Editorial

Special Issue on Machine Learning and Natural Language Processing

Maxim Mozgovoy [1,*] and **Calkin Suero Montero** [2]

1 Active Knowledge Engineering Lab, University of Aizu, Aizuwakamatsu 965-8580, Japan
2 School of Educational Sciences and Psychology, University of Eastern Finland, FI-80101 Joensuu, Finland
* Correspondence: mozgovoy@u-aizu.ac.jp

The task of processing natural language automatically has been on the radar of researchers since the dawn of computing, fostering the rise of fields such as computational linguistics and human–language technologies. Recent advances in machine learning (ML) technologies and the growing power of computing hardware have significantly pushed forward the state of the art in computational linguistics and related fields. The present Special Issue was conceived as a specialised forum for discussing these advancements and charting further potential research directions. We note the high interest of the scientific community on ML-supported natural language processing, resulting in the publication of twenty research papers in the present issue. We also note a growing diversity of natural languages addressed in the studies of this Special Issue, including English, Catalan, Arabic, Chinese, Korean, Vietnamese, Spanish and Portuguese.

Attempting to categorise the research topics addressed here, we highlight:

- **Analysing tweets and Twitter profiles**. Kasthuriarachchy et al. [1] propose an improved method for understanding noisy English texts, while Alshalan and Al-Khalifa [2] design a system for hate speech detection in Arabic tweets. Prada and Iglesias [3], in their contribution, analyse Twitter profiles of the users to predict their reputation on an online marketplace.
- **Annotated text corpora**. Ruiz-Dolz et al. [4] develop a corpus of debate transcripts, suitable for multilingual computational argumentation research. Looking at social media, Bel-Enguix et al. [5] present a corpus focused on negation structures found in Twitter. Additionally, Vu et al. [6] build parallel Korean–English and Korean–Vietnamese datasets targeted at machine translation research, whereas Shaikh et al. [7] exploit text generation models to balance highly biased text corpora in the English language.
- **Chatbots**. Chen et al. [8] improve chatbot performance by augmenting the transformer-based architecture with a memory-based deep neural attention model and Kim et al. [9] propose a method for retaining the dialog context by using a specially designed attention mechanism.
- **Error detection**. Madi and Al-Khalifa [10] address the topic of error detection in the work dedicated to the task of grammar checking in Modern Standard Arabic texts.
- **Named entity recognition (NER)**. Syed and Chung [11] fine tune a BERT model to improve its performance on food menu data in English. Additionally, relying on a finely tuned BERT, Wang et al. [12] present their work, achieving high quality Chinese NER. Kim and Kim [13], on the other hand, design a system able to perform both morphological analysis and NER in Korean, while Dias et al. [14] propose a combination of methods to build a NER system for Portuguese.
- **Natural language understanding**. Son et al. [15] introduce a Sequential and Intensive Weighted Language Modelling scheme that is used together with multi-task deep neural network to outperform state-of-the-art approaches on the standard natural language understanding benchmarks. On the other hand, Zeng et al. [16] provide a survey of existing machine reading comprehension tasks, evaluation metrics, and datasets.

- **Text classification and summarisation.** Long et al. [17] propose a novel graph convolution network-based classifier for the task of relation classification. Focusing on the specialised task of classifying industrial construction accident reports, Zhang et al. [18] present their approach and promising results. Additionally, Kim et al. [19] design a topical category-aware text summariser able to consider the topic of the input document. Zeng et al. [20], furthermore, implement a self-matching mechanism to improve the memory capacity of the document summarisation system.

