The Recent Advances and Applications of Natural Language Processing

Prakash Chandra Bhoi[1], Dilip Singh[2], Bishwa Ranjan Das[3]

[1,2]Department of Computer Science and Engineering
Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar
[3]Department of Computer Science and Engineering
Aryan Institute of Engineering and Technology, Bhubaneswar

prakashbhoi@gmail.com, dillipsingh0306@gmail.com, biswadas.bulu@gmail.com,

Abstract

This paper presents the recent trends and applications in the current scenario of the society using Natural Language Processing (NLP) text. NLP is a subset of AI which finds growing importance due to the increasing amount of unstructured language data. The rapid growth of social media and digital data creates significant challenges in analyzing vast user data to generate insights. Further, interactive automation systems such as chatbots and ChatGPT are unable to fully replace human's intelligence due to their lack of understanding of semantics and context. To tackle these issues, natural language models are utilizing advanced machine learning (ML) and Deep Learning (DL) techniques to better understand unstructured voice and text data. This paper provides an overview of the recent advances and applications of the top global natural language processing in 2023. Many research works based on text data have been doing and developing different tools using intelligent computing techniques such as Part of Speech Tagging, Text summarization, Sentiment Analysis, Morphological Analysis, Named Entity Recognition, Machine Translation etc. In the current situation, the majority of the tools have already been developed with excellent precision.

Keywords: Odia, Part of Speech, Summarization, Sentiment, Morphology, Machine and Deep Learning.

Introduction

Natural Language Processing (NLP) is a branch of Artificial Intelligence and Linguistics, devoted to make computers understand natural language text or words written by human. Natural language processing came into existence to ease the user's work and to satisfy the wish to communicate with the computer in natural language. Since all the users may not be well-versed in machine specific language, NLP caters those users who do not have enough time to learn new languages or get perfection in it. A language can be defined as a set of rules or a set of symbols. Symbols are combined and used for conveying information or broadcasting the information. Symbols are tyrannized by the rules. Natural Language Processing basically can be classified into two parts i.e. Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text. Linguistics is the science of language which includes Phonology that refers to sound, Morphology that refers to word formation, Syntax referring to sentence structure, Semantics referring to syntax and Pragmatics which refers to understanding. Few of the researched tasks of NLP are Automatic Text Summarization, Co-Reference Resolution, Discourse Analysis, Machine Translation, Morphological Segmentation, Named Entity Recognition, Optical Character Recognition, Part Of Speech Tagging etc. Some of these tasks have direct real world applications such as Machine translation, Named entity recognition, Optical character recognition etc. Automatic

summarization produces an understandable summary of a set of text and provides summaries or detailed information of text, of a known type. Co-reference resolution refers to a sentence or a large set of text that determines which words refer to the same object. Discourse analysis refers to the task of identifying the discourse structure of connected text. Machine translation refers to the automatic translation of text from one human language to another. Morphological segmentation refers to separate a word into individual morphemes and identify the class of the morphemes. Named Entity Recognition (NER) describes a stream of text, determines which items in the text relate to proper names. Optical Character Recognition (OCR) gives an image representing printed text, which helps in determining the corresponding or related text. Part of speech tagging describes a sentence, determines the part of speech for each word. Though NLP tasks are obviously very closely interweaved but they are used frequently, for convenience .Some of the tasks such as automatic summarization, co-reference analysis etc. act as subtasks that are used in solving larger tasks. The goal of Natural Language Processing is to accommodate one or more specialties of an algorithm or system. The metric of NLP assess on an algorithmic system that allows for the integration of language understanding and language generation. It is even used in multilingual event detection[2]. A novel modular system is proposed for cross-lingual event extraction for English, Dutch and Italian texts by using different pipelines for different languages. The system incorporates a modular set of foremost multilingual Natural Language Processing (NLP) tools. The pipeline integrates modules for basic NLP processing as well as for more advanced tasks such as cross-lingual named entity linking, semantic role labelling and time normalization. Thus, the cross-lingual framework allows for the interpretation of events, participants, locations and time, as well as the relations between them. Output of these individual pipelines is intended to be used as input for a system that obtains event centric knowledge graphs. All modules behave like UNIX pipes: they all take standard input, to do some annotation, and produce standard output which in turn is the input for the next module. Pipelines are built as a data centric architecture so that modules can be adapted and replaced**.** Furthermore, modular architecture allows the different configurations and dynamic distribution. Most of the work in Natural Language Processing is conducted by computer scientists while various other professionals like linguistics, psychologist and philosophers etc. have also shown interest. One of the most ironical aspects of NLP is that it adds up to the knowledge of human language. The field of Natural Language Processing is related with different theories and techniques that deal with the problem of natural language of communicating with the computers. Ambiguity is one of the major problems of natural language which is usually faced in syntactic level with further subtasks like lexical and morphology concerned with the study of words and word formation. Each of these levels can produce ambiguities that can be solved by the knowledge of the complete sentence. The ambiguity can be solved by various methods such as Minimizing Ambiguity, Preserving Ambiguity, Interactive Disambiguity and Weighting Ambiguity.

