Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/22/20231224

Challenges and strategies for sentiment analysis of irony and humor in social media based on machine learning

Zijun Li

School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2GN, United Kingdom

vergil_lzj@hotmail.com

Abstract. Sentiment analysis of irony and humor in social media poses a formidable challenge owing to the complexity and context-dependency of such expressions. This report offers a comprehensive overview of diverse machine learning techniques utilized in the analysis of irony and humor, encompassing traditional machine learning algorithms and deep learning approaches. Moreover, this review scrutinizes the challenges and future prospects for the sentiment analysis of irony and humor in social media. Future research pathways involve cross-lingual and cross-cultural analysis, multimodal information integration, autonomous identification of novel patterns, adversarial training, and augmenting the explainability and interpretability of sentiment analysis models. The report highlights the importance of these challenges and potential directions, unveiling their impact on this advancing field of research.

Keywords: machine learning, irony, humor, sentiment analysis.

1. Introduction

The rapidly growing popularity of social media platforms has necessitated understanding user sentiment through text for various applications, including marketing, customer service, and opinion mining. Irony and humor, prevalent in social media, pose significant challenges for sentiment analysis due to their context-dependency, diverse forms, and intricate language features. As science and technology advance, artificial intelligence increasingly permeates specialized domains of everyday life. The Verta Insights Study reveals that among investment strategies across six distinct spending categories for 2022 and 2023, the AI innovation technology category remains the top priority, with 54% and 58% of respondents ranking it as such, respectively [1]. This underscores the importance of artificial intelligence technology in social development. Given the rising significance of artificial intelligence and the challenges presented by irony and humor in sentiment analysis, this review aims to offer a comprehensive overview of methods and strategies to address the sentiment analysis of irony and humor in social media using machine learning techniques.

This review initiates with an in-depth explanation of irony and humor within the social media landscape, followed by an exploration of various machine learning techniques, along with their applications, benefits, and drawbacks. The discussion incorporates both conventional algorithms, including Support Vector Machines, Decision Trees, Random Forests, Naive Bayes Classifier, and modern deep learning ones, including Convolutional Neural Networks, Recurrent Neural Networks,

DOI: 10.54254/2755-2721/22/20231224

and Transformers. It concludes by addressing the challenges and future of sentiment analysis for irony and humor, emphasizing the necessity of ongoing advancements in methodologies for enhanced sentiment analysis in social media text. This research serves as a beneficial reference for both practitioners and academics in the realm of social media analytics and sentiment analysis, thereby promoting informed decision-making and efficacious strategy formulation.

2. The basic concepts of irony and humor sentiment analysis

2.1. The concepts of Irony

Irony is a rhetorical device that involves a discrepancy between what is expected and what actually occurs, or between the intended meaning and the literal meaning of words. It is commonly used in literature, language, and everyday conversation to create humor, emphasize a point, or express subtle criticism. Among the various types of irony, verbal irony and situational irony are the most common [2-3].

• Verbal irony

Verbal irony occurs when a speaker's intended meaning is opposite or contrary to the literal meaning of their words.

For instance:*"Oh, great! Just what I needed—a flat tire!"*

The speaker's words seem to express enthusiasm, but their actual intention is to convey frustration or disappointment about the flat tire.

• Situational irony

Situational irony arises when there is a discrepancy between the expected outcome of a situation and the actual outcome. This type of irony is often used to create a dramatic effect, add depth to a story, or convey a moral lesson, featuring prominently in literature. *For example: In Shakespeare's Romeo and Juliet, the two lovers plan to elope and live happily ever after. However, their pursuit of happiness ultimately leads to their tragic demise.*

2.2. The concepts of Humor

Humor is a multifaceted and subjective concept, often defined as a form of communication that elicits laughter and entertainment, typically prompted by unexpected or incongruous situations. Humor can be verbal, visual, or physical. This discussion concentrates on verbal humor, which can be further classified into the following categories: self-irony, exaggeration/hyperbole, phonetics assisted, semantic opposites, and secondary meaning [2].

