Literature Review: Topics Relevant to Speech and Accessibility

Andrew DiStasi
Rochester Institute of Technology
Rochester, NY
add5980@rit.edu

1. RELATED WORKS

The fields of Speech Recognition, Natural Language Processing, and Natural Language Generation have seen a recent surge in attention as the relevant technology becomes more and more sophisticated and the demand for improved accessibility increases. While many are not familiar with it, there is an important distinction between the fields of Natural Language Processing and Natural Language Generation. Natural Language Processing is, essentially, the ability of a computer program to understand human language as it is spoken. Alonso et al. defines Natural Language Generation as a field that "uses analytics, AI, and NLP to obtain relevant information about non-linguistic data and to generate textual summaries and explanations of these data which help people understand and benefit from them" [1]. Alonso et al. conducted a survey of several papers devoted to "recent and prominent developments in the field of NLG with Computational Intelligence" [1], examining how Computational Intelligence and Soft Computing allow NLG Systems to represent and deal with the inherent imprecisions and uncertainties of human language. They explore the benefits of combining CI and NLG approaches to improve the handling of vague tasks within the NLG Pipeline and to help developers quickly adapt NLG systems to a new domain.

Another field being explored in conjunction with NLG is that of Linguistic Description of Data (LDD). Ramos-Soto et al. explore the current state of the Natural Language Generation and Linguistic Description of Data fields, examine how these fields generate easily understandable information from data, and explore "potential points of mutual interest and convergence between both fields" [7]. They explore both fields across multiple applications within the meteorology domain, using it as a means to examine the current state of each field before turning to potential areas of intersection between the two. Ramos-Soto et al. posits that "deeper insight into NLG will greatly benefit LDD researchers, especially regarding the development of applied approaches for practical problems" and that "LDD can be a field of interest for NLG researchers in several respects, including quantified sentences and potential derived extensions, evaluation criteria, algorithms, and more importantly, imprecision handling" [7]. These potential collaborations set the stage for a more complete application that addresses some weaknesses and improves upon some strengths found within applications in these two domains.

Kumagai et al. conducted another exploration into Natural Language Generation, this time focusing on employing a Monte Carlo Tree Search to account for situational nature of the structure and content of human speech. Specifically, they "build a search tree of possible syntactic trees to generate a sentence, by selecting proper rules through numerous random simulations of possible yields" [2]. The primary means of evaluation for this work involves giving an NLG machine the ability to effectively evaluate whether or not a constructed sentence is natural. To achieve this, Kumagai et al. utilize a dual-method evaluation score, combining an evaluation of syntactic structure and of the n-gram language model. These two results can be combined to "score" a system on pass/fail criteria to determine if the sentence is natural. This ability offers a key component of functionality to a speech based system, as how natural using it feels is a large indicator of the system's potential acceptance by users.

Now that speech and Natural Language Generation have been examined, we must examine the other relevant component: Accessibility. As software prevalence becomes denser every day and human life expectancy increases, more individuals are using technology and software more frequently, for longer. "Accessible software does not simply mean that a person with some impairment will be able to use it, it means much more. With accessible software, people who were not able to do simple everyday things, which make up the daily life of the majority of the population, are now able to accomplish them" [10]. Silva et al. explore a plan for improving the awareness of relevant accessibility issues and implementation, and exploring a relation between "relevant accessibility documentation and its appropriate type of user interface" [10]. They motivate their research with a cursory search through two paper databases: The Web of Science and Science Direct. Searches of the phrases "Web Accessibility" and "Software Accessibility" show that approximately 25 papers were related to these topics out of millions of stored papers.

Silva et al. create a first slide proposal "in order to spread e-accessibility among developers" [10]. The primary deliverable from this proposal was a series of relationships mapping types of User Interfaces to a relevant document on Accessibility to consult. They conclude by proposing future tasks for improved accessibility development, such as automated software testing. Also supporting the increased presence of Accessibility in the development sphere, Ladner argues that "Accessibility is becoming mainstream" and that accessibility is slowly and rightfully gaining a larger place in the spheres of mainstream industry and in academia [4]. Ladner examines some of the initial attention accessibility gained in the research sphere in the 1970's and 1980's, before turning his attention to the recent rise into "mainstream" relevancy. He explores several relevant accessibility technologies that have garnered mainstream attention, including closed caption, screen readers, speech recognition, and speech synthesis. To close, Ladner looks forward to the expected continued growth of accessibility devices and technology, positing that "moving forward accessibility will be provided by mainstream companies" [4].