The top nine trends in natural language processing for 2023

1. Virtual Assistants
2. Sentiment Analysis
3. Multilingual Language Models
4. Named Entity Recognition
5. Language Transformers
6. Transfer Learning
7. Text Summarization
8. Semantic Search
9. Reinforcement Learning

Natural Language Processing Applications [27]

With the invention of AI bots such as Alexa, Cortana, Siri, and Google Assistant the use of NLP has surged many folds. Scientists are currently focused on building models that can better comprehend human languages such as English, Hindi, Mandarin, Japanese, Spanish, etc. which are formally known as *natural languages*. Here are some trends projected to dominate in the natural language processing sectors in the upcoming years:

- With growing innovations in the field of AI, machine learning is expected to play a vital role in the natural language processing techniques, particularly in text analytics. In the future years, machine learning engine can be used to perform more thorough analysis by means of supervised and unsupervised learning.
- The continuous evolution of social media platforms clearly indicated that these platforms are going to take over an even more significant role in how companies make decisions. For example, at the time of a quarterly report, a company can depend on several NLP tools to monitor the customer reviews, feedbacks, and responses about their company on social media platforms and in the news.
- The rising pragmatic use of NLP can allow companies with large quantities of unstructured text or spoken data to tackle dark data problems and efficiently identify and collate them for insights.
- NLP is also expected to become more common in areas that need to understand user intent such as intelligent chatbots and semantic search. Along with the growing use of deep learning as well as unsupervised and supervised machine learning, the plethora of natural language technologies are expected to endure to mold the communication capacity of cognitive computing.
- NLP is likely to play a vital part in tracking and monitoring market intelligence reports to mine intelligent data for companies for forming upcoming strategies. From 2021, NLP is predicted to find applications in a wide range of business areas. Presently, this technology is extensively used in financial marketing. It is helpful in sharing comprehensive insights into tender delays, market views, and closings and extracting information from large data sources.

Machine learning models such as reinforcement learning, transfer learning, and language transformers drive the increasing implementation of NLP systems. Text summarization, semantic search, and multilingual language models expand the use cases of NLP into academics, content creation, and so on. The cost and resource-efficient development of NLP solutions is also a necessary requirement to increase their adoption.

The Natural Language Processing Trends & Startups outlined in this report only scratch the surface of trends that we identified during our data-driven innovation & startup scouting process. Among others, transfer learning, semantic web, and behaviour analysis will transform the sector as we know it today. Identifying new opportunities & emerging technologies to implement into your business goes a long way in gaining a competitive advantage.

Recent Advances in the Field of Natural Language Processing

The adoption of natural language processing is rapidly growing owing to rising demand for big data, data analytics, powerful computing, and enhanced algorithms. Several players in the natural language processing market are profoundly investing in various developments related to natural language processing.

For instance:

- In July 2019, Facebook AI and researchers from the University of Washington formulated methods to improve Google's BERT language model and attain performance on par or beyond state-of-the-art results in SQuAD, GLUE, and RACE benchmark data sets.
- In April 2020, researchers from the Stanford University NLP have built Stanza, a multi-human language tool kit. This is tool kit is useful for those working with text from many locales—for instance, social media. It offers support for operating numerous precise natural language processing tools on 60+ languages and for retrieving the Java Stanford Core NLP software from Python.
- In April 2020, researchers from Microsoft Research and Google AI have introduced new benchmarks for cross-language natural-language understanding (NLU) tasks for AI systems like named-entity recognition and question answering systems. Google's XTREME covers 40 languages and consists of 9 tasks, whereas Microsoft's XGLUE covers 27 languages and 11 tasks.

Conclusion

Advances and Applications for natural language processing (NLP) are expanding at a breakneck pace, and NLP itself is undergoing rapid development. With so much data at our disposal, it's critical to comprehend, watch over, and, occasionally, filter it. The availability of low-code, no-code tools and ready-to-use pre-trained models will help NLP grow even more in the next years. Businesses will continue to gain from NLP, from better operations and customer happiness to cost savings and better decision-making. Every business area can be improved through the use of NLP by utilizing AI approaches in relation to NLP. Now NLP helps people to improve in the field of education, finance, health sector with minimum cost. NLP is being progressively integrated into computer science and artificial intelligence to develop systems and software, capable of processing human languages. NLP is no longer limited to linguistics. Various other sectors today integrate NLP in their systems to enhance their processes very quickly and efficiently.

