Natural Language Processing Technology Optimization for Human-Computer Interaction and Impacts on Decision Making

Word Count: 2102

Introduction

Human language. The origins of human language date back almost 200,000 years ago, and it continues to evolve even today, given the numerous ways humans have developed communication standards and ways to interact with people from different cultures. However, the topic keeps researchers under check, not only because of its nuances but because of semantic differences and similarities across various forms of language and how those can be overcome to enable efficient interactions. Thus, the need for the emergence of natural language processing research, and in simple terms, it refers to ways in which computer systems are trained to interpret and understand the meaning behind human language, which involves a wide variety of tasks such as speech recognition, sentiment analysis, and knowledge extraction. One of the most overlooked aspects of this field of research is that researchers are constantly working towards resolving ambiguity in language and optimizing language models to tackle many downstream problems efficiently. In a TED talk by Julia Watson, an industry expert in indigenous nature-based technologies 2020, Watson discusses how "...the Tofinu tribe has developed the largest lake city in Africa. Ganvie, meaning "We survived," is built of stilted houses that are organized around a canal system that you can navigate by dugout canoe" (Watson 4:51). The main thing to consider is how various indigenous societies are exploring innovative design techniques as well as implementing them to combat the harmful effects of climate change. Similarly, human decision-making plays a significant factor in improvements in NLP technology because, at the end of the day, and in the face of evolving technology, it is up to humans to decide the amount of trust they want to put into such systems. In a journal called The Meditations, which belonged to the late Roman emperor, Marcus Aurelius Antoninus, Aurelius primarily discusses his ideas regarding Stoic philosophy and brings up an idea that states how "That which is evil for you exists not in the soul of another; nor in any change or alteration of the body…apprehend what evil is" (Aurelius 39). The excerpt discusses the idea of how shifts and changes in the world around a person cannot necessarily harm one, which is true in the sense of NLP research since it provides people with the means to make informed decisions regarding how they want to exercise free will when interacting with the technology. To date, however, the question remains: How can natural language processing technology be truly optimized for human-computer interaction, and what are the potential consequences for decision-making? It may seem surprising that with developments in the last few years, such as bidirectional transformers for language understanding and low-resource machine translation, there is still a long way to go in developing NLP models to have constraint-free conversations with humans. It is still important to consider that through the amplification of biases present in the training data, including a lack of context and misinterpretation of human input, there are numerous ways in which NLP systems can negatively impact human decision-making, but to optimize the technology for human-computer interaction, not only should models be developed that closely align with human intent, but ones that generate truthful and helpful responses, and prioritize transparency to minimize potential errors.

