Guest Editorial: Neurosymbolic AI for Sentiment Analysis

Frank Xing, Iti Chaturvedi, Erik Cambria, Amir Hussain, and Björn Schuller

I. Introduction

NEURAL network-based methods, especially deep learning, have been a burgeoning area in AI research and have been successful in tackling the expanding data volume as we move into a digital age. Today, the neural network-based methods are not only used for low-level cognitive tasks, such as recognizing objects and spotting keywords, but they have also been deployed in various industrial information systems to assist high-level decision-making. In natural language processing, there have been two milestones for the past decade: one is word2vec [1], a group of neural models that learn word embeddings (vector representations of words) from large datasets; and one is the most recent GPT-based models [2], which combine reinforcement learning with a generative transformer in order to enable multi-round end-to-end conversations.

While producing highly accurate predictions on datasets and generating human-like utterances, those neural network-based artifacts provide little understanding of the internal features and representations of the data. Many problems and concerns subsequently emerge from this black-box issue. Because some of the problems and concerns are also relevant in the context of sentiment analysis, we list five of them below:

- **Interpretability:** Neural networks, particularly deep learning models, often have complex architectures and millions of parameters. Understanding how these models arrive at their predictions or decisions can be challenging. Although some sentiment analysis applications are result-driven, there are other critical and sensitive applications such as finance, healthcare, or legal domains, where explanations and justifications are essential. For instance, traders would like to first know what are the triggering entities and events, and why they cause market sentiment to fluctuate [3]; doctors and medical staff analyze sentiment from health social media, but the ultimate purpose is to leverage the sentiment to discover medication non-adherence reasons [4]. These tasks need more goal-driven neural architecture design, which transcends the domain of neural network and machine learning. Although much effort has been devoted to opening the black-box of neural networks, e.g., sensitivity analysis, the interpretability problem generally worsens as the model complexity grows.

- **Bias from annotations:** Neural network models typically require large amounts of labeled data for training. Gathering and annotating such datasets can be time-consuming, expensive, or even impractical for certain languages or specialized domains. Annotation practices for sentiment analysis, e.g., [5], [6], have shown that there is only limited degree of agreement for some ambiguous expressions. This data reliance can also introduce biases present in the training data, which is more difficult to find than in a structured knowledge base. These biases would finally lead to erroneous predictions and understanding of individual's sentiment.

- **Diagnozability:** Sentiment analysis models trained from open data usually need adaptation before being deployed to a specific use environment. During this phase, multiple rounds of tests are conducted and feedback is gathered. This feedback, even clearly sorted as in [7] or [8], are difficult to fix. In fact, neural networks can struggle with generalizing to data that differs significantly from their pre-training distribution. When dealing with the feedback, such as rare words, novel sentence structures, or domain-specific jargon, neural networks are constrained with the format of "instance-label", and the outcomes of such fixes are unreliable.

- **Lack of commonsense:** Neural network-based models learn from patterns, but often lack explicit knowledge of common sense and world knowledge. For example, to understand the sentiment in the sentence "The movie had stunning cinematography, but the plot fell flat.", one needs to know a commonsense that movie is expected to be dramatic and "fell flat" carries a negative sentiment toward its plot. Without commonsense, the sentence becomes a pure description and its link to the elicited-sentiment relies on commonsense. Neural network-based models learn from the various patterns associated with this sentence, after which, some overfitted patterns may be generalized to other sentences, causing hallucinated sentiment from objective descriptions.

- **Cost:** Sentiment analysis in areas such as customer relationship management works on a great amount of data. When the focus is a final statistical result, the speed and cost of sentiment analysis on a single piece of text becomes important. It is well known that modern large language models are expensive to train and run (e.g., training GPT-3 and GPT-4 involves tens of thousands of GPUs running continuously for months [9], and are estimated to cost over 4 million USD; when running, its API is priced at around 0.1 USD per thousand words). Large-scale or real-time sentiment analysis jobs with time/cost constraints, therefore, becomes challenging.