The diversity of topics, languages and approaches to text analysis points to the large number of underexplored areas in this field and the possibility of further progress. We wish to thank all the authors of the present Special Issue and encourage new contributions to natural language processing technologies.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kasthuriarachchy, B.; Chetty, M.; Shatte, A.; Walls, D. From General Language Understanding to Noisy Text Comprehension. *Appl. Sci.* **2021**, *11*, 7814. [CrossRef]
2. Alshalan, R.; Al-Khalifa, H. A Deep Learning Approach for Automatic Hate Speech Detection in the Saudi Twittersphere. *Appl. Sci.* **2020**, *11*, 8614. [CrossRef]
3. Prada, A.; Iglesias, C.A. Predicting Reputation in the Sharing Economy with Twitter Social Data. *Appl. Sci.* **2021**, *10*, 2881. [CrossRef]
4. Ruiz-Dolz, R.; Nofre, M.; Taulé, M.; Heras, S.; García-Fornes, A. VivesDebate: A New Annotated Multilingual Corpus of Argumentation in a Debate Tournament. *Appl. Sci.* **2021**, *11*, 7160. [CrossRef]
5. Bel-Enguix, G.; Gómez-Adorno, H.; Pimentel, A.; Ojeda-Trueba, S.L.; Aguilar-Vizuet, B. Negation Detection on Mexican Spanish Tweets: The T-MexNeg Corpus. *Appl. Sci.* **2021**, *11*, 3880. [CrossRef]
6. Vu, V.H.; Nguyen, Q.P.; Shin, J.C.; Ock, C.Y. UPC: An Open Word-Sense Annotated Parallel Corpora for Machine Translation Study. *Appl. Sci.* **2020**, *10*, 3904. [CrossRef]
7. Shaikh, S.; Daudpota, S.M.; Imran, A.S.; Kastrati, Z. Towards Improved Classification Accuracy on Highly Imbalanced Text Dataset Using Deep Neural Language Models. *Appl. Sci.* **2021**, *11*, 869. [CrossRef]
8. Chen, J.; Agbodike, O.; Wang, L. Memory-Based Deep Neural Attention (mDNA) for Cognitive Multi-Turn Response Retrieval in Task-Oriented Chatbots. *Appl. Sci.* **2020**, *10*, 5819. [CrossRef]
9. Kim, S.; Kwon, O.W.; Kim, H. Knowledge-Grounded Chatbot Based on Dual Wasserstein Generative Adversarial Networks with Effective Attention Mechanisms. *Appl. Sci.* **2020**, *10*, 3335. [CrossRef]
10. Madi, N.; Al-Khalifa, H. Error Detection for Arabic Text Using Neural Sequence Labeling. *Appl. Sci.* **2020**, *10*, 5279. [CrossRef]
11. Syed, M.H.; Chung, S.T. MenuNER: Domain-Adapted BERT Based NER Approach for a Domain with Limited Dataset and Its Application to Food Menu Domain. *Appl. Sci.* **2021**, *11*, 6007. [CrossRef]
12. Wang, Y.; Sun, Y.; Ma, Z.; Gao, L.; Xu, Y. An ERNIE-Based Joint Model for Chinese Named Entity Recognition. *Appl. Sci.* **2020**, *10*, 5711. [CrossRef]
13. Kim, H.; Kim, H. Integrated Model for Morphological Analysis and Named Entity Recognition Based on Label Attention Networks in Korean. *Appl. Sci.* **2020**, *10*, 3740. [CrossRef]
14. Dias, M.; Boné, J.; Ferreira, J.C.; Ribeiro, R.; Maia, R. Ferreira, J.C.; Ribeiro, R.; Maia, R. Named Entity Recognition for Sensitive Data Discovery in Portuguese. *Appl. Sci.* **2020**, *10*, 2303. [CrossRef]
15. Son, S.; Hwang, S.; Bae, S.; Park, S.J.; Choi, J.H. A Sequential and Intensive Weighted Language Modeling Scheme for Multi-Task Learning-Based Natural Language Understanding. *Appl. Sci.* **2021**, *11*, 3095. [CrossRef]
16. Zeng, C.; Li, S.; Li, Q.; Hu, J.; Hu, J. A Survey on Machine Reading Comprehension—Tasks, Evaluation Metrics and Benchmark Datasets. *Appl. Sci.* **2020**, *10*, 7640. [CrossRef]
17. Long, J.; Wang, Y.; Wei, X.; Ding, Z.; Qi, Q.; Xie, F.; Huang, W. Entity-Centric Fully Connected GCN for Relation Classification. *Appl. Sci.* **2021**, *11*, 1377. [CrossRef]
18. Zhang, J.; Zi, L.; Hou, Y.; Deng, D.; Jiang, W.; Wang, M. A C-BiLSTM Approach to Classify Construction Accident Reports. *Appl. Sci.* **2020**, *10*, 5754. [CrossRef]
19. Kim, S.E.; Kaibalina, N.; Park, S.B. A Topical Category-Aware Neural Text Summarizer. *Appl. Sci.* **2020**, *10*, 5422. [CrossRef]
20. Zeng, B.; Xu, R.; Yang, H.; Gan, Z.; Zhou, W. Comprehensive Document Summarization with Refined Self-Matching Mechanism. *Appl. Sci.* **2020**, *10*, 1864. [CrossRef]