2.2.1. Self-irony. Referring to self-irony, frequently employed to defuse tense situations or alleviate embar- rassment.

For example: *"I hate it when I go to hug someone really sexy and my face smashes right into the mirror. "*

The use of self-irony in this expression indicate that the "someone sexy" is the speaker.

2.2.2. Exaggeration/Hyperbole. This category use extreme exaggeration to emphasize a point or add humor. Example:

"I'm so hungry, I could eat a horse!"

Normally, a human cannot eat an entire horse. This extreme exaggeration is used to convey the speaker's intense hunger.

2.2.3. Phonetics Assisted. Phonetics Assisted refers to humor generated by the manipulation of phonetic and pronunciation features. The production of this humor usually depends on the syllables, rhythm, and intonation of the language.

Example: *"I'm reading a book on the history of glue – I just can't seem to put it down. "*

DOI: 10.54254/2755-2721/22/20231224

The double meaning of the phrase "put it down" is employed because it can refer to physically placing a book down and stopping reading a book.

2.2.4. Semantic Opposites. Semantic Opposites refers to the use of words with opposite meanings to create a humorous effect.

Example: *"I love sleeping, it's like being dead without the commitment."*

This example underscores the speaker's love for sleeping by using the contrasting idea of death without its finality.

2.2.5. Secondary Meaning. Secondary Meaning is used to create a humorous effect by utilizing a word or phrase with multiple meanings.

Example: *"Why do we tell actors to 'break a leg'? Because every play has a cast."*

The phrase "break a leg" possesses a secondary meaning in the theater world as a way of wishing someone good luck before a performance, even though the literal meaning is negative.

3. Machine learning strategy for sentiment analysis of irony and humor

After comprehending the fundamental classification of irony and humor, it is crucial to collect data sets, preprocess the data, extract features, and train various models to achieve recognition. Machine learning-based sentiment analysis methods are typically classified into traditional machine learning and deep learning approaches. This section explores traditional machine learning methods, such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Naive Bayes Classifier, as well as deep learning models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformers. These techniques are analyzed to determine their efficacy in sentiment analysis tasks related to irony and humor.

3.1. Machine learning algorithms

Machine learning algorithms, a subset of artificial intelligence, enable machines to learn and enhance their performance from experience without explicit programming. These algorithms are extensively employed in sentiment analysis, including the analysis of irony and humor.

Feature engineering is a prevalent approach in sentiment analysis of irony and humor, where relevant features (such as sentiment words, part-of-speech tags, and syntactic patterns) are identified and employed to train the classifier. For instance, detecting the presence of positive or negative sentiment words in the text can aid in classifying humor, while recognizing ironic expressions like understatement or overstatement can be utilized to classify irony.

Another approach involves ensemble methods, which integrate multiple classifiers to enhance the accuracy of the analysis. For example, combining decision trees and support vector machines (SVMs) could be employed to classify text as humorous or non-humorous. Ensemble methods have demonstrated improved performance in sentiment analysis of irony and humor compared to individual classifiers.

3.1.1. Support Vector Machines. Support Vector Machines(SVM) is a supervised learning algorithm commonly used for classification and regression tasks. The core principle of SVM is to divide the dataset into distinct categories by identifying an optimal hyperplane. In the case of binary classification, SVM aims to find a hyperplane during training that maximizes the margin between the two different categories, effectively partitioning the data.

The Decision function is known as

$$f(x) = w^T \emptyset(x) + b \quad (1)$$

- $f(x)$: the output of the decision function, indicating the classification of the input sample x.
- w: the weight vector, which is learned during training and used to make classifications.

DOI: 10.54254/2755-2721/22/20231224

• ∅(x): the feature vector that represents the input sample x. This feature vector is derived from the original data using a transformation function.

• b: the bias term, which shifts the decision boundary away from the origin. It is also learned during training.

When the dataset is linearly separable, SVM identifies a unique hyperplane that separates the data into two classes, as Figure 1 shows. However, if the dataset is not linearly separable, SVM employs a technique called the kernel function to transform the data from the original space to a higher-dimensional space, where the data becomes linearly separable [4]. Figure 2 is a visual demonstration of the process.

