Mam Muddassira

Natural Language Processing

Name: Muhammad Usman Khan

Roll No: **BSEF20M540**

Article on my website: https://www.codcrafters.com/blog/our-blog-1/natural-language-processing-33

Evolution of Natural Language Processing (NLP)

Fascinating topic of natural language processing (NLP) has undergone substantial development and growth over time. Objective of this field, which lies at the nexus of Artificial Intelligence and Linguistics, has been to enable computers to understand and produce human language. In this thorough investigation, we'll delve into the fascinating past of NLP, from its origin in the 1950s to its most recent frontiers in the twenty-first century.

Birth of NLP (1950-1970)

Alan Turing's Pioneering Vision

Renowned mathematician and computer scientist Alan Turing established the idea of the Turing test as a gauge of artificial intelligence in 1950, which is when NLP's origins can be found. This fundamental notion served as the basis for NLP, igniting interest in the automatic comprehension and creation of natural language.

[https://www.deep-talk.ai/post/history-and-present-of-natural-language-processing]

Early Endeavors in Machine Translation

The Georgetown Experiment: NLP's First Steps

Machine translation (MT), which was an early kind of natural language processing (NLP), began to take off in the 1950s. A key turning point was the Georgetown experiment of 1954, a partnership between IBM and Georgetown University. It sparked expectations that MT might soon break through language boundaries as it attempted to automatically translate sentences from Russian to English. The trip turned out to be harder than expected, though. [https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/]

Rule-Based Era (1970-1990)

Conceptual Systems and Handwritten Rules

Conceptual systems, which intended to organize real-world information into formats readable by computers, began to gain popularity in the 1970s. By the 1980s, complex handwritten rule sets played a significant role in NLP systems. However, Chomskyan linguistic theories' supremacy and computational limits placed restrictions on this era.

[https://www.interaction-design.org/literature/book/the-social-design-of-technical-systems-building-technologies-for-communities-2nd-edition/the-evolution-of-computing]

The Machine Learning Revolution

Unleashing Machine Learning in NLP

As rule-based techniques started to be replaced by machine learning algorithms in the late 1980s, NLP underwent a revolution. With this change, a new paradigm where statistical models and algorithms were at the center of natural language processing was developed. The development of NLP underwent a sea change during this time.

Statistical NLP Takes Center Stage (1990-2000)

Hidden Markov Models and Probabilistic Decisions

The use of statistical models in NLP became increasingly common in the 1990s. With the advent of Hidden Markov Models (HMMs), language processing gained a new perspective. As a result, researchers moved their attention to statistical inference, which enables NLP systems to manage unknown or incorrect inputs. [https://www.linkedin.com/advice/0/what-some-common-applications-challenges-hidden]

Unsupervised and Semi-Supervised Learning

Navigating the Challenges of Language Learning

The 1990s saw a rise in interest in unsupervised and semi-supervised learning techniques as NLP developed. These methods enabled NLP systems to gain knowledge from data that had different levels of annotation. They offered a potential technique to utilize the enormous amount of unannotated data available, including the huge breadth of the World Wide Web, despite being more difficult than supervised learning.

The Neural Revolution (2000-2020)

From 2000 through 2020, neural networks are back in style.

Using a "FeedForward" network, Yoshio Bengio and his group proposed the first neural network-based "language" model in 2001. Data in this kind of network only travels in one direction, from the input nodes to the output nodes via the hidden nodes. In 2011, Apple's Siri rose to prominence as one of the first NLP / AI assistants to be widely adopted by users. The automated speech recognition feature of Siri converts the user's words into concepts that are digitally understood. Following a match between those concepts and specified commands, the voice command system starts the designated actions. For instance, Siri might recognize a "Yes" or "No" response to the question "Do you want to listen to your balance?" and respond appropriately.

[https://deep-talk.medium.com/history-and-present-of-natural-language-processing-f19280866497]

Future of NLP: NO-CODE NLP (2020~∞)

1. No-Code NLP Revolution: Learn how no-code NLP tools are changing the landscape of NLP, making it accessible to a broader audience without the need for programming skills.