Now that each component has been explored, we can begin looking at some applications of these ideas into a practical domain. Lacey et al. and Norman et al. both explore NLG and NLP technology in generating human readable result overviews from differing types of clinical data. Lacey et al. focused on utilizing natural text descriptions of doctor observations in

Epilepsy clinic letters to “extract meaningful and technically correct clinical information from free text sources” [3]. Making use of IBM Watson Content Analytics software (ICA), they defined annotations based on language characteristics to create parsing rules and an NLP pipeline that highlighted and extracted relevant items from clinic letters, including “symptoms and diagnoses, medication and test results, as well as personal identifiers” [3]. A series of epilepsy clinic letters, containing a mix of “new patient” letters and “follow-up” letters across 12 different doctors, were anonymized and fed into the ICA system. Lacey et al. focused on extracting a discretized epilepsy type, cause, age of onset, medical test results, prescribed medication, and clinic date. Their results show a startlingly high accuracy for all extracted features (the lowest of which was 95%), indicating that, at least in this limited domain example, ICA is capable of properly extracting information from unstructured text.

Norman et al. explore a slightly different type of data, attempting to extract a Pediatric Appendicitis Score (PAS) from a combination of structured and unstructured data. The Pediatric Appendicitis Score is used to aid physicians by automatically generating a score ranking the likelihood that a pediatric patient had appendicitis. This is valuable because the harmful effects of excessive exposure to radiation dictate that diagnostic imaging be minimized, especially in child patients [6]. To that end, a PAS Score below 4 indicates that there is a low suspicion for appendicitis, meaning that imaging is not required unless additional symptoms present themselves. A score above 8 also removes the need for imaging, as this score should automatically lead to a surgery consultation because the likelihood of appendicitis is very high. Norman et al. created a software application that performed NLP preprocessing and feature extraction on a set of text before feeding that information into a model that utilizes a series of classifiers to extract and tag relevant textual data [6]. They found that a Logistic Regression classifier gave them an F-score of 0.9874, indicating that it was very effective at correctly extracting and classifying PAS data from both structured and unstructured tests.

Another experiment conducted by Sauer et al. focused on a combination of structured and semi-structured (rather than unstructured, like the previous two papers) data. Sauer et al. utilized an NLP Tool to extract Pulmonary Function Test (PFT) Reports from Veteran Affairs data of these types. These PFTs are “objective estimates of lung function, but are not reliably stored within the Veteran Health Affairs data systems as structured data” [8]. Data was extracted from the reports of patients at seven VA Medical Centers who suffered from asthma and were fed into a NLP tool Sauer et al. developed. Performance was judged against a human reference standard over 1,001 randomly sampled documents. They found that the tool demonstrated a precision of 98.9% in the validation set, indicating that it can observably improve identification of PFTs in medical research and treatment. However, Sauer et al. caution that it would be erroneous to assume that “a single domain of clinical data can completely capture the scope of a disease, treatment, or clinical test” [8].

The final healthcare related study collected for this paper focuses on using an NLP tool for large-scale data extraction from Echocardiography Reports. Nath et al. observed that because Echocardiography is one of the most commonly ordered diagnostic tests in cardiology, “large volumes of data are continuously generated from clinical notes and diagnostic studies catalogued in electronic health records (EHRs)” [5]. One of the major barriers to leveraging this unstructured data to improve the quality of care for patients is that there are few viable tools that allow accurate extraction of high-quality data from such a large volume of various forms of unstructured data [5]. To that end, Nath et al developed an NLP tool called EchoInfer, that allows for automatic extraction of “data pertaining to cardiovascular structure and function from heterogeneously formatted echocardiographic data sources” [5]. Data elements were extracted and structured into various data formats before being preprocessed and having a series of document and sentence segmentations performed to isolate text relating to certain features and generate relationships between these sentences and the relevant features. Nath et al. analyzed 15,116 echocardiography reports from 1,684 patients, extracting 59 quantitative and 21 qualitative data elements per report [5]. EchoInfer achieved a precision of 94.06%, a recall of 92.21%, and an F1-Score of 93.21% across all data elements in a test subset of 50 reports. Given these results, we can conclude that EchoInfer’s NLP Processing permits large-scale extraction across various data types pertaining to echocardiographic reports with a high degree of precision, accuracy, and recall.