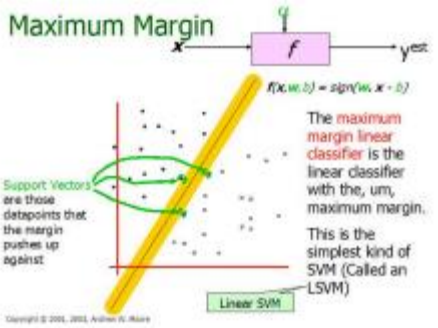

Figure 1. SVM: Linear SVM [7].

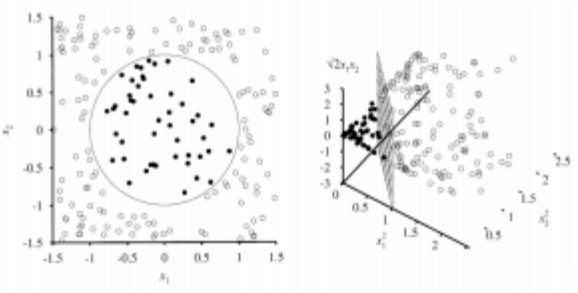

Figure 2. SVM: Mapping data into a high-dimensional space [5].

The selection of appropriate parameters is critical for achieving effective performance and generalization capabilities when training an SVM for sentiment analysis. Different combinations of parameters can have a significant impact on the model's performance.

• Kernel function: The selection of an appropriate kernel function is vital to achieving optimal SVM performance. For datasets that are linearly separable, a linear kernel function typically results in better classification outcomes. However, for nonlinear datasets, polynomial and Radial Basis Function (RBF) kernels are commonly used. Polynomial kernels can capture polynomial relationships, while RBF kernels can handle complex nonlinear relationships. Different kernel functions can be compared using cross-validation during the selection process [4].

• C: C is a regularization parameter in SVM, and its value is an important factor in the objective function, which can be expressed as

$$\min_{w,b} \frac{1}{2}|w|^2 + C\sum_{i=1}^{n}\xi_i \quad (2)$$

A larger value of C implies a greater penalty for misclassification, increasing the risk of overfitting. On the other hand, a smaller value of C indicates a lower penalty for misclassification, which heightens the risk of underfitting. In practice, various values of C can be tested, and the optimal value is typically found using cross-validation on a logarithmic scale [6].

DOI: 10.54254/2755-2721/22/20231224

3.1.2. Decision Trees & Random Forests. A decision tree is a tree-structured classification algorithm that partitions the dataset, with each internal node representing a feature and each leaf node representing a category. The decision tree is built recursively, selecting each node based on a specific characteristic, typically using information gain or the Gini index. Information gain measures a feature's importance for classification, while the Gini index measures the purity of samples.

During the decision tree's classification process, starting from the root node, input samples are categorized according to the characteristics represented by each node, then assigned to corresponding child nodes until reaching the leaf nodes. Each leaf node represents a category, and the sample is ultimately assigned to a specific leaf node, which constitutes the sample's classification result as illustrated in Figure 3 [7-8].

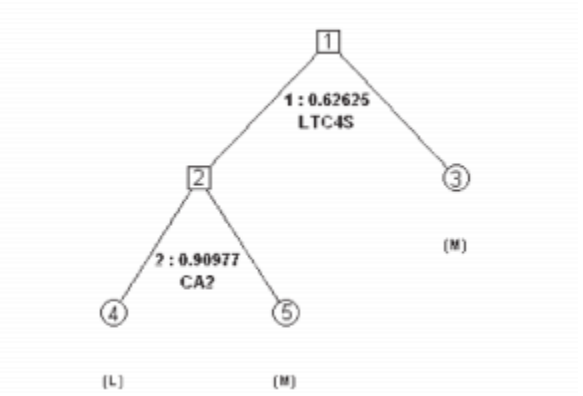

Figure 3. Decision tree: Decision tree corresponding to the partitioned feature space in Figure 3 [8].

