Advertising

USMAN AHMED and JERRY CHUN-WEI LIN, Western Norway University
of Applied Sciences, Norway
GAUTAM SRIVASTAVA, Brandon University, Canada, China Medical University, Taiwan, and Lebanese
American University, Lebanon

Social media platforms have made increasing use of irony in recent years. Users can express their ironic thoughts with audio, video, and images attached to text content. When you use irony, you are making fun of a situation or trying to make a point. It can also express frustration or highlight the absurdity of a situation. The use of irony in social media is likely to continue to increase, no matter the reason. By using syntactic information in conjunction with semantic exploration, we show that attention networks can be enhanced. Using learned embedding, unsupervised learning encodes word order into a joint space. By evaluating the entropy of an example class and adding instances, the active learning method uses the shared representation as a query to retrieve semantically similar sentences from a knowledge base. In this way, the algorithm can identify the instance with the maximum uncertainty and extract the most informative example from the training set. An ironic network trained for each labelled record is used to train a classifier (model). The partial training model and the original labelled data generate pseudo-labels for the unlabeled data. To correctly predict the label of a dataset, a classifier (attention network) updates the pseudo-labels for the remaining datasets. After the experimental evaluation of the 1,021 annotated texts, the proposed model performed better than the baseline models, achieving an F1 score of 0.63 on ironic tasks and 0.59 on non-ironic tasks. We also found that the proposed model generalized well to new instances of datasets.

CCS Concepts: • **Computing methodologies → Lexical semantics**; **Natural language processing**; • **Applied computing → Psychology**;

Additional Key Words and Phrases: Information retrieval, Emotional Intelligence, Natural Language Processing, irony detection, unsupervised learning

Authors' addresses: U. Ahmed and J. C.-W. Lin (corresponding author), Department of Computer Science, Electrical Engineering and Mathematical Sciences Western Norway University of Applied Sciences, Inndalsveien 28, 5063, Bergen, Norway; emails: usman.ahmed@hvl.no, jerrylin@ieee.org; G. Srivastava, Department of Math and Computer Science, Brandon University, Brandon, MB R7A 6A9, Canada; Research Centre for Interneural Computing, China Medical University, Taichung 40402, Taiwan and Department of Computer Science and Math, Lebanese American University, Beirut 1102, Lebanon; email: SRIVASTAVAG@brandonu.ca.

1 INTRODUCTION

Several applications, including human–computer interaction, computational learning, computer-assisted interviewing, autonomous agents, and crowd simulations involve the automatic recognition of overt feelings, emotions, sarcasm, and humor. Indeed, chatbots have now entered the realm of context-aware communication. Chatbots based on **Natural Language Processing (NLP)** drive growth and profitability, improve the customer experience, and streamline business operations. NLP is the general term for natural language understanding, natural language generation, and dialogue management. Typically, NLP is about finding the meaning or intent hidden in human language. Together with Artificial Intelligence/Cognitive Computing, Natural Language Processing enables an easy understanding of the meaning of words in context, taking into account abbreviations, acronyms, slang, and so on. In the field of Artificial Intelligence, NLP enables robots and applications to interpret the intent of human speech input and generate appropriate responses, resulting in a natural flow of conversation. This is a fantastic opportunity for companies to gather strategic information such as preferences, opinions, purchasing patterns, and emotions. Companies can use this information to identify trends, uncover operational risks, and gain actionable insights.

Among the symbolic and creative language seen on social media platforms is irony [1, 4]. This phenomenon is studied in the fields of linguistics, philosophy, and rhetoric as a multifaceted, controversial, and fascinating English phenomenon. Detecting irony impacts sentiment analysis, hate speech detection, and fake news detection. Automatic irony detection can enhance the performance of such systems [12].

According to the *New Princeton Encyclopedia of Poetry and Poetics*, classical, romantic, tragic, cosmic, verbal, situational, dramatic, and poetic are the most prevalent types of irony in social networks. In general, situational and linguistic irony is the most prevalent [13]. On the one hand, situational irony is related to paradoxical scenarios regarding specific events, such as "A security company is the most recent victim of a malware attack." On the other hand, other authors have described verbal irony as imparting a meaning that contradicts the literal meaning, as in, "Oh, look, there is another storm in Sydney. How weird." Sarcasm is an oft-studied form of verbal irony in which the speaker's intention is to criticize an individual or subject in an ironic manner. As one of the most widely used forms of irony, sarcasm has been the subject of much scholarly analysis. In text messages, it is a subjective, complex, and event-dependent skill [12]. Several essential aspects, such as polarity discrepancies, common-sense knowledge, similes with "over" or "as" structures, punctuation or entreaty, emotional features, negation and contextual features, context incongruence, and so on have been identified to solve the ironic detection problem. However, computational techniques that adhere to the text compositionality principle cannot explain textual irony based solely on the word composition of a message. This is mainly because numerous ironic indicators, both kinesthetic (facial expressions and hand movements) and verbal characteristics, are lost in text messages (voice tone, rhythm, silence, etc.).

1.1 Motivation

Pool-based active learning algorithms are used when the learner has access to a pool of examples from which to learn. The idea is that the algorithm learns from the examples in the pool and is able to generalize to new examples that are not in the pool. Active learning algorithms based on a pool have many advantages. One advantage is that they can help reduce data needed to train a **Machine Learning (ML)** model. This is because the algorithm can learn from several examples and then generalize to new data. This can be helpful when data are scarce or expensive. Another advantage of pool-based active learning algorithms is that they can help improve the performance of the machine learning model. This is because the algorithm can learn from its mistakes and avoid those mistakes in the future. This can lead to a more accurate and efficient machine learning

model. It is known that active learning can effectively reduce the amount of data needed to train a machine learning model. Pool-based uncertainty sampling is a specific active learning technique that is particularly effective in various situations. Pool-based uncertainty sampling can be used with various machine learning algorithms, including supervised and unsupervised methods. This flexibility may be necessary for situations where data is very complex and it is unclear which algorithm is most effective.

1.2 Contribution

This study shows how NLP and attention-based contrastive active learning can support irony detection learning. A context extraction method was used to represent semantic vectors. The method takes relevant marginal elements from the text that are not labelled and brings them into the active learning process. The strategy adds more examples to train the model. It is repeated until the best solution is found. A pool of texts with no labels is added to the training set. The study seeks to understand how people learn to add text data. By using the proposed method, the amount of labelling work required can be reduced, and the learning system can learn better from new situations. The semantic vectors of the network and synonym expansion help achieve high accuracy without affecting data annotations. Besides labelling, irony detection also requires key phrase extraction. In particular, this paper includes the following contributions:

(1) Propose a lexicon expansion to extend the syntactic and semantic relations between words conveying the same context as part of the proposed method.
(2) Show a method for using pre-trained word embeddings in a context generated from unlabeled data, the active learning function uses a pool of unlabeled data to find examples that differ.
(3) Introduce model development and evaluation using the attention network learned representation by combining the contrast set, entropy, least confidence and random sub-sampling techniques.

2 RELATED WORK

Several types of approaches to irony detection can be distinguished: rule-based methods, traditional machine learning methods that utilize features, and deep neural network methods [4, 6, 23]. In Twitter irony detection, linguistic features are typically used, such as sentiment lexicons or hashtags, which alter the literal meaning of tweets using hashtags. Among the most commonly used hashtags on Twitter are "Irony," "sarcasm," and "not" to show irony [21]. People think that using hashtags like "#sarcasm" is a way to replace language markers like exclamations and words that mean "very" or "very much." Classic feature-based machine learning methods use hand-crafted features to find irony, such as a sentiment lexicon, a subjectivity lexicon, features for emotional categories, features for emotional dimensions, or structural features [23].