To explore a domain outside of healthcare, we turn to research performed by Schlunz et al., where the examine the accessibility in text-to-speech synthesis for South African Languages. They examine three use cases where multilingual individuals using some form of Augmentative and alternative communication were observed to measure a “baseline integration of the existing Qfrency TTS voices into a selected AAC system and to evaluate the user experience” [9]. Grid 3, an AAC system sold and commonly used in South Africa was integrated with the Qfrency TTS voices and customized to build text interfaces with simple South African sentences. Literate AAC users were recruited to perform acceptance testing on this new tool, focusing on how natural and intelligible the TTS voices were when using the application. Users were asked to utilize closed-form answering machines to rank the prosody, pronunciation, and intelligibility of the system. Intelligibility was scored high fairly consistently, but naturalness ratings were “more spread out between the two poles of robotic and human-like synthetic speech” [9]. This shows that current technology is capable of making systems that can be understood, there are still steps to be taken to improve how natural utilizing such a system feels. Additionally, none of the previous studies focused on the appropriateness of a given data type as a variable in the success or failure of the application. This motivates a line of inquiry into determining how the use of these different data types could affect the naturalness of NLP, NLG, and TTS Systems.

2. REFERENCES

[1] Alonso, Hose M, and Alberto Bugarin. “Natural Language Generation with Computational Intelligence.” *IEEE Xplore*, Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), 19 July 2017, ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7983468.

[2] Kumagai, Kaori, et al. “Human-like Natural Language Generation Using Monte Carlo Tree Search.” *Proceedings of the INLG 2016 Workshop on Computational Creativity and Natural Language Generation*, no. 2016, Sept. 2016, pp. 11–18., www.aclweb.org/anthology/W16-5502. Association for Computational Linguistics.Human-like Natural Language Generation Using Monte Carlo Tree Search

[3] Lacey, Arron S., et al. “Obtaining structured clinical data from unstructured data using natural languageprocessing software.” *International Journal of Population Data Science*, Aug. 2016, pp. 1–2., ijpds.org/article/view/381/362. IJPDS.

[4] Ladner, Richard E. “Accessibility is Becoming Mainstream.” *ASSETS '16 Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 26 Oct. 2016, pp. 1–1., dl.acm.org/citation.cfm?id=2982180. ACM Digital Library.

[5] Nath C, Albaghdadi MS, Jonnalagadda SR (2016) A Natural Language Processing Tool for Large-Scale Data Extraction from Echocardiography Reports. PLOS ONE 11(4): e0153749. https://doi.org/10.1371/journal.pone.0153749

[6] Norman, Brittany, et al. (2017). Automated identification of pediatric appendicitis score in emergency department notes using natural language processing. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7897310

[7] Ramos-Soto, A., et al. “On the role of linguistic descriptions of data in the building of natural language generation systems.” *Fuzzy Sets and Systems*, vol. 285, 15 Feb. 2016, pp. 31–51. *Research Center on Information Technologies*, (CiTIUS), www.sciencedirect.com/science/article/pii/S0165011415003085.

[8] Sauer, Brian C. et al. “Performance of a Natural Language Processing (NLP) Tool to Extract Pulmonary Function Test (PFT) Reports from Structured and Semistructured Veteran Affairs (VA) Data.” *eGEMs* 4.1 (2016): 1217. *PMC*. Web. 26 Oct. 2017.

[9] Georg I. Schlünz et al. 2017. Applications in accessibility of text-to-speech synthesis for South African languages: initial system integration and user engagement. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists* (SAICSIT '17). ACM, New York, NY, USA, Article 32, 10 pages. https://dl.acm.org/citation.cfm?id=3129445

[10] de Sousa e Silva J., et al. (2017) Making Software Accessible, but not Assistive: A Proposal for a First Insight for Students. In: Rocha Á., Correia A., Adeli H., Reis L., Costanzo S. (eds) *Recent Advances in Information Systems and Technologies*. WorldCIST 2017. Advances in Intelligent Systems and Computing, vol 570. Springer, Cham