In order to partition the feature space into distinct regions, a decision tree leverages a method known as "recursive partitioning," illustrated in Figure 4. This methodology entails iterative testing and partitioning of the data, based on the specific values of the features, and considers a hypothetical scenario encompassing a two-dimensional feature space, represented by features 1 and 2. Initially, the decision tree conducts a test on feature 1 (for instance, when feature 1 is less than a specific threshold), and contingent upon the result, divides the data into unique groups. Subsequently, within each of these just-formed partitions, the algorithm executes a test on feature 2. This iterative process of testing and partitioning persists until a predefined stopping criterion, typically associated with a statistical measure of homogeneity within groups, is satisfied.

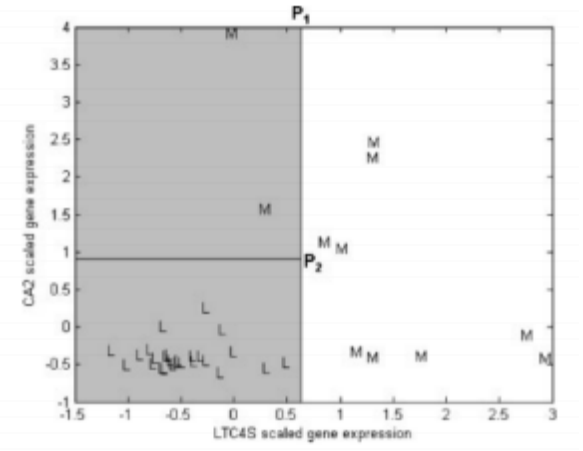

Figure 4. Decision tree: Recursively- partitioned feature space (features 1 and 2) of XT [8].

DOI: 10.54254/2755-2721/22/20231224

To prevent overfitting, the decision tree undergoes continuous pruning during training based on the training data. Pruning is generally divided into two methods: pre-pruning and post- pruning [9]. Pre-pruning limits each node's division during tree construction to avoid overfitting, while post-pruning trims a complete tree after establishment to reduce overfitting.

Random Forests is an ensemble learning method that enhances classification or regression accuracy by constructing multiple decision trees. Each decision tree is obtained by randomly subsampling the training data and training with a random subset of features. At testing time, each decision tree classifies or regresses the input data, and the results from all decision trees are integrated to produce a final prediction.

We use the random forest algorithm to train the training set, generating multiple decision trees. At each decision tree node, we select a feature to classify and divide the dataset into two subsets. This process repeats until specific conditions are met.

3.1.3. Naive Bayes classifier. The Naive Bayes classifier is a machine learning algorithm based on Bayes' theorem, which assumes feature independence, meaning the values of each feature under a given category are independent.

Specifically, the Naive Bayes classifier transforms a text data sentiment classification problem into a conditional probability calculation. Given text data D, the probability $P(c|D)$ of it belonging to a certain sentiment category c must be calculated. According to Bayes' theorem, $P(c|D)$ can be expressed as:

$$P(c|D) = \frac{P(D|c)P(c)}{P(D)} \tag{3}$$

Here,

$P(D|c)$ denotes the probability of text data D occurring under the given emotion category c.

$P(c)$ represents the probability of emotion category c occurring.

$P(D)$ is the probability of text data D occurring. $P(D)$ is identical for all sentiment categories, it can be disregarded [10].

The Naive Bayes classifier presumes that for a given class c, the values of each feature x_1 are independent. Thus, the conditional probability $P(D|c)$ can be expressed as:

$$P(D|c) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \tag{4}$$

Here, $x_1, x_2, \ldots, x_n$, represent the features of text data D, and n is the number of features [10].

Goel et al. propose a method that combines SentiWordNet and Naive Bayes to enhance the accuracy of sen- timent analysis for tweets [11]. SentiWordNet is a sentiment lexicon that assigns sentiment scores to words, reflecting their positivity, negativity, and objectivity. The findings demonstrate that incor- porating SentiWordNet with the Naive Bayes classifier can lead to improved sentiment analysis accuracy.

3.2. Deep learning algorithm

Deep learning algorithms have demonstrated considerable potential in numerous natural language processing tasks, encompassing sentiment analysis of irony and humor.