There have been extensive uses of literary devices in online communication, including tweets on Twitter and other forms of sarcastic communication. The irony is characterized by being out of place in the exchange context [23]. Practical sentiment analysis requires an accurate identification of irony. The irony in phrases like "I love being ignored" is established by the conflict between the word "love" and its negative connotation, "being ignored." Context incongruence is currently detected using supervised learning text categorization that relies on explicit expressions. Irony detection is a transfer learning task in research when knowledge obtained from external sentiment analysis resources is added to supervised learning on material labelled with irony. Unlike the previous strategy, which relies on statements such as "I prefer to think of myself as a broken Justin Bieber-my philosophy professor," this approach offers three techniques based on transfer

learning for improving the learning algorithms of recurrent neural models to detect irregular hidden patterns by leveraging emotional information. The study's fundamental discovery is that using sentiment data from outside sources is highly effective in enhancing irony identification. Further, it appears that neural networks are more efficient at detecting implicit incongruity than cutting-edge neural models in detecting irony.

The topic of computationally based irony detection on Twitter has been the subject of several recent workshops. Irony is a figurative language that influences human communication in natural language, mainly social media. A model for detecting Irony in Twitter posts using Transformer architecture that contextualizes pre-trained word embeddings is presented in this study. In addition to using the same robust architecture as BERT, this method can also use embeddings within the domains [12]. As part of the study, extensive analysis was conducted of both Spanish and English corpora. Using these results, we believe the concept is sufficiently and accurately formulated. In the Spanish corpus, the system ranked at the top, whereas the system ranked the second highest in the English corpus. This study examined and evaluated multiple-headed self-attention mechanisms as to their specialization in irony identification by examining the polarity and meaning of individual words and their connections to one another. The results of this study provide a better understanding of the multi-head self-attention processes used by the transformer architecture to handle the ironic detection problem as a consequence of these findings.

The use of transfer learning has been successfully applied to several real-world problems that classical ML algorithms cannot handle, such as image processing, audio recognition, and NLP [19]. Transfer learning is a NLP technique in which a model trained for one task is reused for a related but different task using well-labelled data and data formats. Even if the model has never been exposed to a particular sentence, it can be used to predict the part of speech of words in a sentence by using a model trained on a large corpus of English texts. A transfer learning approach can be used to improve the performance of NLP models when data is limited, since features learned from the model (e.g., the order in which words appear before other words) can be generalized to other tasks. Suppose there are limited training data for a new task. It may be beneficial to first train the model on a similar task with more data and then apply the learned features to the new task. In addition to fitting a model to a new domain, this technique can also be used to fine-tune a model that was initially trained on news articles [7]. Data and the task used to perform transfer learning will determine the method used. Fine-tuning pre-trained models and using those models as feature extractors are some examples of standard methods.

A semi-supervised method for studying cyber-racism would involve analyzing a dataset of online comments and posts for instances of cyber-racism using a combination of supervised and unsupervised learning techniques. This strategy requires manually labelling a small amount of data to develop a supervised machine learning model for detecting racist content. The unsupervised portion of the research would involve applying techniques such as topic modelling to data clusters and searching for cyber-racism patterns as part of the unsupervised portion of the research. The application of this methodology would allow for automated analysis of large amounts of data, while manual labelling could still be monitored by humans, facilitating the study of cyber-racism [5]. The recent coronavirus outbreak was studied using a semi-supervised method. A major goal of this study was to develop machine learning models to detect cyber-racism and model topics using **Latent Dirichlet Allocation (LDA)**. In March 2020, 7,454 clean tweets with the Chinese hashtags "virus" and "Kung Flu" were collected on Twitter. As part of the training process for machine learning models, unfavourable tweets (racism, sarcasm/irony, and others) were flagged using sentiment analysis. The study examined the efficacy of different machine learning algorithms in distinguishing between racism and sarcasm/irony. A comparison was made between Random Forest, J48, and Support Vector Machine models. Random Forest was found to be the most successful

model, achieving 78.1% accuracy in distinguishing racism from sarcasm/irony and 77.7% accuracy in distinguishing racism from other sentiments. These results indicate that Random Forest is the best option for automated racism detection. The LDA distinguishes between three categories of racist tweets: Eating Habits, Political Hostility, and Xenophobia. Based on their consistent performance in detecting cyber-racism patterns based on text exchanges, the evaluated models can be considered reliable.

Irony, sarcasm, and satire are among the most important applications of sentiment analysis. This is because these linguistic devices often involve some sort of emotional or mental state that can be detected and analyzed by a computer. With irony, for example, you usually say something that is the opposite of what you actually mean to make a point or to be funny. Sarcasm is similar but is often used to be mean or hurtful to someone. And satire is often used to criticize or make fun of something. Since sarcastic criticisms are positive in their words but negative in their emotions, it is difficult to determine irony in satire reviews [17]. This study was conducted to identify irony in Amazon product reviews posted on Twitter. We examined lexicon-based features using N-gram and skip-gram algorithms to detect irony. Using a novel **deep learning (DL)** technique, the study collected and analyzed 22,000 tweets that contained both ironic and non-ironic messages about Amazon products to detect and predict irony. The implementation of the recommended irony detection work, which includes **decision trees (DT)**, support vector machines, logistic regressions, and **random forest (RF)**. Compared to the traditional models DT and RF, the proposed model DL has average results.

If a restaurant review mentions service, food, and ambiance, then the mood level can be calculated for each of these three aspects [22]. The advantages of categorizing mood at the aspect level are twofold. First, it allows for a more nuanced understanding of the sentiment expressed in a rating or sentence. Second, it can provide insights that can be used to improve the product or service being rated. For example, if the sentiment toward food is negative, then the restaurant can work to improve its food. An aspect-level sentiment categorization is a valuable tool for sentiment analysis, and its use is likely to increase as researchers develop more sophisticated methods for sentiment analysis. The small datasets of neural network models limit their ability to classify feelings at the aspect level, because it is difficult to label such data.

In addition to the data spontaneously provided by social media, most studies require time-consuming content and sentiment analysis. Due to the huge amount of data and the development of data science, a transfer learning-based method for analyzing public perceptions of **alternative meat (AM)** in China can be considered a valuable alternative. Using the annotated sample, Naive Bayes and Support Vector Machines are compared with the **BERT-based alternative meat (BAMK)** model [10]. When applied to the entire dataset, the BAM model outperforms all other models in terms of macro F1 score and precision. The result of the sentiment analysis of 41,782 linked posts is that 42.10% of the posts express negative, neutral, and positive attitudes, in the proportions of 28.77%, 22.9%, and 48.32%, respectively, although previous surveys indicate that the majority of Chinese people have positive attitudes toward AM and only a few have negative attitudes. People's tendency to try or buy AM is influenced by several variables, including gender, location, price, vegetarianism, and food safety. For the first time, conspiracy theories were identified as a major reason for Chinese consumers' rejection of AM. This strategy adds to the growing evidence suggesting a link between these parameters.

When extracting data about organized events from unstructured text, event detectors identify triggers and classify these events based on predefined event types [9]. This process is critical for extracting data about organized events. Because of the multi-hop process of Dynamic Memory Network, the model proposed by Deng et al. [11] is more robust and is capable of obtaining contextual information from a large number of event references. However, vanilla networks gather

event mentions only once and compute event prototypes by simply using the arithmetic mean. Compared to a collection of baseline models, the network performs more consistently with a relatively broad range of event types and is more tolerant of sample scarcity. The network is capable of accommodating the limited number of samples due to the low number of occurrences.

Social context and distribution-based techniques could be an alternative or complement to content-based techniques. The study aims to construct a model that can detect veracity in Arabic news [14]. The method introduces a deep neural network strategy that uses "Convolutional Neuron Networks" to classify bogus and true news claims. This technique solves the problem from the fact-checking point of view by predicting whether an assertion in a news text is legitimate or fraudulent. The method uses a balanced Arabic corpus, because it combines point-of-view identification, point-of-view inference, document retrieval, and fact-checking. The fact-checking ability to detect fake Arabic news has been extensively evaluated. The model outperforms the state of the art with 91% accuracy on the same Arabic dataset.