Advances in and Optimization of NLP Technology

When there are so many different applications of Natural Language Processing technology, it may seem challenging to determine which task lays the groundwork for most models to work off of. However, for any model to adequately interpret and understand human language, it must first understand the dependencies between various sentences, otherwise referred to as Natural Language Inference. In 2018, the research community, especially those affiliated with Google's AI Language Research Group, released a model known as BERT: Bidirectional Encoder Representations from Transformers, which primarily aims to handle the limitations of standard language models being unidirectional, meaning that they can only process text in a single direction, thereby rendering the application of fine-tuning approaches to token-level tasks harmful. The researchers specifically mention that "...BERT uses masked language models to enable pre-trained deep bidirectional representation" (Devlin et al. 4172) and that "BERT is the first fine-tuning based representation model that archives state-of-the-art performance on a large suite of sentence-level and token-level tasks" (Devlin et al. 4172). As such, the model is more adept at predicting the vocabulary IDs of words in the masked input tokens purely based on context obtained from surrounding sentences. It is also important to consider that in addition to the development of text-based transformers that can efficiently process text to establish a linkage between various pieces, there have also been advancements in approaches to process text according to downstream learning standards, such as models that utilize an encoder to process input text and generate output using a decoder. In a comprehensive study conducted by yet another group of researchers affiliated with Google Language Research on how transfer learning standards can be used to convert text-based language problems into a text-to-text format, the authors propose a framework that can be used in "...feeding our model text as input and training it to generate some target text. This allows us to use the same model, loss function, hyperparameters, etc. across our diverse set of tasks" (Raffel et al. 3). Through this approach, not only can the limits of transfer learning objectives for NLP be explored, but also unlabeled data sets, in addition to the benefit of being able to scale such models optimally. It is without a doubt that to maximize the performance of NLP models, they need to be trained on large sets of data and fine-tuned using a select combination of hyperparameters, but what is a matter of interest to most researchers is how they can effectively elicit the models' knowledge during pretraining without the need for additional parameters. To capitalize on this effort, researchers from the University of California, Irvine, and the University of California, Berkeley teamed up to propose a solution that not only "...creates a prompt by combining the original task inputs with a collection of trigger tokens according to a template" (Shin et al. 2), but one for which the prompt's language model predictions, "are converted to class probabilities by marginalizing over a set of associated label tokens… would any other classifier" (Shin et al. 2). The minimal effort that is needed on users' ends to come up with their queries and the rise of more personalized interactions between the system and the user as a result of such efforts has also given rise to efforts aimed at reducing language barriers in HCI such as utilizing a multilingual labeled sequence translation model to help transfer knowledge from high-resource to low-resource languages. A joint effort by researchers from the University of Hong Kong, Beihang University, and Microsoft Research Asia has resulted in a framework "... advised by multilingual corpora and phrase-level alignment pairs…effectively transferred to target languages by a round-trip translation and label projection" (Yang et al. 9). The implications are huge as it can not only tremendously improve the way people across cultures communicate with one another, but it can bridge the gap for translation and label projection being used to classify groups to text into various categories. It should be evident that with advancements in algorithms to enable NLP models to process and classify text, communication barriers have been reduced among various populations, and the efficiency of interactions between humans and such systems has significantly improved. The same applies to downstream tasks such as sentiment analysis and speech recognition, owing to widespread modern voice assistant and chatbot implementations. From the standpoint of researchers such as Guimin Hu and his colleagues at the Harbin Institute of Technology in Shenzhen, China, the fact that "...most existing works treat MSA and ERC as separate tasks, ignoring the similarities and complementaries between sentiments and emotions" (Hu et al. 7837) prompted the need for a multimodal sentiment knowledge-sharing framework that formalizes "...MSA and ERC labels into Universal Labels (UL) to unify sentiment and emotion" (Hu et al. 7838). As such, this framework can provide grounds for enabling language systems to recognize human behavior better and tailor their actions. Similarly, members of the Institute of Electrical and Electronics Engineers(IEEE) have also proposed methods for developing adaptation algorithms for neural-network-based speech recognition because of how "NNs require fewer constraints on the input than a Gaussian-based system, along with the gradient-based discriminative training" (Bell et al. 1). By focusing their efforts on proposing how the supervised learning of speech representations, in addition to deep probabilistic generative modeling, can address the issues of adaptation in E2E-based models, one can see how these algorithms can be used in the context of human, accent, and domain adaptation, which plays a crucial role in helping distinguish among various human qualities. The question of what novel NLP algorithms are capable of has been discussed thus far, but the underlying question that still needs to be answered is what will the future of human-computer interaction look like because of the continuous development and integration of scalable software solutions.

Future of Human-Computer Interaction from NLP Technology Optimization

Given that the vast majority of existing NLP infrastructures are built on language queries, questions regarding ethicality and biases are bound to arise, which have been discussed extensively by Professors Emily M. Bender and Batya Friedman from the University of Washington, who have proposed data statements as a means of, "...allowing developers and users to understand better how experimental results might generalize…what biases might be reflected in systems built on the software" (Bender and Friedman 587). Thus, data statements can make NLP systems more ethically responsive and improve scientific outcomes related to software deployment and regression testing. In relation to the idea of software testing as well, researchers from UC Irvine, The University Of Washington, and Microsoft Research have come up with a behavioral testing framework called CheckList that "...encourages users to consider how different natural language capabilities are manifested on the task at hand, and to create tests to evaluate the model on each of these capabilities" (Ribeiro et al. 2). By generating test cases for various benchmarks and detecting critical bugs even in extensively tested models, human interaction with such systems is improved over time. Another metric that researchers have developed to model the performance of NLP systems, but this time across various languages, is through typological linguistics, as demonstrated by how it helps in "bridging the gap between the interpretable, language-wide, and discrete features of linguistic typology… probabilistic models of NLP" (Ponti et al. 591). The significance of this approach lies in the fact that typological combinations can be integrated with the continuous nature of NLP algorithms used in NLP frameworks, which contributes to optimization efforts in the line of feature engineering for HCI. Moreover, the fact that large-scale NLP models can assist with "correcting ambiguity in language and for some adds significant numerical design to information" (Pancholi et al. 791) as stated by Aditi Pancholi, a B.Tech Scholar at the Indore Institute of Science and Technology and her colleagues, creates grounds for further analysis on how they can be further optimized to align with users' interests and how the regressive effects through being trained on open-source datasets can be minimized. Furthermore, researchers affiliated with OpenAI, the company behind the largely successful language models GPT-3.5 and GPT-4, released a study that aims to develop language models that are harmless and helpful based on fine-tuned human feedback. By utilizing

RLHF(Reinforcement Learning from Human Feedback) “as a reward signal to fine-tune our models” (Ouyang et al. 2), the trained models are then able to “retain some alignment even on tasks for which they get very little direct supervision signal” (Ouyang et al. 4). The implications are massive as it entails models that can act according to human input and are adept at following explicit and implicit instructions.