3.2.1. Convolutional Neural Networks (CNNs). Although CNNs are predominantly utilized in image processing, they can also be employed for text classification tasks, including irony and humor analysis. They operate by applying con- volutional filters to extract features from text and pooling layers to diminish the dimensionality of the extracted features.

The application of CNNs in humor and irony sentiment analysis hinges on convolution and pooling operations. Convolution operations capture local features in the text, while pooling op- erations reduce feature dimensions and enhance feature robustness. CNNs accept text input as a matrix, convolve it

DOI: 10.54254/2755-2721/22/20231224

with multiple convolution kernels, and then use pooling operations for fea- ture extraction and dimensionality reduction. Ultimately, a fully connected layer classifies the features [12].

The training procedure of CNNs unfolds in the sequence illustrated in Figure 5, entailing the following steps [13]:

1. Input layer: The input layer receives the text data for analysis.
2. Convolutional layer: This layer applies a set of filters to the input data to detect specific features, such as word se- quences, patterns, and structures.
3. Pooling layer: The pooling layer down samples the activation layer's output by summarizing the information in neighboring activations.
4. Fully connected layer: This layer takes the pooling layer's output and applies a set of weights to generate a final output.
5. Output layer: The output layer yields a probability distribution over possible sentiment labels (e.g., positive, negative, ironic, humorous).

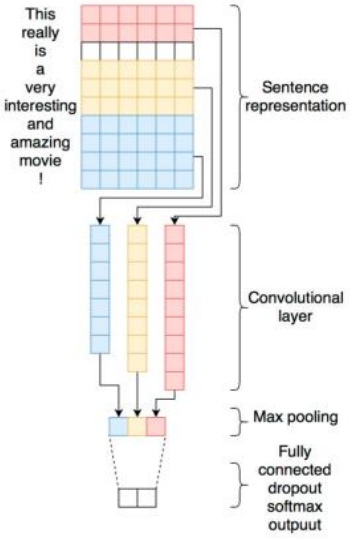

Figure 5. CNN: Example architecture [13].

3.2.2. Recurrent Neural Networks (RNNs). Recurrent Neural Networks (RNNs) constitute another deep learning algorithm suitable for humor and irony sentiment analysis. RNNs excel at analyzing sequential data, crucial for com- prehending the context of humorous or ironic statements [14].

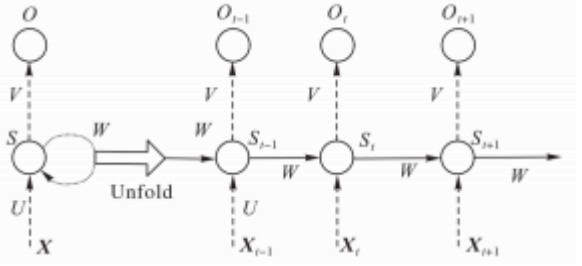

Figure 6. RNN: Example architecture [14].

The fundamental principle of RNNs involves using the output from a previous time step as input to the current time step, enabling the network to maintain a form of memory for previous inputs. This feature is especially benefi- cial for analyzing text sequences, where a word or phrase's meaning may rely on the context of preceding text.

DOI: 10.54254/2755-2721/22/20231224

A significant application of RNNs in humor and irony analysis involves generating new text that is humorous or ironic. By training an RNN on a dataset of humorous or ironic text, the model can learn to generate new text that mirrors the style and tone.

Long Short-Term Memory (LSTM) Long Short-Term Memory (LSTM) represents a specialized RNN model designed for processing sequential data and has gained widespread use in irony and humor sentiment analysis [14].

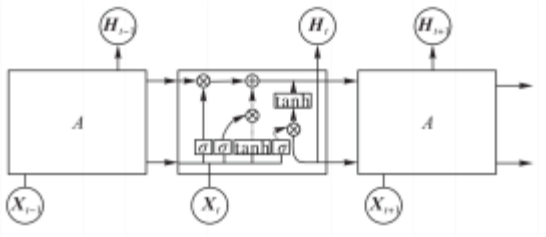

Figure 7. LSTM: Example architecture [14].