Sentiment Analysis (SA) is concerned with the computational processing of text sentiments, viewpoints, and subjectivity on Web-based data in Indian languages, including Hindi, Marathi, Kannada, Tamil, and so on. It is important to study these data and extract useful information [15]. Since Hindi is the native language of most Indians, SA in Hindi has become a priority for businesses and government agencies. This article's main contribution is categorizing different articles based on SA methods in Hindi. Most approaches used lexicon-based, machine learning, deep learning, and hybrid methods. These methods influence SA problems, analytical levels, and performance evaluation metrics. This study helps researchers access annotated datasets and linguistic and lexical resources. Regional language research is limited by a lack of resources. Existing large datasets should be expanded. In addition, researchers should make their resources available for the future development of SA. Researchers should also provide new, optimized lexical and linguistic resources to build a strong SA system.

As a result of its numerous applications, sentiment analysis has gained prominence in recent years in conversations, including sentiment analysis, recommender systems, and human–robot interactions. Due to the fact that context can influence an utterance's sentiment, conversational sentiment analysis differs from single-sentence sentiment analysis. Due to the difficulty of conveying context via speech, existing methods utilize deep learning to differentiate conversation participants and model context [16]. In this method, party-ignorant sentiment analysis of the conversational text can be performed in a rapid, concise, parameter-effective manner. A neural tensor block and a two-channel classifier are responsible for context compositionality and sentiment classification in this methodology, respectively. On three example datasets, the technique outperforms current best practices in terms of performance.

The related summary of the literature review is mentioned in Table 1. In this article, we have found that the chance of a phrase being associated with a keyword increases with the frequency of the term in a sentence-level text representation. Thus, our study is unique in that we evaluate both implicit and explicit ironic expressions. This is done using different text material parts with different information values. We weigh the three sentence parts differently depending on their individual meanings. Active learning helps machine learning mine unlabeled data for labelling. Labelling the most valuable data points improves the performance of the model and reduces the cost of labelling. This study found that data points should be collected near the model's feature space (such as comparable related speech and conscious temporal text). This study provides a new acquisition tool that searches unlabeled data for unique examples. **Contrastive active learning (CAL)** selects unlabeled data points whose prediction probability differs from their neighbors in the training set. Instead of clusters, CAL generates communities from the feature space. CAL ranks unlabeled data by accuracy.

Table 1. Related Methods Summary

| Paper	Data-set	Method	Unsupervised	Adaptive learning	Performance
[17]	Online product	Statistical machine learning	No	No	0.87
[22]	Online review	Attention networks	No	No	0.80
[10]	Sentiment analysis | online review	BERT attention network	No	No	0.89
[9]	English news article (2017–2019)	Embedding Deep learning Attention	Yes	No	79.94
Proposed	Irony detection	Lexicon Expansion with Deep active learning	Semi-Supervised	Yes	0.61
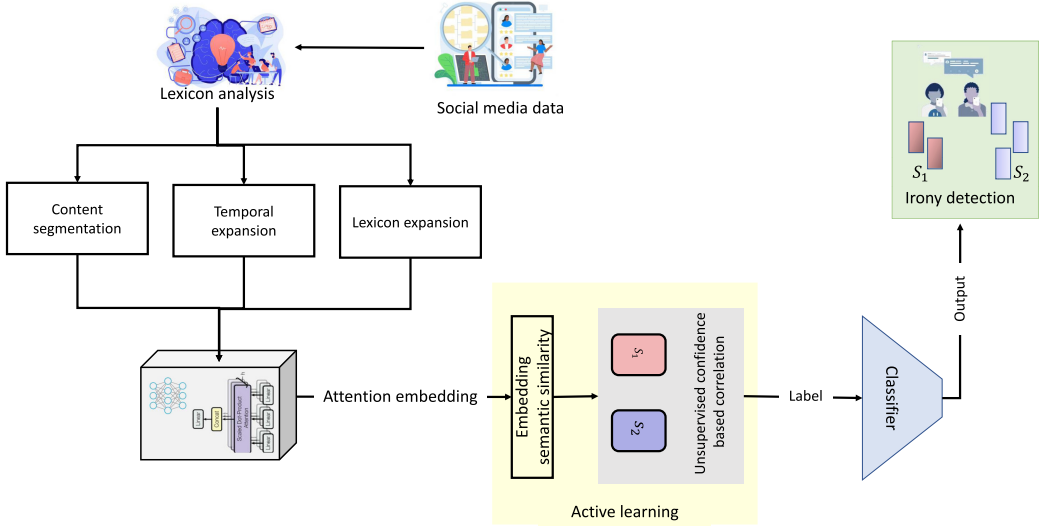
Fig. 1. A workflow of the proposed approach.

3 METHODOLOGY

Figure 1 explains in detail how the irony detection process works. This section explains the steps in preprocessing the documents, expanding the lexicon, and profiling the content segmentation. We also elaborate on obtaining the dataset, annotating the data, and expanding the lexicon knowledge for the co-training model.

In the first step, lexical and linguistic analyzes are performed, as shown in Figure 1. In the lexical analysis, the syntactic structure of a string is analyzed to determine its word components and its semantic words, prefixes, suffixes, phrases, numeric constants, and punctuation marks. Then the semantic words are analyzed to determine linguistic relations, such as part-of-speech (verb, noun, adjective), and so on. After performing a (syntactic and linguistic) lexicon analysis, we can extract all possible concepts from the input text, as mentioned in Section 3.1. Each concept has one or more semantic words associated with it, and each semantic word has a particular semantic relationship with the concept. For a discussion of this relationship, see the following section on content segmentation (Section 3.2), temporal profiling, expansion (Section 3.3), and attention network creation (Section 3.4). An intermediate concept lattice is formed by extracting a set of extracted concepts.

3.1 Lexicon Analysis

The part-of-speech tagging system identifies nouns, verbs, adjectives, and adverbs that contain the tagged part-of-speech as part of words. Using WordNet, we extracted synonyms, hyponyms, morphemes, and physical meanings from a text corpus. Using this method, temporal keywords can

be identified directly for each paragraph, and the terms used to train the model are then included as part of the vocabulary construction process. A trained model (pre-trained network) can be used to convert lexicons (cosine similarity) to vector format (the dimension of the phrase vector). We use trained-to-embed to convert texts into vectors based on semantic awareness by computing whether the focus time and the quarry creation time are related by computing the similarity between them. We propose a hybrid model where the provided text is first modelled as an unsupervised embedding for both text and extended text (for both classes). Word sequences are generated depending on the size of the nearby window. Each embedding generates a feature vector that can classify the data based on the feature vector. An embedding is created based on a class mapping, which is then used to measure the degree of similarity between the class extension and the forum content.

3.2 Content Segmentation

Using this method, the text is divided into three parts following the number of terms and words included in each section. Each segment has a percentage of words that is different from the total number of words in the text. In the first segment, 50% of the words are contained. In the second segment, 30% are contained. Furthermore, in the third segment, 20% are contained. As a result of this division, it is evident that the first portion of the paragraph is the most crucial, as it provides the most helpful information regarding irony as a whole. As you will see at the beginning of the article, this section has been crafted to be highly lengthy to minimize the chances that you will miss important details. We have decided to maintain the original length of the third section. However, it is shorter, since the final section of a book does not generally contain a great deal of background material. Before dividing the irony into three parts, it is essential to break it down into sentences. It is estimated that 40% of the entire phrases are contained in the first segment, followed by segments that contain 40% and 20%, respectively.