The potentially broad societal impacts of neural network-based methods alert people to a dystopian future and re-ignite research on neurosymbolic AI (also known as hybrid AI): a key idea to mitigate unexpected model behavior and inject interpretability by combining learnable parameters (neuro) with predefined knowledge bases or templates (symbolic). In traditional symbolic AI, knowledge and reasoning are represented using explicit symbols and rules, and logical inference is used to manipulate these symbols to derive new knowledge. These features supplement neural models with explicit symbolic representations and reasoning capabilities. In neurosymbolic AI, symbolic representations and reasoning are used to provide high-level context and logic, while neural networks are employed for pattern recognition, learning from data, and capturing low-level features. This hybrid approach allows neurosymbolic AI systems to leverage the power of neural networks for tasks such as perception, language understanding, and pattern recognition, while also benefiting from symbolic reasoning for tasks that require explicit knowledge representation, logical inference, and explainability.

In the context of sentiment analysis, we are witnessing three main research directions. The *first direction* focuses on interpretability and explainability: instead of only aiming for high accuracy, neurosymbolic AI systems also focus on how accurately a machine learning model can associate a cause to an effect; and possibly also on the the ability of the parameters (often hidden in deep neural networks) to justify the results. The *second direction* is about leveraging existing lexicons and knowledge bases [3] to fuse factual information [10] or some high-level decision making mechanism at some layer of the neural model. The *third direction* is about leveraging theories developed in other reference disciplines to guide the neural architecture design, so that the sub-module functions become clear and the whole model's interoperability is enhanced. All three directions are achieving sound improvements in sentiment analysis and are deepening our understanding of affective computing and the cognitive root of human emotion.

In the context of this background, we launch this special issue of IEEE Transactions on Affective Computing that focuses on presenting some representative recent advances in designing, using, and evaluating neurosymbolic AI for sentiment analysis.

II. Contents of This Special Issue

Out of the 50 submissions received, only 9 were accepted to appear in this special issue. All articles have been reviewed by at least three reviewers and handled by the guest editors, except for [11] and [12], which were handled independently by other Associate Editors of the journal. Among the selected articles, [11], [13], [14] belong to the *first direction* of interpretability and explainability; [12], [15], [16] belong to the *second direction* of injecting external knowledge into a neural model; finally, [17], [18], [19] belong to the *third direction* of migrating theory from reference disciplines.

The special issue also presents four new datasets that future researches could use: ConvECPE for emotion-cause pair extraction from conversation [11], Emotion-Cause-in-Friends for multimodal emotion-cause pair extraction [13], PerceptSent for visual sentiment analysis [16], and TWISCO for suicidal intent detection [19].

A. Interpretability and explainability [11], [13], [14]

Emotion-cause pair extraction (ECPE) is a task that extracts a segment from emotion words' context to provide an explanation for the emotion detected. ECPE in conversations is especially challenging because the cause expression can span multiple rounds of texts.

In the first article [11], the authors propose a new two-step framework for extracting multiple emotion-clause pairs simultaneously in a conversation. The future plan is to integrate the speaker's personality into the prediction.

In [13], a special type of emotion-rich conversation is studied. The conversations in sitcoms such as 'Friends' are found to contain more emotions and cause clauses compared to movies. In this article, the authors propose multi-modal emotion-cause extraction based on audio, video and text features. Previous studies have constructed an annotated multi-modal conversation dataset from 10 famous TV series. In this article, the text and facial expressions are combined when annotating the cause of an emotion.

In [14], an end-to-end approach for the ECPE task is explored to tackle cascading errors that are often associated with pipeline models. In this article, the authors solved the issue of capturing the implicit co-occurrence or exclusion patterns between multiple pairs of emotions. This is achieved by finding the cause clauses after finding the emotion clauses. They also have a pre-training phase that can reduce the computation time needed for limited datasets.

B. Knowledge selection and utilization [12], [15], [16]

Humans often rely on context and commonsense to understand emotions in utterances. Hence, it is necessary for machines to effectively integrate commonsense databases into their models. Transformers are commonly used in dialogue systems to model words and their context. In addition, an emotional dictionary is used where each word has a score. In [12], the authors propose automatic selection of emotional knowledge to reduce the size of the database. This allows better performance of a sequence transformer on datasets of different sizes and from different domains.

On the other hand, incorporating external knowledge is not a "the more the merrier" game. Previous studies have shown that removing/eliminating ineffective knowledge can improve model performance. In [15], the authors propose to eliminate external knowledge from an utterance that may have negative effects on the classifier. Many previous studies have suggested organizing external knowledge using relationships. For example, in ConceptNet [20], the nodes are concepts, and the edges are the relationships. In this paper, the authors randomly eliminate words whose emotion is inconsistent with the utterance of emotion. Recently, recurrent neural networks are being used to capture inter-speaker dependencies in a conversation. Instead, here they use graph attention to refine the weights of knowledge at both syntactic and semantic levels.

