Rahul Agasthya

Professor Katherine Johnston

CSE 300: Technical Writing

11 April 2017

Natural Language Processing:

Text to Speech Conversion and Speech Recognition

Natural Language Processing is a branch of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human language i.e. natural language. Natural languages are languages spoken by humans. Currently, we are not yet at the point where these languages in all of their unprocessed forms can be understood by computers. Natural Language Processing includes a collection of techniques employed to try and accomplish that goal. It is a collection of techniques used to extract grammatical structure and meaning from input in order to perform a useful task as a result, natural language generation builds output based on the rules of the target language and the task at hand. It is useful in the tutoring systems, duplicate detection, computer supported instruction and database interface fields as it provides a pathway for increased interactivity and productivity.

The research work in the natural language processing has been increasingly addressed in the recent years. The natural language processing is the computerized approach to analyzing text and being a very active area of research and development. The literature distinguishes the main application of natural language processing in the field of Text to Speech Conversion and Speech Recognition and the methods to describe it.

The Speech Synthesis approach is based on the text to speech conversion in which the text data is the first input into the system. It uses the sentence segmentation which deals with

punctuation marks with a simple decision tree. Text to Speech synthesis makes use of Natural Language Processing techniques extensively since text data is first input into the system and thus it must be processed in the first place. L. R. Bahl, P. F. Brown, V. D'Souza and R. L. Mercer, in their work, describes the different high-level modules involved in this sequential process (L. R. Bahl).

The work of Alpa Reshamwala, Direndra Mishra, and Prajakta Pawar further contemplates the aspects that are normally taken for granted when reading a text. They write that the sentence segmentation can be achieved through dealing with punctuation marks with a simple decision tree. However, in more confusing situations require more complex methods. Some examples of these more puzzling situations are the period marking, the disambiguation between the capital letters in proper names and the beginning of sentences, the abbreviations, etc. The tokenization separates the units that build up a piece of text and normally splits the text of the sentences at white spaces and punctuation marks. Finally, they write, nonstandard words such as certain abbreviations like Mr., Dr., etc., date constructs, phone numbers, acronyms or email and URL addresses need to be expanded into more tokens in order to be synthesized correctly. Rules and dictionaries are of use to deal with non-standard words. Part-of-Speech Tagging assigns a word-class to each token. Thus, this process pursues the Text Normalization. Part-of-Speech taggers have to deal with unknown words, which is also called as the Out-Of-Vocabulary problem and words with ambiguous POS tags, usually the same structure in the sentence such as nouns, verbs, and adjectives (Alpa Reshamwala).

On one hand, Y. Y. Wang, M. Mahajan, and X. Huang write that the Grapheme-to-Phoneme Conversion assigns the correct phonetic set to the token stream. It must be stated that this is a continuous language dependent process since the phonetic transcriptions of the token

boundaries are influenced by the transcriptions of the neighboring token boundaries. Thus, accounting for the influence of morphology and syllable structure can improve the performance of Grapheme-to-Phoneme conversion (Y.Y. Wang).

While on the other, L. Zhou and D. Zhang have mentioned in their paper, that an application called Word Stress assigns the stress to the words, a process tightly bound to the language of study. The phonological, morphological and word class features are essential characteristics in this assignment whereas, the stress is mostly determined by the syllable weight (L. Zhou).

The Speech Recognition approach uses Natural Language Processing Techniques in a fairly restricted way and is based on grammars. Alpa Reshamwala, Direndra Mishra, and Prajakta Pawar have said that grammar refers to a set of rules that determine the structure of texts written in a given language by defining its morphology and syntax. Hence, for Speech Recognition the incoming speech must follow the predetermined set of rules established by the grammar of a particular language, which is common in many formal languages. In such cases, Context Free Grammars play an important role as they are capable of representing the syntax of the language while being efficient at the analysis of the sentences. Therefore, such a language cannot be considered as a Natural Language. The Speech Recognition systems assume that a large enough grammar rule set enables any language to be taken for natural. Natural Language Processing techniques are of use in Speech Recognition where modeling the language or domain of interaction in question (Alpa Reshamwala).