3.3 Temporal Profiling and Expansion

Inverted pyramids are used to classify media pieces based on certain criteria into three distinct groups. According to Figure 1, the first segment focuses on the questions "what," "when," and "where," in addition to "who." For the purpose of obtaining information that can be used for the following inquiries, it is first necessary to search for terms that could be translated [18]. Which individuals or entities are involved in the detection? When will irony occur? Where will it occur? What is real irony? And to whom does it pertain? Because the context reveals the ironic subject matter, we extract keywords from it, which answer the first question, "what."

3.4 Temporal Focus Time Attention Network

The model is trained using a limited number of labelled samples, which are then applied to unlabeled data for training. Using both labelled and unlabeled data in this method reduces human annotation and data preparation costs and time. It is possible to synthesize multi-modal data at several levels to make predictions using theory-based machine learning. Furthermore, probability formulations can be used to stimulate evidence collection for dynamic model improvement in addition to quantifying prediction uncertainty. Figure 1 illustrates the possibility of incorporating data into predictive models that identify emerging threats and provide individualized prevention strategies with the advent of the Digital Twin era. Humans can recognize and utilize both form features and contextual relationships as a species. In studies, the ability of machine learning to utilize contextual information inaccessible to humans may account for the fact that it performs similarly or better and has lower agreement than human readers. It is possible to overcome human biases and errors through machine learning that is both robust and reproducible.
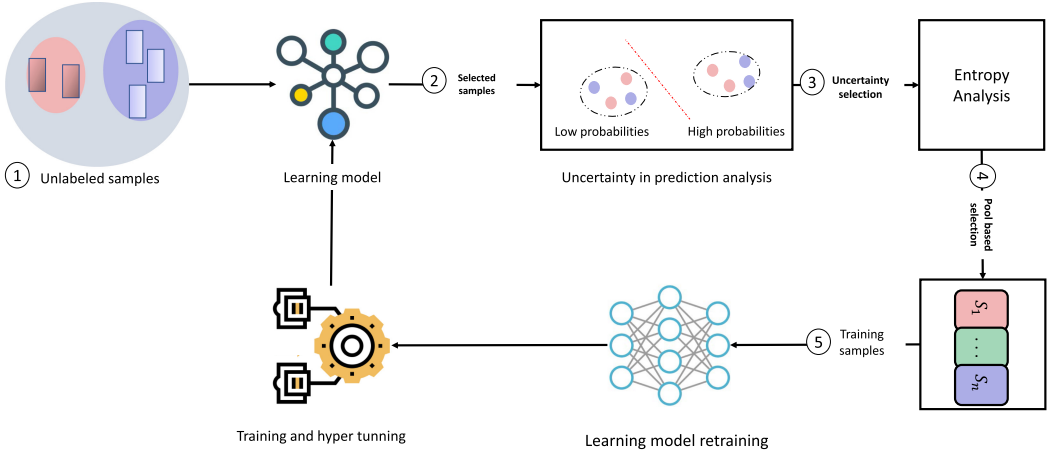
Fig. 2. Active learning process.

3.5 Pool-based Uncertainty Active Learning

By using an entropy-based active learning method, the learner is uncertain about the correct labelling of instances, so the pre-train model can determine which instances to label using glove embedding (e.g., transfer learning). Uncertainty helps minimize the learner's expected entropy and maximize the information the learner is expected to gain. Using the entropy of distribution over labels is a standard method of analysis. Entropy is highest when the probability is uniform (i.e., the learner is maximally uncertain about the correct label) and lowest when the probability is concentrated on a single label (i.e., the learner is certain of the correct label). As a result, the selected samples show the distribution that would be predicted if the learner had perfect knowledge about the instance. Entropy is highest when the predicted distribution is uniform and lowest when the predicted distribution is centered on a single label. There are many other ways to measure learner uncertainty. A learner's behaviour is influenced by the choice of uncertainty measure. The goal is to minimize the expected entropy of the posterior distribution. Therefore, the active learner will tend to select instances that are difficult to label (i.e., instances where the posterior distribution is nearly uniform).

An active learning model envisions a learner interacting with the environment to obtain feedback that can inform future decisions. As shown in the semantic similarity embedding in Figures 1 and 2, the proposed technique first trains a classifier (model) for each tagged dataset using a temporally pre-trained neural network [2] classifier. The proposed technique creates a tagged dataset from unlabeled inputs over multiple rounds. In pool-based selection, a subset of instances is selected from a larger pool of instances for learning. The instances can be generated based on a learned model prediction (e.g., by sampling from a distribution). Empirical analysis has shown that this method can be effective in situations where there is limited labelled data available and where the learner uses a heuristic to select instances that it assumes are informative. Pool-based selection allows a small number of instances to be selected for labelling to ensure a high degree of accuracy in training a model. As a result, we were able to process a large number of classes with high-dimensional data.

In the uncertainty sampling method, instances are selected as expected to provide the most informative information for learning. In representative sampling, a diverse set of instances is selected to ensure that all classes are represented. The core set selection method focuses on selecting a small

number of instances that contain sufficient learning information. Depending on the specific data and task, different pool-based selection methods may be more or less effective. The uncertainty-based active learning approach is powerful, because it can be used with any supervised learning algorithm. Moreover, it can be used with any uncertainty measure, which gives great flexibility to the active learner.

3.6 Attention Network

As the name implies, attention network embedding is a type of neural network architecture that maps input data points to output data points. After lexicon analysis is completed, the network is used to train the embedding. This dataset consists of labelled and unlabeled data points and the attention network is trained with input data points and unlabeled data points [3]. Finally, the attention network can convert input data points into output data points. To calculate the loss function, the squared errors of the assigned and output data points must be added. Attention is given to a text based on recognizing its meaning in a particular context. An attention approach was created by adding a layer of LSTMs to the network [2]. For supervised learning to be effective, a large amount of labelled data is required. This function can extract practical terms from the dropout layer to analyze a result. Using transfer learning to expand the field of lexical analysis and label the dataset, attention output vectors were used as inputs to construct a dropout layer in this study. Using a previously described method that is mentioned again here, this lexicon was constructed. An attention-based contrast phrase was used to map continuous vectors to their corresponding labels. We used attention weights from each tagged dataset to categorize temporal focus according to attention. The extraction, detection, and classification processes utilize attention-learning techniques to create contrast sets. Using the cosine similarity method, we mapped unsupervised semantic similarity to irony and non-irony labels.

A positional attention network is used by Ahmed et al. [3]. Text tensors have an embedding dimension of 300 for each size applied to them. The term "size" refers to the number of words processed simultaneously, and the matrix has the same number of columns regardless of the word. The size of the output is identical to that of the max-pooling layer. This layer applies the Max operation to all the outputs generated by the different filters with the same size. As a result, the Max procedure ensures that we have captured the most significant part of the sentence or document through a single feature that contains all outputs. A major benefit of merging resources is that the number of parameters or weights has been drastically reduced, thereby reducing computation costs and reducing overfitting. To create the final feature map, the feature vectors individually derived from the text tensors are concatenated together. A map containing the most prominent and significant features was created after the extraction of feature vectors. Four dense layers were then employed to reduce dimensionality uniformly. The first and second thick layers each have their own set of neurons, followed by a third dense layer containing 64 neurons. Neurons are present in the last dense layer, so the binary classification can be supported by them (i.e., Irony and Non-Irony). The probability of failure is halved in each of the dense layers. The binary classification is performed by the ReLU activation function, which is supported by the first three dense layers. The last dense layer, however, is supported by the softmax function.

4 EXPERIMENTAL RESULT AND ANALYSIS

A text with ironic content is preprocessed as part of creating a lexicon expansion. Then, various neural networks are trained, and we chose the Glove network for transfer learning. As a result of incorporating a new lexicon into the network, the learned embedding can now support a more excellent range of lines. In the end, the text model is converted into temporal lexicons, which are then used to label previously unlabeled data with vectors after completing this step. As a result

of the labelled data, several patterns are then trained, and once they have been trained, they are compared. By computing the ROC curve, accuracy, recall, and F-measure, we determine how well each pattern performed, and the Adam optimizer was utilized to reduce the number of training cycles required.