In contrast to traditional RNN models, LSTM incorporates three gate mechanisms: input, forget, and output gates. These gates control the influence of information passed from previous moments and current moment in- put on subsequent moments. This mechanism effectively addresses the vanishing gradient problem associated with long sequence data, thereby enhancing the model's performance

In humor and irony sentiment analysis, LSTMs frequently model text sequences to capture contextual and semantic information within language. By learning long-term dependencies, LSTMs gain a better understanding of implicit semantics and logic in humor and irony, leading to improved sentiment analysis.

LSTM models typically require extensive data for training to achieve optimal performance. During the training process, the back propagation algorithm updates the model's parameters, while techniques such as cross-validation evaluate the model's performance and generalization capabilities. Ultimately, the trained model performs sentiment analysis on new text data, classifying and recognizing humor and irony.

In the study conducted by Krupa Patel, Manasi Mathkar, Sarjak Maniar, Avi Mehta, and Shachi Natu, the researchers presented a comprehensive approach for employing an LSTM (Long Short-Term Memory) model to detect humor in text. The model achieved an impressive accuracy of 94.62% [15].

AttBiLSTM (Attention-based Bidirectional Long Short-Term Memory) AttBiLSTM (Attention-based Bidirectional Long Short-Term Memory) extends the LSTM model by incorporating an additional attention mechanism and bidirectionality (Figure 8) [16].

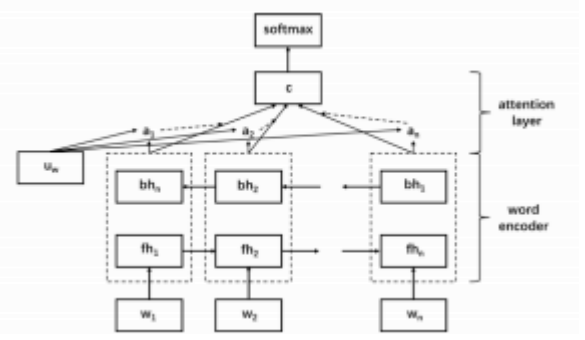

Figure 8. AttBiLSTM: architecture [16].

In AttBiLSTM, a reverse LSTM processes the re- verse part of the sequence at each time step, in addition to the forward LSTM. Furthermore, AttBiLSTM employs an attention mechanism to calculate

DOI: 10.54254/2755-2721/22/20231224

the importance of each time step in the sequence, assigning importance to the corresponding hidden state to better capture key information within the sequence.

Compared to LSTM, AttBiLSTM offers superior sequence modeling and expressive capabili- ties, handles long sequences more effectively, and demonstrates higher performance in numer- ous natural language processing tasks.

3.2.3. Transformers. Transformers represent a neural network model based on the self-attention mechanism, orig- inally proposed by Google. The core concept involves acquiring a comprehensive feature repre- sentation of the input sequence by executing self-attention weighted aggregation on the features of all positions within the input sequence.

Transformers generally comprise two stages: pre-training and fine-tuning. During the pre training stage, the model learns to represent text data from a vast text corpus. In the fine tuning stage, the model undergoes refinement on specific humor and irony datasets, enhancing its sentiment analysis capabilities for humor and irony. Distinct from traditional RNNs and CNNs, Transformers consider the semantics of the entire sentence when processing text, resulting in superior performance for humor and irony that require an understanding of contextual language forms.

4. Challenges and Prospects of Sentiment Analysis for Irony and Humor

The undertaking of sentiment analysis concerning irony and humor in social media presents notable challenges, making it a complex task. The primary challenges lie in:

• Context-dependency: The proper comprehension and interpretation of irony and humor require specific contexts and background knowledge. Without such context, the true senti- ment of a text may be misinterpreted.

• Diversity: Irony and humor appear in diverse forms such as sarcasm, puns, and hyperbole, which complicate the creation of a comprehensive approach for sentiment analysis.

• Language features: Irony and humor involve sophisticated language features like wordplay, ambiguity, and rhetorical devices, which challenge conventional NLP techniques.

• Cross-cultural differences: The variation of irony and humor across cultures adds complex- ity to the development of universally applicable models.