Another way of curating external knowledge is by storing metadata with the dataset. In [16], the authors explore convolutional neural networks and investigate the influence of user subjectivity for visual sentiment analysis. To accurately judge baseline performance, they consider different experimental settings of the number of target classes and voting. A vector space of emotions is found to be more accurate than categorical labelling. This is because people may have the same sentiment for an utterance, but the reason may differ. To leverage a psychological theory, a new dataset for sentiments in Flicker images was created along with perceptions and evaluator's metadata (age, gender, socioeconomic status, education background, and psychological hints).

C. Multidisciplinary interoperability [17], [18], [19]

Automatic text summarization helps reduce the time of filtering large amounts of text. One challenge in text summarization is to capture long distance sentiment flows. In [17], the authors use a meta-design to consider global changes of sentiments in a document with the aim of providing better summaries. Previously, joint-sentiment-topic models have been used to predict the relationship between topic and sentiment. Here, they leverage common structures in writing for organizing sentences such as hierarchical or circular. The output of topic models can also be used to create evaluation indicators. The proposed method shows over 2% improvement over baselines in text summarization.

In [18], the authors tackle the challenge of a large volume of texts in the metaverse, which are very resource-consuming when it comes to emotion processing. A recurrent voting generator was proposed to analyze text in virtual reality using three different algorithmic modules. Here, a dynamic system is needed for content creation in a 3D space. In the past, a multi-combination method was used to control the flow of communication in space and time. In this article, they propose automatic generation of words using an English book such as a drama by Shakespeare in a rapidly evolving metaverse.

Finally, suicidal intent detection in textual data is closely related to sentiment analysis. Previous studies have used manually annotated tweets together with social graphs to predict suicide. However, limited ethical guidance is currently available on this problem. In [19], the authors propose a feature based graph convolutional network for this task. For feature development, the authors apply a coding framework developed in psychology to distinguish between different mentions of suicide. They also investigate the relationship between feelings of dominance and expressions of suicide. Experiments on benchmarks show that a high false negative rate is desirable to detect rare events where a person does not show these intentions.

Acknowledgment

This special issue would not have been possible without the help of many patient and generous reviewers, and responsive IEEE staff. The guest editors would like to thank the current Editor-in-Chief of the IEEE Transactions on Affective Computing, Jesse Hoey, for making the special issue possible, and Elisabeth André, former Editor-in-Chief, for her persistent encouragement.

References

[1] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *27th Annual Conference on Neural Information Processing Systems*, 2013, pp. 3111–3119.

[2] OpenAI, "Gpt-4 technical report," 2023.

[3] K. Du, F. Xing, and E. Cambria, "Incorporating multiple knowledge sources for targeted aspect-based financial sentiment analysis," *ACM Transactions on Management Information Systems (TMIS)*, doi:10.1145/3580480 2023.

[4] J. Xie, X. Liu, D. Dajun Zeng, and X. Fang, "Understanding medication nonadherence from social media: A sentiment-enriched deep learning approach," *MIS Quarterly*, vol. 46, no. 1, pp. 341–372, 2022.

[5] P. Malo, A. Sinha, P. J. Korhonen, J. Wallenius, and P. Takala, "Good debt or bad debt: Detecting semantic orientations in economic texts," *Journal of the Association for Information Science and Technology*, vol. 65, no. 4, pp. 782–796, 2014.

[6] M. Gref, N. Matthiesen, S. H. Venugopala, S. Satheesh, A. Vijayananth, D. B. Ha, S. Behnke, and J. Köhler, "A study on the ambiguity in human annotation of german oral history interviews for perceived emotion recognition and sentiment analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 2022–2031.

[7] F. Xing, L. Malandri, Y. Zhang, and E. Cambria, "Financial sentiment analysis: An investigation into common mistakes and silver bullets," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 978–987.

[8] Z. Liu, Y. Guo, and J. Mahmud, "When and why a model fails? a human-in-the-loop error detection framework for sentiment analysis," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*, 2021, pp. 170–177.

[9] B. Yuan, Y. He, J. Davis, T. Zhang, T. Dao, B. Chen, P. Liang, C. Ré, and C. Zhang, "Decentralized training of foundation models in heterogeneous environments," in *NeurIPS*, 2022.

[10] E. Cambria, Q. Liu, S. Decherchi, F. Xing, and K. Kwok, "Senticnet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 3829–3839.