4.1 Dataset

The dataset[1] is custom Urdu ads sent to mobile phones in Pakistan for advertising and custom promotions. The dataset contains 1,021 pieces of textual data and is subsequently labelled by the annotators. The text contains fake information about the gifts and the valuable sweepstakes, i.e., people are excited and pay a small amount to get a prize. The scheme is not valid and is used to hack innocent users. Most ironic messages do not require context to be understood and are based on conveying contradictory meanings. The additional source [4] is used for expansion. For experimentation, we split the original training partition into a training partition and a development partition in the ratio of 80% to 20%.

4.2 Pretrain Embedding

There have been numerous strategies for detecting irony in the literature on natural language processing. Recently, knowledge-based systems have gained considerable attention because of their ability to detect irony. A lexicon consists of a collection of words, their meanings and context anchors that have been learned. Affective knowledge is composed of words that can provide context for the user. Our goal is to develop an embedding method that prioritizes input from online discussion forums and utilizes words from a contextually appropriate lexicon (based on their meaning). The Glove is a pre-trained 300-dimensional global vector model for word representation [20]. Each unique word token is embedded within the text's word embedding. To project context into the vector space, we use glove-based vector embedding. Once the embedding has been restored, the semantic structure of the text may be preserved, since the embedded component presents the newly discovered sentence structure [8].

LSTM networks are commonly used to predict time-series data. A bidirectional LSTM network, the Bi-LSTM network, is used to estimate and reduce noise [2, 3]. Forward LSTM networks are used for learning from previous data in the upward direction, while backward LSTM networks are used for learning from future values in the downward direction, where the inputs are the weights of the LSTM cell gates and the outputs are forward or backward. The output gate contains step information in both directions.

4.3 Performance Metrics

It can be seen from the trend map that the performance of the model improves when the training, development, and test sets approach the upper left corner of the accuracy curve. To assess the performance of the model, the ROC curve, precision, recall, and F-measure were calculated. When comparing multiple classifiers, it may be advantageous to integrate each classifier's performance into a single statistic. Calculating the area under the ROC curve, also known as the AUC, is standard practice [2, 3]. It is comparable to the probability that a randomly selected positive case would have a higher score than a randomly selected negative case or the Wilcoxon rank sum statistic for two samples. High AUC classifiers sometimes perform worse in a given scenario than lower AUC classifiers. In practice, however, the AUC proves to be a good indicator of predictive accuracy. Because the number of occurrences for each symptom is different, we used macro and micro techniques. The difference between macro- and micro-averaging is that each class is weighted

---

[1]https://home.mycloud.com/action/share/fca47920-3a3f-48ed-9d78-772dfe1e9a98.

(a) Precision recall curve analysis

(b) ROC curve analysis.

(c) Confusion Matrix analysis.
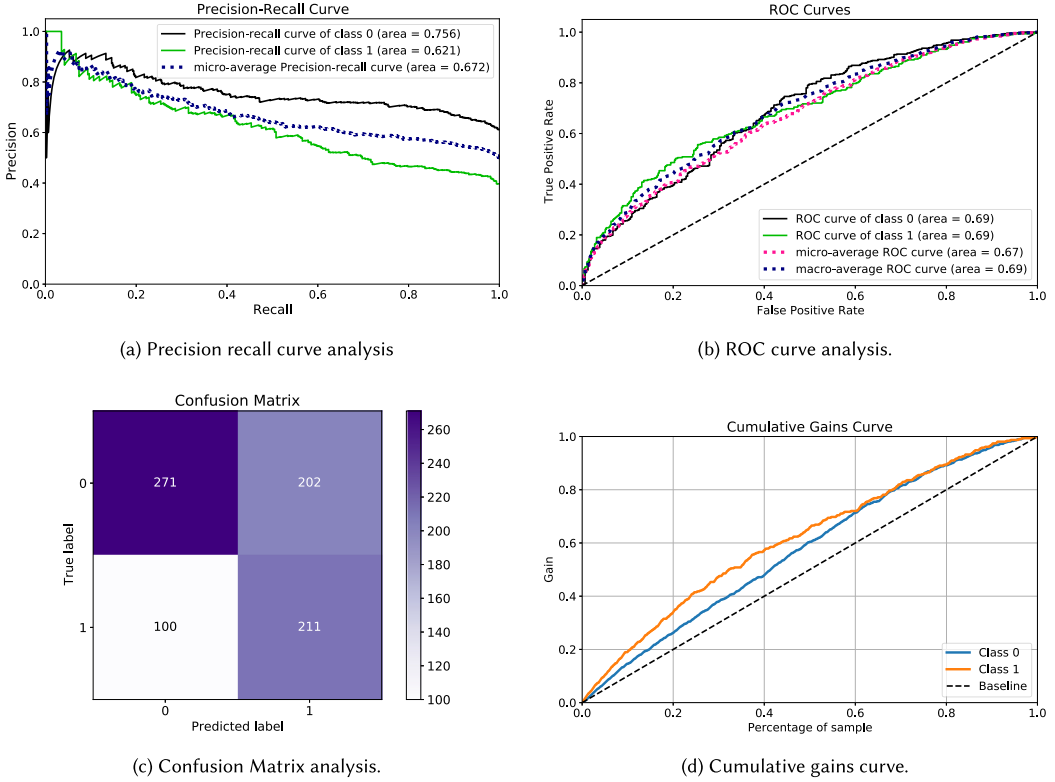
(d) Cumulative gains curve.

Fig. 3. LSTM model analysis under different metrics.

equally in macro-averaging, whereas each sample is weighted equally in micro-averaging. The macro and micro scores are similar when each class contains the same number of samples.

4.4 Evalution

The best F1 score we can achieve on the test set is 73% when this model is re-implemented and applied to the pre-processed data, as shown in Figure 3. We may be able to attribute the huge performance discrepancy between the models to differences in data preparation, since the models use the same partitioning strategy for the data. To test if a dataset model works, you need identical data that has been pre-processed. The hyperparameters for the models used in this study were adjusted based on the development datasets, giving us the following values: one layer for the BiLSTM model with 128 hidden units in the layers; 128 dimensions for the hidden vectors; 128 hidden units for the layers of all feed-forward networks; and an adjusted learning rate for Adam optimizers.

For each size applied to the text tensors, an embedding dimension of 300 was specified. The matrix has the same number of columns regardless of the word, and the term "size" refers to the number of words processed simultaneously. The result is exactly the same size as the max-pooling layer. All the outputs produced by the different filters that had the same size are subjected to the Max operation in this layer. A feature that includes all outputs is the final result. The max operation ensures that we have captured the sentence or feature of the document that is most important to us. The primary benefits of resource sharing are as follows. First, using fewer parameters or weights drastically reduces computing costs and reduces overfitting. Second, a final feature map is created

by concatenating the feature vectors generated independently from the text tensors, which contain the most relevant and important features. Finally, following the extraction of all feature vectors, a uniform dimension reduction was achieved by employing four dense layers.

Filters are first applied to the text portion, as the content of online exchanges determines certain emotional keywords. Then, based on features from both branches corresponding to the same text modality, feature vectors are concatenated. Last, numerous dense layers with a gradually decreasing number of neurons are added for smooth dimensionality reduction to converge the system to a multiple-label classification system. In this way, it is possible to obtain all the different features needed for categorization. By using labelled and unlabeled corpora, optimal feature engineering can be achieved using a semi-supervised neural network and a self-assembling architecture. The final configuration of the neural network was achieved by learning the dimensions of the layers.