Future research avenues in this area should focus on:

• Cross-cultural and cross-lingual analysis: Designing models that can effectively detect irony and humor across different languages and cultural contexts is a persistent challenge.

• Multimodal information fusion: Future models should efficiently incorporate multimodal features like images and videos to enhance the accuracy of detection.

• Pattern discovery: The development of unsupervised or weakly supervised methods for the automatic discovery of new irony and humor expressions would substantially contribute to the field.

• Adversarial training: This could improve the robustness and generalization of sentiment analysis models for irony and humor.

• Explainability and interpretability: It's vital to enhance the explainability and interpretabil- ity of models to foster trust and comprehension.

5. Conclusion

This review scrutinizes machine learning strategies, including supervised, unsupervised, and deep learning approaches, for analyzing irony and humor in social media sentiment. Though strides have been made, challenges persist due to context-dependency, diversity, language intricacies, and cross-cultural variances. Future promising research could focus on cross-lingual and cultural analysis, multimodal information integration, discovery of novel patterns, adversarial training, and enhancing model interpretability. Addressing these hurdles and exploring these avenues will bolster accuracy and robustness in sentiment analysis of irony and humor.

Nevertheless, this review has its limitations. The primary shortcoming is the lack of practical case studies to illustrate the efficacy of the presented machine learning strategies. Future work should strive

DOI: 10.54254/2755-2721/22/20231224

to supplement theoretical discussions with empirical analyses to demonstrate these strategies' real-world applicability. Moreover, this review did not adequately delve into the challenges of cross-lingual and cross-cultural sentiment analysis. Future studies should focus more extensively on this area, considering the rapidly increasing diversity of social media users worldwide.

References

[1] King, R. (2023). Companies Continue to Push AI/ML Investments. Verta Blog. Retrieved April 23, 2023, from https://blog.verta.ai/2023-ai/ml-investments

[2] Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. Data & Knowledge Engineering, 74, 1-12.

[3] Barbieri, F., & Saggion, H. (2014). Modelling Irony in Twitter (pp. 56–64). Association for Computational Linguistics. https://aclanthology.org/E14-3007.pdf

[4] Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine learning applications based on SVM classification a review. Qubahan Academic Journal, 1(2), 81-90.

[5] Gavrilov, Z. (n.d.). SVM Tutorial. https://web.mit.edu/zoya/www/SVM.pdf

[6] Jordaan, E. M., & Smits, G. F. (2002, May). Estimation of the regularization parameter for support vector regression. In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (Vol. 3, pp. 2192-2197). IEEE.

[7] Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine. In 2016 international conference on computing communication control and automation (ICCUBEA) (pp. 1-7). IEEE.

[8] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275-285.

[9] Esposito, F., Malerba, D., Semeraro, G., & Kay, J. (1997). A comparative analysis of methods for pruning decision trees. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5), 476–493. https://doi.org/10. 1109/34.589207

[10] Kaur, G., & Oberai, E. N. (2014). A review article on Naive Bayes classifier with various smoothing techniques. International Journal of Computer Science and Mobile Computing, 3(10), 864-868.

[11] Goel, A., Gautam, J., & Kumar, S. (2016, October). Real time sentiment analysis of tweets using Naive Bayes. In 2016 2nd International Conference on Next Generation Computing Technologies (NGCT) (pp. 257-261). IEEE.

[12] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8(1). https://doi.org/10.1186/s40537-021-00444-8

[13] Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understand- ing based on sentiment analysis of twitter data. Procedia Computer Science, 111, 376–381. https://doi.org/10.1016/j.procs.2017.06.037

[14] Wang, Y., Zhu, J., Wang, Z., Bai, F., & Gong, J. (2022). Review of applications of natural language processing in text sentiment analysis. Journal of Computer Applications, 42(4), 1011. https://doi.org/10.11772/j.issn.1001-9081.2021071262

[15] Patel, K., Mathkar, M., Maniar, S., Mehta, A., & Natu, S. (2021, July). To laugh or not to laugh–LSTM based humor detection approach. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[16] Li, D., Rzepka, R., Ptaszynski, M., & Araki, K. (2020). HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. Information Processing & Management, 57(6), 102290.