Through the production of an accurate set of rules for the grammar, the structures for the language are defined. These rules can either be either of the following:

1. hand-crafted; or

2. derived from the statistical analyses performed on a labeled corpus of data.

The former implies a great deal of hard work since this process is not simple as it has to represent the whole set of grammatical rules for the application. However, the latter is generally chosen of a tradeoff between the complexity of the process, the accuracy of the models and the volume of the training and test data. P. Clarkson and R. Rosenfeld continue to write that hand-crafted grammars depend solely on linguistics for a particular language and application, hence they have little interest in machine learning research in general. Thus, the literature is extensive on the data-driven approaches bearing in mind that by definition, a grammar based representation of a language is a subset of a natural language. Hence, the paper suggests to build N-gram language models aiming at a flexible enough grammar to generalize the most typical sentences for an application (P. Clarkson).

N-grams model a language through the estimates of sequences of N consecutive words. While the former tackles the problem with a binary decision tree, the latter chooses to use more conventional Language Modeling theory also makes use of N-gram structures but it pursues a unified model integrating Context Free Grammars. The work of J. R. Bellegarda presents a means of dealing with spontaneous speech through the spotlighting addition of automatic summarization including indexing, which extracts the gist of the speech transcriptions in order to deal with Information Retrieval and dialogue system issues (Bellegarda).

The future of Natural Language Processing is being redefined as there is a push to create more user-friendly systems and face new technological challenges. This is enhancing Natural Language Processing to migrate to an Open Source Environment to make it less proprietary and less expensive. A. Guerra writes, that systems' components thus built will be easily replaceable, which takes less time to develop and is more user-friendly (Guerra). L. Zhou and D. Zhang have

categorically stated that Text to Speech Conversion is tightly bound to the performance of the previous text-processing modules, while the use of the rules of Natural Language Processing is complementary in Speech Recognition (L. Zhou).

Natural Language Processing is a relatively new area of research and development and there have been many success stories already. With the migration to a more Open Source Environment, Natural Language Processing in especially the field of Text to Speech Conversion and Speech Recognition continue to be a major area of Research and Development.

Works Cited

Alpa Reshamwala, Dhirendra Mishra, Prajakta Pawar. “Natural Language Processing.” *Engineering Science and Technology: An International Journal*. vol. 3, no. 1, 2013, pp. 115 - 116.

Bellegarda, J. R. “Statistical language model adaptation: Review and Perspectives.” vol. 42, ed. 1, 2004, pp. 93 - 108.

Guerra, A. “T. Rowe Prime to hone in on voice systems.” *Wall Street and Technology*. vol. 19, ed. 3, 2000.

L. R. Bahl, P. F. Brown, P. V. D’Souza, R. L. Mercer. “A tree based statistical language model for natural language speech recognition.” *Acoustics, Speech and Signal Processing*. vol. 7, ed. 37. Institute of Electrical and Electronics Engineers Transactions, Yorktown Heights, 1989, pp. 1001 - 1008.

L. Zhou, D. Zhang. *NLPIR: a theoretical framework for applying natural language processing to information retrieval*. vol. 2, ed. 54. Association for Information Science and Technology, Silver Spring, 2003.

P. Clarkson, R. Rosenfeld. “Statistical Modeling using the Cum-cambridge Toolkit.” *Euro Speech. Ed. N. F. G. Kokkinakis and E. Dermatas*. Rhodes, Greece, 1997. pp. 2707 - 2710.

Y.Y. Wang, M. Mahajan, and X. Huang. “A unified context-free grammar and n-gram model for spoken language processing.” *IEEE International Conference on Acoustics, Speech, and Signal*. Vol. III. Institute of Electrical and Electronics Engineers, Piscataway, 2000. pp. 1639 - 1642.