References

[1] A. Bermingham and A. F. Smeaton, "Classifying sentiment in microblogs: is brevity an advantage" in Proceedings of the 19th ACM international conference on Information and knowledge management, 2010, pp. 1833-1836.

[2] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford, Vol. 1, pp. 12, 2009.

[3] Ahonen, Helena, et al. "Applying data mining techniques for descriptive phrase extraction in digital document collections." Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on. IEEE, 1998.

[4] Baccianella, S., Esuli, A. and F. Sebastiani, "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", The International Conference on Language Resources and Evaluation (LREC). Vol. 10, 2010.

[5] Bangalore, Srinivas, Owen Rambow, and Steve Whittaker. "Evaluation metrics for generation." Proceedings of the first international conference on Natural language generation-Volume 14. Association for Computational Linguistics, 2000.

[6] Benson, Edward, Aria Haghighi, and Regina Barzilay. "Event discovery in social media feeds." Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011.

[7] Bondale, Nandini, et al. "Extraction of information from open-ended questionnaires using natural language processing techniques." Computer Science and Informatics 29.2 (1999): 15-22.
[8] Drucker, Harris, Donghui Wu, and Vladimir N. Vapnik. "Support vector machines for spam categorization." IEEE Transactions on Neural networks 10.5 (1999): 1048-1054.
[9] Kaushik, C.and Mishra A., "A Scalable, Lexicon Based Technique for Sentiment Analysis", International Journal in Foundations of Computer Science and Technology (IJFCST), Vol.4, No.5, September 2014.
[10] McCallum, Andrew, and Kamal Nigam. "A comparison of event models for naive bayes text classification." AAAI-98 workshop on learning for text categorization. Vol. 752. No. 1. 1998.
[11] Nakagawa, T., S. Kurohashi, "Dependency Tree based Sentiment Classification using CRFs with Hidden Variables", Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.
[12] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Volume 1, Pp 655-665, 2014.
[13] Papineni, Kishore, et al. "BLEU: a method for automatic evaluation of machine translation." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.
[14] Patil, S. and T. Patil, "Automatic Generation of Emotions for Social Networking Websites using Text Mining", 4th International Conference on Computing, Communication And Networking Technologies (ICCCNT) IEEE, pp4-6, July 2013.
[15] Pimpalkar, A., Wandhe, T., Swati Rao M., Kene M., "Review of Online Product using Rule Based and Fuzzy Logic with Smileys", IJCAT – International Journal of Computing and Technology, Volume 1, Issue 1, February 2014.
[16] Porter, Martin F. "An algorithm for suffix stripping." Program14.3 (1980): 130-137.
[17] Ritter, Alan, Sam Clark, and Oren Etzioni. "Named entity recognition in tweets: an experimental study." Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2011.
[18] Sager, Naomi. "Natural language information processing: A computer grammar of English and its applications." (1981).
[19] Sager, Naomi, et al. "Medical language processing: applications to patient data representation and automatic encoding." Methods of information in medicine 34.1-2 (1995): 140-146.
[20] Sha, Fei, and Fernando Pereira. "Shallow parsing with conditional random fields." Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003.
[21] Sharma, S., Srinivas, PYKL, & Balabantaray, RC (2016). Emotion Detection using Online Machine Learning Method and TLBO on Mixed Script. In Proceedings of Language Resources and Evaluation Conference 2016 (pp. 47-51).
[22] Sun, Xu, et al. "Modeling latent-dynamic in shallow parsing: a latent conditional model with improved inference." Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2008.
[23] Varghese,R., Jayasree, M., "A Survey on Sentiment Analysis and Opinion Mining", IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11, October 2013
[24] Vohra, S. M. and Teraiya, J. B., "A Comparative Study of Sentiment Analysis

Techniques", Journal of Information, Knowledge and research in Computer Engineering Volume – 02, Issue – 02, October 2013.
[25] Wang, Dingding et. al., "Integrating document clustering and multidocumanet summarization", ACM Transactions on Knowledge Discovery from Data (TKDD) 2011.
[26] X. Wang, Y. Liu, C. Sun, B. Wang, and X. Wang, "Predicting Polarities of Tweets by Composing Word Embeddings with Long Short-Term Memory," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, vol. 1, pp. 1343–1353.
[27]https://www.researchdive.com/5343/Analyst-Review/natural-language-processing-market.

.