The growth of the lexicon determines the effectiveness of a content segmentation strategy. According to this, irony detection occurring within a shorter time period than the time of query has fewer wrong years because of the effect of distance between irony detection and query. Significant variables include detection frequency and persistence over time. By using a scoring method that assigns each message a temporal score, messages are ordered in descending order based on their temporal expressions. To determine a document's attention time, the first two temporal expressions are examined; the higher the score, the higher the ranking position of the expression. Based on the separation of the documents, the attention time of a message document can be evaluated differently. In contrast, the time between enquiry and irony detection does not affect accuracy ratings. The mean number of years of errors is the second type of evaluation used.

Figure 3 and 4, illustrates how time has passed despite the fact that the trigger words in these phrases are directly associated with a wide variety of terms (including irrelevant terms). By assigning higher relevance values to the most informative terms, temporal lexicons, however, are characterized by variety and consistency. As a result, vector representation is able to learn more effectively, even when the trigger words are closely connected to numerous topics. The aggregation strategy of representation can uncover additional true vectors of information that are hidden within temporal lexicons by assigning higher relevance ratings to the most informative terms.

Figures 5 and 6 illustrate that lexicon expansions have the highest score, and any statement that contains all relevant expressions and trigger words is likely to indicate that this occurrence occurs. We investigate the correctness of the automatically generated labelled data, which is anticipated to play a significant role in this event, first and foremost. In this case, the temporal expression of this sentence seems to be appropriate. A dataset of 500 instances was selected from the collection, which was automatically tagged, that included highlighted triggers, labelled arguments, and corresponding event types and roles. Each instance has a trigger, labelled arguments, and the appropriate event types and roles.

A dataset is enhanced by adding autonomously generated and tagged data. Using these enhanced training sets, we evaluate whether the event extractor has improved its performance. Using the temporal expansion technique, distracting verbal triggers were filtered out, and nominal triggers were expanded, demonstrating the effectiveness of the method. In light of the fact that event identification results are improved when the feature set is used, it may also be beneficial to use the feature set to identify temporal information with expressive neural techniques. The significant differences between the arguments indicate that the features defined to extend the argument language are beneficial.

The cycles of semi-supervised learning are depicted in Figure 6. Because of the limited space and a wide variety of forms, which frequently consist of lengthy sentences containing jargon, it is challenging to identify irony. As a result, the attention technique enhances the model results, which utilize the predictive potential of word placements. Moreover, the addition of lexicon-based

(a) Precision recall curve analysis.

(b) ROC curve analysis.

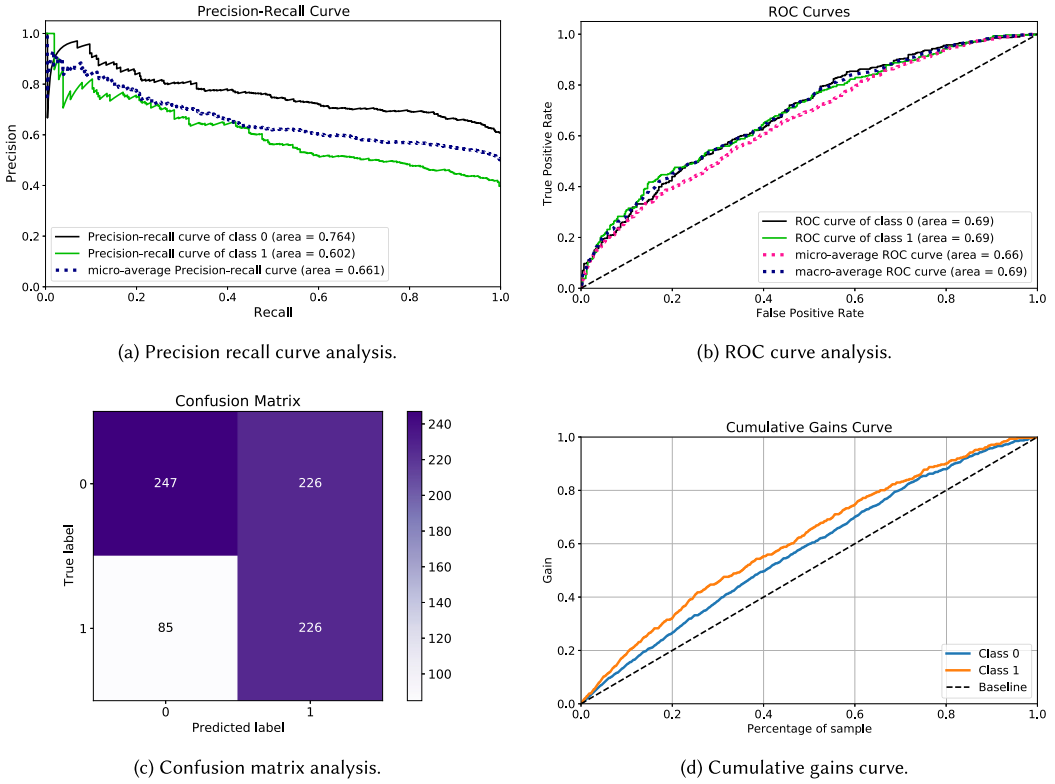(c) Confusion matrix analysis.

(d) Cumulative gains curve.

Fig. 4. BI-LSTM model analysis under different metrics.

neighbours improves the model's overall utility. To discover unlabeled occurrences near the decision border when applying the attention technique, uncertainty sampling can be utilized. The level of confidence in the model is used as a criterion of categorization as it can be used to identify unlabeled occurrences.

4.5 Active Method Comparison

We compare the model with different sampling methods, as shown in Figure 7. Contrast set-based active learning is a method for selecting instances to learn from to produce the most accurate model possible. In this approach, the instances that differ the most from the others in the dataset are selected. This ensures that the model can learn from a variety of different data points, ultimately resulting in a more accurate model. In active learning, entropy is used to determine how well a model can learn from the data. If a model has high entropy, then it is unable to learn from data and is therefore ineffective. Active learning algorithms strive to reduce entropy to improve the learning process. In pool-based active learning, the learner is given a set of instances (the pool) and can choose which instances to label. The learner is then trained on the labelled instances and can repeat the process of selecting and labelling instances until a specific termination criterion is met. Active learning based on pools has many advantages. First, it can help reduce the amount of labelling required, because the learner can focus on the most informative instances. Second, it can help improve the quality of the learning model, as the learner can focus on the most important instances for learning. Finally, it can help reduce the cost of labelling, since the learner

(a) Precision recall curve analysis.

(b) ROC curve analysis.

(c) Confusion matrix analysis.
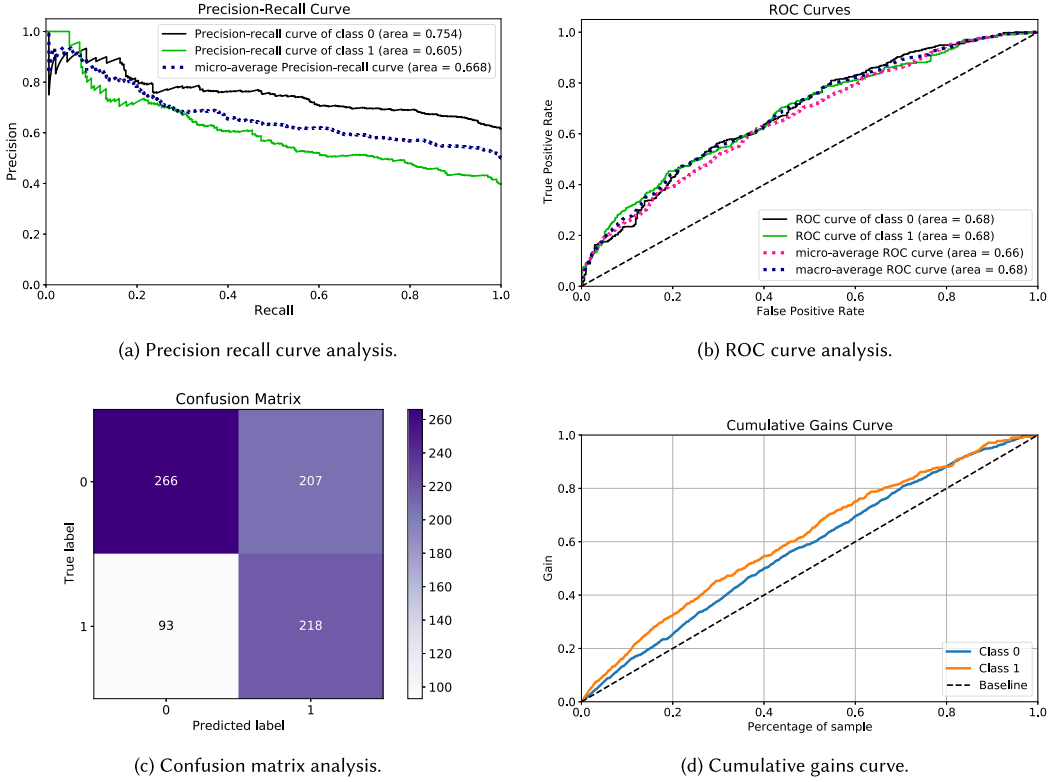
(d) Cumulative gains curve.