Conclusion

Through the application of novel natural language processing algorithms in a world of ever-increasing dependence on automation, not only can the computational efficiency of downstream tasks such as entity extraction, sentiment analysis, and natural language generation be enhanced, but the extent to which researchers can determine how exclusion, underexposure, and bias impact how humans perceive NLP systems and interact with them. For example, frameworks such as CheckList, developed for collaborative test creation of NLP models, can enable “...evaluation of NLP models that is much more robust and detailed, beyond just accuracy on held-out data” (Ribeiro et al. 4911). Similarly, the issue of human decision-making getting negatively impacted by such systems has also been addressed by researchers at OpenAI through a technique aimed at fine-tuning large language models based on human feedback, which has the potential to limit the model’s tendencies to generate harmful, toxic, and biased output (Ouyang et al. 19). The core functionality of language models revolving around being able to predict the next word in sequence accurately may undoubtedly be one aspect to consider when scaling such models, but the fact that training these models to follow user instructions or making them open-source may lead to severe abuse should also be considered as a means of being able to improve transparency and human trust.

Works Cited

Aurelius, Marcus. *"The Meditations of the Emperor Marcus Aurelius Antoninus, Book IV."* *Google Books*, Google, 2018, www.google.com/books/edition/Meditations/xp5CvgAACAAJ?hl=en.

Bell, Peter, et al. "Adaptation Algorithms for Neural Network-Based Speech Recognition: An Overview." *IEEE Open Journal of Signal Processing*, vol. 2, Institute of Electrical and Electronics Engineers (IEEE), 2021, pp. 33–66. *Crossref*, doi:10.1109/ojsp.2020.3045349.

Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, vol. 1, June 2019, pp. 4171–4186., doi:10.18653/v1/n19-1423.

Emily M. Bender, Batya Friedman; Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics* 2018; 6 587–604. doi: https://doi.org/10.1162/tacl_a_00041

Hu, Guimin, et al. "UniMSE: Towards Unified Multimodal Sentiment Analysis and Emotion Recognition." *Association for Computational Linguistics*, Dec. 2022, pp. 7837–51, aclanthology.org/2022.emnlp-main.534.

Ouyang, Long, et al. "Training Language Models to Follow Instructions with Human Feedback." *Advances in Neural Information Processing Systems*, 31 Oct. 2022.

Pancholi, Aditi, et al. “A Review Paper on Applications of Natural Language Processing - Transformation from Data-Driven to Intelligence-Driven.” *International Journal of Engineering Research and Technology*, vol. 10, no. 05, May 2021, pp. 791–796., doi:10.17577/IJERTV10IS050427.

Ponti, Edoardo Maria, et al. “Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing.” *Computational Linguistics*, vol. 45, no. 3, Sept. 2019, pp. 559–601., doi:10.1162/coli_a_00357.

Raffel, Collin, et al. “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.” *Journal of Machine Learning Research*, vol. 21, no. 1, 1 Jan. 2020, pp. 5485–5551.

Ribeiro, Marco Tulio, et al. “Beyond Accuracy: Behavioral Testing of NLP Models with CheckList.” *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 2020, pp. 4902–4912., doi:10.18653/v1/2020.acl-main.442.

Shin, Taylor, et al. “AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts.” *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7 Nov. 2020, doi:10.18653/v1/2020.emnlp-main.346.

Watson, Julia, director. *How to Build a Resilient Future Using Ancient Wisdom*, TedEd, 21 Aug. 2020, https://www.youtube.com/watch?v=_DOOJx7AZfM. Accessed 24 Mar. 2023.

Yang, Jian, et al. "CROP: Zero-Shot Cross-Lingual Named Entity Recognition with Multilingual Labeled Sequence Translation." *Transactions of the Association for Computational Linguistics*, Dec. 2022, pp. 486–496.