The Evolution of NLP for the last few years

A computer program's ability to understand natural language, or human language as it is spoken and written, is known as natural language processing (NLP). It is a part of machine intelligence (AI). NLP has been around for over 50 years and has linguistic roots. It has numerous practical uses in a range of industries, including corporate intelligence, search engines, and medical research.

What Is the Process of Natural Language Processing?

Human language is broken down into pieces in natural language processing so that sentence structure and word meaning can be examined and understood in relation to one another. This enables computers to read and comprehend spoken or written text in a manner similar to that of people.

Before NLP technologies to understand human language, data scientists must first conduct a few key NLP preprocessing tasks:

- **Tokenization:** Text is tokenized when it is divided up into smaller semantic components or single clauses.

- **Part-of-speech-tagging:** Tag words with their parts of speech, such as nouns, verbs, adjectives, adverbs, pronouns, etc.
- **Lemmatization and stemming:** Reducing words to their basic forms to standardize vocabulary
- **Stop word removal**: Removing common words like prepositions and articles that don't contribute much original content (at, to, a, the).
  NLP tools can only then convert the text into a form that a machine can comprehend.

Real-world applications of NLP:

Email Filters:

You have undoubtedly profited from NLP technology if you have used an email account in the past ten years. The latter of which is typically deleted from your account before you even notice it, once an algorithm has been properly trained on email textual data, it can correctly recognise, categorize, and label emails as regular emails, spam emails, or harmful emails. Similarly, depending on the email provider, different email filter kinds, such social or promotional, may be employed. Strong email filtering can also dramatically lower the chance that someone will click on and open a malicious email at work, minimizing the exposure of critical data.

Smart Assistants:

Smart assistants like Siri, Alexa, and Cortana have perhaps gained the most notoriety as the most well-known examples of NLP. They dissect language into word stems, parts of speech, and other linguistic aspects using NLP. The remainder will be handled by natural language generation (NLG), which is what gives machines the ability to "talk," and natural language understanding (NLU), which is what enables machines to understand language. Ideally, this results in the reaction you want. The same principles apply to NLP-Powered chatbots, another type of smart assistant, but they respond to text input from users instead of speech recognition. Given their usefulness as a source of information, the majority of internet businesses now use them as their main means of communication on their websites.

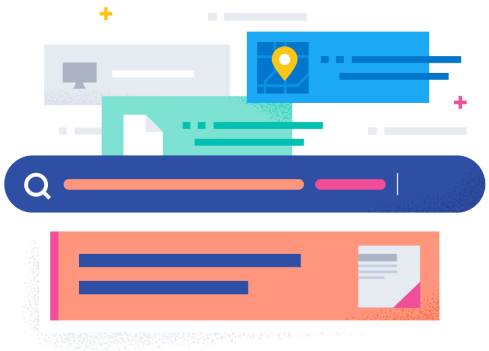
Text Analytics:

Since many businesses have more data than they know what to do with, it can be difficult to derive insightful conclusions. Because of this, a lot of companies now rely on NLP and text analytics to help them interpret their unstructured data. Users have the ability to recognise important information like names, dates, currency values, and even phone numbers thanks to fundamental NLP features like named entity extraction. After that, the entities are grouped into preset

categories to make it simple to find this crucial information in documents of various sizes and types, including files, spreadsheets, web pages, and social messaging. Even the insurance industry uses NLP text analytics to help with claims and risk management decision-making.

Search Results:

People currently access information mostly through online searches. As a result, more businesses are coming to understand the benefits of including NLP search capabilities in their software. Leading NLP innovators like expert.ai are utilizing search to its full potential. Customers can design custom enterprise search solutions, such as those that enable search through message or search for documents connected to a certain topic, using a cognitive search engine (i.e., one that is AI-enabled). For instance, since SUVs, station waggons, and corvettes are all related to "cars," searching for the keyword "car" would also show up documents regarding them.