Fig. 5. BI-LSTM-Attention model analysis under different metrics.

can choose to label only the most important instances. Random sampling is a method of active learning where the learner is given a set of data points to learn from. The learner is not given any information about the target function or distribution. The goal is to learn as much as possible about data so that the learner can generalize to new data. This method is often used in reinforcement learning and unsupervised learning. The least confidence active machine learning algorithm is designed to select the instances to be labelled by querying the user/oracle. The idea is to select the instances where the model has the least confidence. In other words, this algorithm is used to find the instances where the model is most uncertain. However, it is important to point out that the least confident active learning algorithm is not without its weaknesses. One of the main drawbacks is that it can be prone to error. This is because the algorithm relies on the model being accurate in its predictions. If the model is not accurate, then the algorithm cannot select the correct instances.

A contrast set and entropy-based model are powerful compared to other techniques for several reasons. For accurate classification, the contrast set and entropy-based model consider the correlation between words (supported by the attention network). Due to semantic expansion, the contrast set and entropy-based model can handle lexicon expansion well. In addition, contrast sets and entropy-based models can handle the uncertainty of the model parameters. In addition, the contrast set and entropy-based models have been shown to be robust to overfitting through validation. They are generally more accurate, efficient, and scalable than other methods and robust to outliers and data noise. The drawback is that active learning requires more time to find the desired

(a) Precision recall curve analysis.

(b) ROC curve analysis.

(c) Confusion matrix analysis.
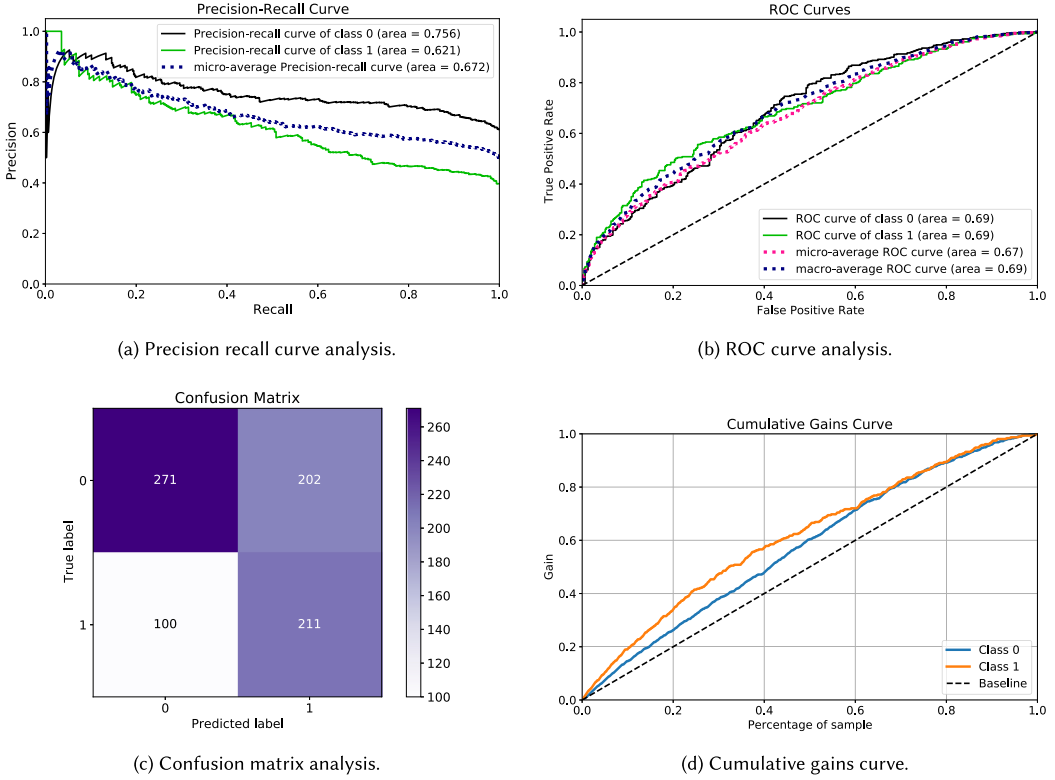
(d) Cumulative gains curve.

Fig. 6. BI-LSTM-Attention with active learning model analysis under different metrics.
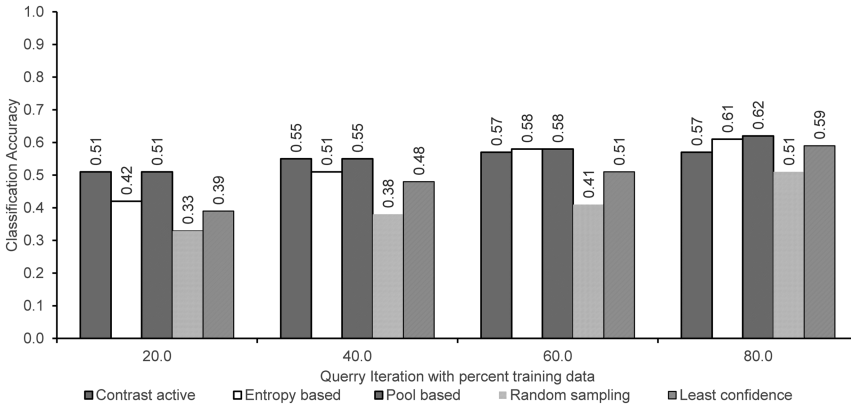

Fig. 7. Model comparison with different sampling collecting methods.

patterns. The above three descriptions of active learning show the process of active learning and its advantages: (1) more selectivity leads to better results because it allows better matching of data and knowledge; (2) an active learner avoids bias due to input data selection; and (3) the improved performance of active learning comes at the cost of increased computation time. It is a data-driven approach that is initiated by the learned model and can be more selective in its use of data, i.e.,

Table 2. Various Strategies Associated with the Model That Has Been Proposed Are Compared

Metrics	LSTM	BILSTM	BILSTM_EX	BILSTM_EX_AL
Accuracy	0.579	0.617	0.615	0.603
ROC AUC Macro Average	0.666	0.682	0.693	0.689
ROC AUC Micro Average	0.643	0.664	0.673	0.66
F1-Score Macro Average	0.578	0.616	0.613	0.603
F1-Score Micro Average	0.579	0.617	0.615	0.603
ROC AUC Irony ADS vs Rest	0.666	0.682	0.693	0.689
ROC AUC Non-Irony ADS vs Rest	0.666	0.682	0.693	0.689
Average Precision Irony ADS vs Rest	0.749	0.754	0.756	0.764
Average Precision Non-Irony ADS vs Rest	0.594	0.605	0.621	0.602
F1-Score Irony ADS vs Rest	0.596	0.639	0.642	0.614
F1-Score Non-Irony ADS vs Rest	0.561	0.592	0.583	0.592

BILSTM_EX: Bi-direction LSTM with lexicon expansion, and BILSTM_EX: Bi-direction LSTM with Active learning.