[11] W. Li, Y. Li, V. Pandelea, M. Ge, L. Zhu, and E. Cambria, "ECPEC: Emotion-cause pair extraction in conversations," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1754–1765, 2023.

[12] D. Jiang, R. Wei, J. Wen, G. Tu, and E. Cambria, "AutoML-Emo: Automatic knowledge selection using congruent effect for emotion identification in conversations," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1845–1856, 2023.

[13] F. Wang, Z. Ding, R. Xia, Z. Li, and J. Yu, "Multimodal emotion-cause pair extraction in conversations," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1832–1844, 2023.

[14] X. Chen, Q. Li, Z. Li, H. Xie, F. L. Wang, and J. Wang, "A reinforcement learning based two-stage model for emotion cause pair extraction," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1779–1790, 2023.

[15] G. Tu, B. Liang, D. Jiang, and R. Xu, “Sentiment- emotion- and context-guided knowledge selection framework for emotion recognition in conversations,” *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1803–1816, 2023.

[16] C. R. Lopes, R. Minetto, M. R. Delgado, and T. H. Silva, “Perceptsent - exploring subjectivity in a novel dataset for visual sentiment analysis,” *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1817–1831, 2023.

[17] X. Li, C. Zou, P. Wu, and Q. Li, “TSSRD: A topic sentiment summarization framework based on reaching definition,” *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1716–1730, 2023.

[18] W. H. Park, N. M. F. Qureshi, and D. R. Shin, “An effective 3D text recurrent voting generator for metaverse,” *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1766–1778, 2023.

[19] A. M. Schoene, L. Bojanić, M.-Q. Nghiem, I. M. Hunt, and S. Ananiadou, “Classifying suicide-related content and emotions on twitter using graph convolutional neural networks,” *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1791–1802, 2023.

[20] R. Speer, J. Chin, and C. Havasi, “Conceptnet 5.5: An open multilingual graph of general knowledge,” in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4444–4451.

Frank Xing received the bachelor’s degrees in information systems and economics from Peking University, and the Ph.D. degree in computer science and engineering from Nanyang Technological University. He is currently an Assistant Professor with the Department of Information Systems and Analytics, National University of Singapore. His research interests include FinTech and the application of sentiment analysis and text mining in information systems. Dr. Xing has served as an guest editor for IEEE Transactions on Artificial Intelligence, Applied Soft Computing, Information Processing & Management (Elsevier).

Björn Schuller (Fellow, IEEE) received the Doctoral degree in electrical engineering and information technology from TUM/Germany. He is currently a Full Professor at the Imperial College London/UK heading GLAM and at the University of Augsburg/Germany. Dr. Schuller is best known for his works advancing machine learning for the engineering of intelligent audio and multi-sensorial complex information systems, affective computing, mHealth, human-computer/robot interaction, and multimedia retrieval. His publications received more than 50000 citations (according to Google Scholar). He is fellow of the AAAC, BCS, ELLIS, and ISCA.

Iti Chaturvedi received the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore. She is currently a Lecturer of Information Technology with James Cook University, Townsville, Australia, where she teaches information processing and design thinking. Her research interests include signal processing and artificial intelligence on social media.

Erik Cambria (Fellow, IEEE) received the Ph.D. degree in computing science and mathematics from the University of Stirling, UK, in 2012, following the completion of an EPSRC project in collaboration with MIT Media Lab, USA. He is currently a Full Professor at Nanyang Technological University, Singapore. His research focuses on neurosymbolic AI for explainable natural language processing in domains, such as sentiment analysis, dialogue systems, and financial forecasting. Dr. Cambria is recipient of several awards, including the 2018 “AI’s 10 to Watch” and the 2019 “IEEE Outstanding Early Career Award”, and was featured in Forbes as “One of the 5 People Building our AI Future”.

Amir Hussain received the Ph.D. degree in novel neural network architectures and algorithms for real-world applications from the University of Strathclyde in Glasgow, UK. He is the Director of the Centre of AI and Robotics at Edinburgh Napier University and Founding Chief Editor of the Cognitive Computation journal (Springer). His research interests are focussed on cognitive data science and trustworthy AI and NLP technologies to engineer the next-generation of smart and secure industrial and healthcare systems. He is a Senior Member of the IEEE and elected Executive Committee member of the UK Computing Research Committee (UKRC) - the National Expert Panel of the IET and the BCS for UK computing research.