2. Key Players in No-Code NLP: Explore five prominent no-code NLP platforms, including DeepTalk, KNIME Analytics Platform, Orange Data Mining, Obviously, and Levity, and their impact on various industries.

3. The User-Friendly Learning Curve: Understand how visual programming and user-friendly tutorials are reducing the learning curve for individuals without a programming background.

4. Real-World Applications: Discover the practical applications of no-code NLP in areas like Customer Experience, Customer Success, Product Management, and more, and how these tools can yield meaningful insights.

[https://appmaster.io/glossary/no-code-natural-language-processing-nlp]

Section 2

Natural Language Processing (NLP) models have become indispensable tools for data scientists, enabling them to tackle a wide range of language-related tasks. In this overview, we will explore eight prominent NLP models that are extensively used in the field of data science.

[https://www.linkedin.com/pulse/top-8-sota-state-art-pre-trained-nlp-models-data-scientists-momin]

1. **Facebook RoBERTa**

   Masking method of BERT (Bidirectional Encoder Representations from Transformers) is built upon by RoBERTa, which stands for Robustly Optimised BERT. It is a self-supervised NLP pre-training system made to be highly effective at tasks like foretelling purposefully buried text fragments. RoBERTa is an effective tool for recognizing and preventing fake news and other inflammatory content since it has been specifically trained to interpret data from news stories.

2. **ULMFiT (Universal Language Model Fine-tuning)**

   The NLP paradigm called ULMFiT, or Universal Language paradigm Fine-tuning, was created by Sebastian Ruder and Jeremy Howard. This model effectively lowers mistake rates in NLP tasks by 18–24%. Text processing, voice-to-text conversion, and comprehending the context of textual language are only a few of the areas where ULMFiT is useful.

3. **Google ALBERT**

   Google ALBERT functions as an open-source application within the TensorFlow framework and is an improved version of BERT. It delivers exceptional accuracy of about 80.1% with just 12 million parameters. Often outperforming BERT in certain areas, ALBERT excels at tasks like phrase prediction, question answering, and conversational response production.

4. **XLNet**

   The Google Transformer-XL model is expanded upon by XLNet. This pre-trained NLP model is skilled at tasks like question answering, text categorization, and sentiment analysis. It excels at learning functions from contexts in both directions. The performance of XLNet has even surpassed that of BERT in several language processing jobs

5. **ELMo (Embeddings from Language Models)**

   ELMo, or Embeddings from Language Models, explores the linguistic contexts, syntax, and semantics of words. It was created by Allen NLP using a sizable text corpus and learns operations using bi-directional models. ELMo is a flexible NLP tool because of its abilities to perform textual entailment, sentiment analysis, and question-answering.

6. **Microsoft CodeBERT**

   As an NLP framework based on a multi-layer bi-directional neural architecture, Microsoft CodeBERT stands apart. It excels at activities like creating code documentation and conducting code searches. Surprisingly, CodeBERT was trained using the largest dataset of GitHub projects, which included code from six different programming languages.

7. **OpenAI-GPT-3**

   A significant pre-trained NLP model created by OpenAI is called OpenAI-GPT-3. This model stands out for its comprehensive pre-training on a whopping 175 billion parameters and its transformer-based design. OpenAI-GPT-3 excels at problem-solving, translation, and specialized jobs requiring instantaneous logical reasoning, such as word unscrambling, news article generation, and helping programmers create machine learning applications.

Citation:

1. https://www.deep-talk.ai/post/history-and-present-of-natural-language-processing

2. https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/

3. https://www.linkedin.com/advice/0/what-some-common-applications-challenges-hidden

4. https://www.linkedin.com/pulse/top-8-sota-state-art-pre-trained-nlp-models-data-scientists-momin

5. https://www.youtube.com/watch?v=EZMOBbu_5b8