"selecting which data to use to solve a task." The advantage of a technique such as active learning is that many problems are simpler the more data that is used. Thus, using a larger dataset usually leads to better results.

The final neural network design was created with the appropriate layer dimensions, as indicated in Table 2. The ensemble predictions of the unknown labels generated by the semi-supervised architecture are as accurate as the actual labels themselves. Consequently, training the model with a minimal number of labelled samples produces acceptable results, and adding more labelled samples does not adversely affect the performance of the model. Studies on the datasets of the proposed system have shown that the model has excellent classification capabilities and can be trained with a small number of labelled samples to produce excellent results. Although the proposed method has demonstrated its superiority on a number of parameters, it still has some drawbacks or weaknesses.

5 CONCLUSION AND FUTURE DIRECTION

Time plays an essential role as a fundamental component of the surrounding physical universe. To effectively address the issue of finding irony-based advertisements, the ironic emphasis on time records is more crucial than ever. The current state of technology has made irony-focused literature no longer relevant. Moreover, since there are no irony questions in the datasets provided, we cannot verify or evaluate the model's ironic conclusions, because there is no irony question in these datasets. We demonstrate how lexicon expansion approaches can improve remote vector retrieval by combining syntactic information with variety and semantic exploration. In addition, these procedures can also be utilized to expand the lexicon's diversity. In the proposed approach, a pre-trained neural network is used to find classifiers (models) for each tagged record, and then high-confidence pseudo-labels are added to improve the accuracy of the data. Labelled data is used to create pseudo-labels, which are then updated in the event a classifier (attention network) correctly predicts a sample's label. After integrating the modified predictions from the classifier, the unlabeled inputs are labelled in successive rounds. The suggested model obtained a high score of 89%, which is noteworthy, since expansion and profiling performed better with semi-supervised learning than with supervised learning.

We intend to apply the proposed methods to the important task of relevance-based ironic commentary in the future. Heterogeneous data can come from a variety of sources, including social media, blogs, and news articles. They can be used to make predictions about the contexts of

the text. Heterogeneous data varies in structure and also in composition. These types of data are often difficult to predict, because there is no clear pattern to follow. In the future, we will combine heterogeneous graph networks that contain nodes of different types and edges that connect them. These networks can be used to model relationships between different data types.

REFERENCES

[1] Usman Ahmed, Humera Liaquat, Luqman Ahmed, and Syed Jawad Hussain. 2019. Suggestion miner at semeval-2019 task 9: Suggestion detection in online forum using word graph. In *Proceedings of the 13th International Workshop on Semantic Evaluation*. 1242–1246.

[2] Usman Ahmed, Jerry Chun-Wei Lin, and Gautam Srivastava. 2022. Multi-aspect deep active attention network for healthcare explainable adoption. *IEEE J. Biomed. Health Inf.* (2022). DOI: 10.1109/JBHI.2022.3204633

[3] Usman Ahmed, Suresh Kumar Mukhiya, Gautam Srivastava, Yngve Lamo, and Jerry Chun-Wei Lin. 2021. Attention-based deep entropy active learning using lexical algorithm for mental health treatment. *Front. Psychol.* 12 (2021), 642347.

[4] Usman Ahmed, Lubna Zafar, Faiza Qayyum, and Muhammad Arshad Islam. 2018. Irony detector at semeval-2018 task 3: Irony detection in english tweets using word graph. In *Proceedings of the 12th International Workshop on Semantic Evaluation*. 581–586.

[5] Vimala Balakrishnan, Kee S. Ng, and Hamid R. Arabnia. 2022. Unravelling social media racial discriminations through a semi-supervised approach. *Telemat. Inf.* 67 (2022), 101752.

[6] Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, and Viviana Patti. 2016. Overview of the evalita 2016 sentiment polarity classification task. In *Proceedings of 3rd Italian Conference on Computational Linguistics (CLiC-it 2016) and 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA'16).*

[7] Jireh Yi-Le Chan, Khean Thye Bea, Steven Mun Hong Leow, Seuk Wai Phoong, and Wai Khuen Cheng. 2022. State of the art: A review of sentiment analysis based on sequential transfer learning. *Artif. Intell. Rev.* (2022), 1–32.

[8] Walter G. Charles. 2000. Contextual correlates of meaning. *Appl. Psycholinguist.* 21, 4 (2000), 505–524.

[9] Yubo Chen, Shulin Liu, Xiang Zhang, Kang Liu, and Jun Zhao. 2017. Automatically labeled data generation for large scale event extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 409–419.

[10] Yuan Chen and Zhisheng Zhang. 2022. Exploring public perceptions on alternative meat in China from social media data using transfer learning method. *Food Qual. Pref.* 98 (2022), 104530.

[11] Shumin Deng, Ningyu Zhang, Jiaojian Kang, Yichi Zhang, Wei Zhang, and Huajun Chen. 2020. Meta-learning with dynamic-memory-based prototypical network for few-shot event detection. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 151–159.

[12] José Ángel González, Lluís-F Hurtado, and Ferran Pla. 2020. Transformer based contextualization of pre-trained word embeddings for irony detection in Twitter. *Inf. Process. Manage.* 57, 4 (2020), 102262.

[13] Roland Greene, Stephen Cushman, Clare Cavanagh, Jahan Ramazani, Paul Rouzer, Harris Feinsod, David Marno, and Alexandra Slessarev. 2012. *The Princeton Encyclopedia of Poetry and Poetics*. Princeton University Press, Princeton, NJ.

[14] Fouzi Harrag and Mohamed Khalil Djahli. 2022. Arabic fake news detection: A fact checking based deep learning approach. *Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 4 (2022), 1–34.

[15] Dhanashree S. Kulkarni and Sunil S. Rodd. 2021. Sentiment analysis in Hindi–A survey on the state-of-the-art techniques. *Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 1 (2021), 1–46.

[16] Wei Li, Wei Shao, Shaoxiong Ji, and Erik Cambria. 2022. BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing* 467 (2022), 73–82.

[17] S. Uma Maheswari and S. S. Dhenakaran. 2022. Analysis of approaches for irony detection in tweets for online products. In *Innovations in Computational Intelligence and Computer Vision*. Springer, 141–151.

[18] Donald Metzler, Rosie Jones, Fuchun Peng, and Ruiqiang Zhang. 2009. Improving search relevance for implicitly temporal queries. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 700–701.

[19] Shuteng Niu, Yongxin Liu, Jian Wang, and Houbing Song. 2020. A decade survey of transfer learning (2010–2020). *IEEE Trans. Artif. Intell.* 1, 2 (2020), 151–166.

[20] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 1532–1543.

[21] Emilio Sulis, Delia Irazú Hernández Farías, Paolo Rosso, Viviana Patti, and Giancarlo Ruffo. 2016. Figurative messages and affect in Twitter: Differences between# irony,# sarcasm and# not. *Knowl.-Bas. Syst.* 108 (2016), 132–143.

[22] Guixian Xu, Zixin Zhang, Ting Zhang, Shaona Yu, Yueting Meng, and Sijin Chen. 2022. Aspect-level sentiment classification based on attention-BiLSTM model and transfer learning. *Knowl.-Bas. Syst.* 245 (2022), 108586.
[23] Shiwei Zhang, Xiuzhen Zhang, Jeffrey Chan, and Paolo Rosso. 2019. Irony detection via sentiment-based transfer learning. *Inf. Process. Manage.* 56, 5 (2019), 1633–1644.