Natural language processing

Tom Rochette <tom.rochette@coreteks.org>

December 21, 2025 — 77e1b28a

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- How did language emerge?
- Structured vs unstructured languages
- Text summarization techniques
- Text understanding
- Text generation
- Determine if two texts are the same or similar
- Char-RNN (https://github.com/karpathy/char-rnn)
- How to properly classify each type of sentence?
- Is it possible to have a meaningful conversation with a machine without using human language?
    - Can we consider programming languages as such? My first reaction would be no.
        * I think part of what makes human conversation interesting is the unexpected as well as the ability for others to surprise us on their own (by behavior that isn't programmed by the programmer and thus unexpected)

1 Overview

Being able to understand natural language is an important step in building something one might call intelligent. Language is first and foremost how individual communicates with one another using a common syntax.

Some natural languages are not able to express certain concepts at all while others can greatly define that same concept. From this, one will obviously ask: is there such a thing as a language that can express everything?

2 Reading words

Reading a word is basically triggering letter by letter a sub neural network multiple times until the appropriate word is triggered

For instance, reading "word" would trigger all words with the letter "w", then "wo", then "wor" and finally "word". As the same sub neural network gets activated multiple times, its residual activation keeps increasing as more and more letters of the word are read.

3 Requirements

- Linguistics
    - Grammar

     * Syntax
   – Semantics

4 History of approaches

- Hand/hard coding of large sets of rules (if/then)

5 Current *state of the art*

- Based on statistical machine learning
  - Generally grounded in statistical inference
  - Analysis of large corpora of real-world examples
  - Supervised learning using tagged corpus

6 Suggested readings

7 See also

8 References

8.1 General

- https://en.wikipedia.org/wiki/Natural_language_processing

8.2 Summarization/Simplification

- https://en.wikipedia.org/wiki/Automatic_summarization
- https://en.wikipedia.org/wiki/Text_simplification
- http://ultimate-research-assistant.com/
- http://www.wikisummarizer.com/Pages/Default.aspx

8.3 Uncategorized

- https://en.wikipedia.org/wiki/tag:Formal_languages
- https://en.wikipedia.org/wiki/Context-sensitive_grammar