Language Translation:

Years ago, translating a sentence from one language to another would always produce results that were baffling and/or offensive. This was so common that many people questioned if text translation would ever be correct. Thanks to AI and NLP, it is now possible to produce the comparable meaning in another language by training algorithms on texts written in multiple languages. Even non-Romance languages like Russian and Chinese, which have traditionally been more challenging to translate due to their unique alphabetic structures and use of symbols rather than letters, are now supported by this technique.

Approaches to NLP:

1. Heuristic Method Approach(from 1950):

Any strategy to problem-solving or self-discovery that uses a practical method that is not ensured to be ideal, perfect, or rational but is nevertheless sufficient for achieving an immediate, short-term goal or approximation is known as a heuristic or heuristic methodology.

Examples:

1. Pattern matching with regular expressions
2. Wordnet – Organize data about one word with data about other words.
3. open-minded, logical inquiries

Advantages:

1. As a result of human engagement, it is accurate.

2. It is a prompt strategy.

Disadvantages:

1. Effective heuristic application requires knowledge and experience.
2. You should have to combine the outcomes of several experts.
3. Occasionally time-consuming.

2. Machine Learning Approach(from 1990):

The problem is solved when we utilize the heuristic technique, but when an open-ended circumstance arises, the heuristic approach falls short, and the machine learning approach enters the picture. The significant development in NLP began around 1990, and numerous new strategies emerged in the field. The ability to train a model using already-existing data is a major benefit of employing the machine learning approach. Also, this approach produces the best outcomes.

Machine Learning Workflow

1. Problem Statement
2. Data Gathering
3. Text preprocessing
4. Feature Engineering
5. Model Building
6. Model Evaluation
7. Model Deployment

Convert text to number and give this data to the machine learning model for training and evaluation.

Model used

Naïve Bayes

Logistic Regression

SVM

LDA

Hidden Markov Models

3. Deep Learning Based Approach(from 2010):

Since 2010, deep learning has been widely used in the data science industry because deep learning has advantages over machine learning such as

1. In Deep Learning, feature generation is automatic.
2. Works well with Unstructured data.
3. Deep learning has self-learning capabilities.
4. Ability to deliver high-quality results.
5. Cost Effectiveness. etc

The big advantage:

1. Convert text to numbers in their sequence information to retain.
2. Feature importance is not required.

Deep Learning models

LSTM (Long Short-term Memory)

GRU (Gated Recurrent Unit

CNN (Convolutional Neural Networks)

Transformers

Autoencoder, etc

Trends in NLP(2022):

Low-Code Tools Going Mainstream:

In the past, creating an NLP model required a solid background in the field, coding proficiency to leverage open-source libraries, and machine learning proficiency. not anymore. Although low-code/no-code tools have been available for some time, it's anticipated that they will become more popular this year. Thanks to SaaS providers like Monkey Learn, NLP tasks that were previously solely available to data scientists and developers may now be completed by non-technical persons. We can predict a rise in the usage of NLP tools by businesses in 2022 because of MonkeyLearn's point-and-click model builder, which makes it simple to develop, train, and integrate text classification or sentiment analysis models in just a few clicks.

Transfer Learning:

A model is trained for one job and then repurposed for another activity that is related to the primary task using the machine learning process known as transfer learning. Therefore, you will only need to tweak a pre-trained model rather than creating and training a model from the start, which is costly, time-consuming, and requires enormous amounts of data. By employing lesser amounts of labeled data, firms may execute NLP tasks more quickly. Transfer learning, hitherto only employed in computer vision, is now applied to NLP tasks like intent

The key takeaways from this article –

1. Between 1950 and 1960, the Natural Language Processing period began.
2. Numerous uses can be made of natural language processing. Contextual advertisements, spam filtering, social media (removing adult content), chatbots, and other examples are some that we have seen.
3. Although heuristic technique approaches are precise and quick, they must be used efficiently through training and experience.
4. The benefit of employing a machine learning approach is that you may train a model using data that is already available, and this model produces the best results.
5. The majority of NLP data is unstructured, and deep learning is effective with this type of data.
6. The rise of Low-Code Tools, Chatbots, and Customer Assistants Taken To The Next Level, as well as new developments/trends in NLP like Transfer Learning, Fake News and Cyberbullying Detection, were the last things we noticed.