

Capstone Project : Stylistic Language Editing

Vivek Aithal

University of California,
Berkeley

Priyam Srivastava

University of California,
Berkeley

Daniel McAndrew

University of California,
Berkeley

We propose using contemporary natural language processing (NLP) methods to create an intuitive AI assisted platform for writing. In specific, we want to understand how to generate and edit natural language documents to reflect certain styles. Several theories of language suggest that the meaning of utterances is socially constructed and understood through social contexts (Bakhtin and Holquist, 1981). Many factors (like a speaker’s gender, race, age, occupation, social status, relationship with their listener, current emotional state etc.) are known to influence how language is structured and used to refer to concepts. One important factor that drives writing style is the domain, since many domains have specific stylistic restraints on how language ought to be structured. Some of these domain-specific restraints can pose significant challenges to new writers. We are proposing researching ways that algorithmic tools can aid users in such dilemmas by suggesting edits and additions to meet stylistic constraints. Over the years, many tools from various style guides, to Microsoft Word’s Clippy, have been developed to help people to write in better styles. Modern NLP has provided machine language generation capabilities which are now the frontier of such assistance. Our research will focus on leveraging these innovations, which have often focused on general “best” practices, in order to help with writing in specific domains.

To operationalize our research, we will develop a tool which approaches language generation from a machine-in-the-loop framework, where the writing is performed largely by human users, but is aided by algorithmic suggestions. Specifically, we want to focus on the domain of informal scientific writing such as Wikipedia articles and scientific blog posts which aim to be pedagogical and informative. Our system will be able to aid in the generation of new text and suggest edits to exist-

ing documents, given a desired style. It will also provide suggestions to remove certain forms of biases and portray entities which are mentioned in a neutral manner, if neutrality is a desired stylistic trait. We will incorporate aspects of modern neural style transfer (Gatys et al., 2015) and bias neutralization (Pryzant et al., 2020) approaches in our solution. We believe that this problem is challenging but that given recent progress on related problems and the availability of parallel text datasets (Pryzant et al., 2020) and multidimensional models of text styles (Kang and Hovy, 2019) that we can make novel contributions to the field of stylistic language generation and provide a useful tool for writers.

At the School of Information we find ourselves at the intersection of cutting edge research and a will to democratize access to such research. We recognize that technology enables us to leapfrog traditional barriers and make powerful tools accessible to all people, enriching our society and culture in the process. Informal scientific writing, in our opinion, has some of the most significant impact in democratizing knowledge. Unfortunately, this domain famously suffers from issues of bias, accessibility and quality control. Anyone who has ever made a Wikipedia edit or has written a technical blog post immediately recollects the dozens of examples one has to pore over, to emulate the style, objectivity and tenor of the writing style. At best, it leads to a high variance in quality, and at worst, it makes the process of writing entirely prohibitive to newcomers. We believe that a system such as the one we plan to build will help users generate style-conformant content, without having to spend hours mastering the ‘art’. We believe this project will be a culmination of our journey in the MIMS program. This open-ended area of research will be heavily informed by our coursework and projects in NLP, machine learning, informa-

tion retrieval and by our experiences in building user-facing interfaces.

All three of us plan to collaborate equally on the core NLP research problem. In the interface design, Vivek will focus on developing the back-end system, while Daniel and Priyam will focus on the front-end and user study.

References

- M. Bakhtin and M. Holquist. 1981. *The dialogic imagination: Four essays*. Austin: University of Texas Press.
- Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. [A neural algorithm of artistic style](#). *CoRR*, abs/1508.06576.
- Dongyeop Kang and Eduard H. Hovy. 2019. [xs-lue: A benchmark and analysis platform for cross-style language understanding and evaluation](#). *CoRR*, abs/1911.03663.
- Reid Pryzant, Richard Diehl Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, and Diyi Yang. 2020. [Automatically neutralizing subjective bias in text](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):480–489.