6 Questions for Socially Aware Language Technologies

Diyi Yang, Georgia Institute of Technology, Georgia, USA dyang888@gatech.edu

Abstract Over the last few decades, natural language processing (NLP) has dramatically improved performance and produced industrial applications like personal assistants. Despite being sufficient to enable these applications, current NLP systems largely ignore the social part of language. This severely limits the functionality and growth of these applications. This letter discusses 6 questions towards how to build socially aware language technologies, with the hope of stimulating discussion, inspiring more research into Social NLP, and pushing our research field to the next level.

Over the last few decades, natural language processing (NLP) has had increasing success, and has dramatically improved performance and produced industrial applications like machine translation, search, and personal assistants. The recent Generative Pre-trained Transformer 3 (GPT3) learns language through exposure to numerous examples of data more than a human can access during their life (Brown et al., 2020) , and exhibits state of art performances on a wide range of tasks and their zero-shot learning settings.

Despite being sufficient for applications mentioned above, the current NLP systems largely ignore the social part of language, i.e., they pay attention only to what is said, but not to who says it, in what context and for what goals. This limitation severely limits both the functionality and growth of these applications, such as conversational agents' inconsistent personality and incoherent argument when conducting dialogues with humans, the failure of machine translation in generating culturally respectful outputs, the inability of current systems in commonsense reasoning, or the general struggles of current systems with social intelligence. Ultimately, the goal of NLP is to understand language the way a human does. However, it seems hard to argue that NLP models have reached human level capacity, as language is more than just information —it is about how human use a complex system of words, structures and grammar to effectively communicate with others.

I argue that getting the content correct is not enough and we should push forward on how to build socially aware language technologies that can understand and model social factors in language, interpersonal relations around language use, and the context where language is being used. The idea of language as a social construct is not new: linguistics and philosophy have long modeled it this way (Wittgenstein, 2010; Eckert, 2012) . For instance, instead of pure syntax and semantics, systemic functional linguistics (SFL) (Halliday and Matthiessen, 2013) studies language and its functions in social settings. Grice (1975) laid out four maxims that govern effective communication in social situations, *quality*—make your contribution true, do not lie or make unsupported claims, *quantity*—make your contribution as informative as is required (but not more informative), *relevance*, and *manner*—be brief and orderly and avoid obscurity of expression and ambiguity. Our recent work introduced a set of seven social factors in language and their use in NLP (Hovy and Yang, 2021); Nguyen et al. (2021) highlighted ways of learning and representing social meaning in NLP. These frameworks are quite useful in highlighting and formalizing socially aware language technologies, though there are still obstacles. Overall, I envision 6 questions about socially aware language technologies that need to be thought clearly in order to push our field to the next level.

(1) Is theory necessary in the age of data? Deeper and larger neural networks learn over massive amount of data in an end-to-end way. Should social NLP be model and data oriented? Subtle social factors are often difficult to be defined and measured, especially when "*what is said is not what is meant*", such as sarcasm, irony, deception, and any other situation that requires a "social" interpretation. For that, we need *good* theories to characterize these language phenomena, such as using the aforementioned SFL and social factors taxonomy, as well as social or linguistic theories related to individual NLP phenomena like Brown and Levinson's politeness theory (Brown and Levinson, 1987) and the incongruity theory behind humor (Lefcourt, 2001). Such theories provide grounded perspectives of linking social and language phenomena, towards the knowledge we are advancing. On the other hand, data can extend and inform theory, as many prior theories were produced by speakers of a small set of European languages in a narrow social class stratum, with a dearth of exposure to a variety of utterances.

(2) Are benchmarks the right way to go? One key assumption of most NLP tasks is to reason over the provided benchmark. However, a single corpus might not be enough to include and represent the dynamic scenarios associated with a social phenomenon in terms of its size, genre, and population. For instance, compositionality, commonsense, or implications are key to our daily interactions, but it is often difficult to collect these rich natural situations. "*The abilities of a four-year-old that we take for granted ... answering a question*" (Pinker, 2003) do not require enormous computational or data resources to be achieved. How can we enable models to conduct open domain understanding, as social intelligence goes beyond a fixed corpus? Insights from efforts such as Kiela et al. (2021) and Bowman and Dahl (2021) and perhaps **living** benchmarks via crowdsourcing that can grow and allow for flexible input or output could help benchmark social NLP.

(3) Should social NLP models passively learn or proactively experience the world? Current NLP systems that take social factors in account mainly have been using observational data from online media or other user generated data, though there are a few exceptions in actively simulating data. This "passive" fashion only allows models to examine what is in the data and learn from it from an *association* perspective, but not easily adaptable to new scenarios even with simpler tasks. Moravec's paradox (Moravec, 1988) stated that *"it is comparatively easy to make computers exhibit adult level performance on intelligence tests..., and difficult or impossible to give them the skills of a one-year-old when it comes to perception"*. Perhaps one direction is to let NLP systems experience the world and learn, adapt their use from interacting with human. In practice, experience or interaction involves more than exchanging information via language, but also a wide range of aspects related to social and interpersonal factors reflected in rich modalities. When proactively experiencing the world, socially aware NLP also needs to go beyond text to adequately model the complexity for better understandings of language use.

(4) Should the model and evaluation stay the same? Subtle social factors are often hard to be scaled for annotation due to its subjective nature, and social scenarios often produce dynamically changing data. These *socially low resourced and evolving scenarios* poses new challenges for modern neural network techniques, making it hard for gigantic models to comprehend the world reasonably. For instance, GPT3's less encouraging results when it comes to talking about COVID-19 in late 2020 or historical figures such as asking Steve Jobs GPT-3 "*where are you right now*" and being replied as "*I'm inside Apple's headquarters in Cupertino, California*"—coherent but hardly an up-to-date/trustworthy one (Vincent, 2021). This calls for advances in methodologies that can learn with limited data and evolving facts. Not only with models, especially when social factors are involved, it might be intractable to evaluate such systems (Paullada et al., 2020; Flek, 2020). Current NLP models often use deterministic assessments to compare to some standards or ground-truths. However, these may be inadequate in capturing the nuances of social NLP, as there may be little to none ground-truth, and outputs can be various and change depending on the speaker, receiver, or other aforementioned social factors. Discrepancies with ground-truth might still be acceptable, but could also be detrimental when it comes to high stakeholder scenarios, such as inappropriate outputs from chatbots in counselling context.

(5) How can social NLP be responsible and reproducible? Unique bottlenecks for responsible social NLP includes data collection, and the associated questions about privacy, protection, and ethics, all of which we need to be aware of for doing the right things right. We need careful procedures and practices such as Institutional Review Boards or Ethics and Society Review (Bernstein et al., 2021) to ensure users' data can be used in appropriate and ethical ways (Bender et al., 2021), especially when it comes to protected information that is often manifested unconsciously by users in so-called "publicly observable" social interactions. It is necessary and essential to share data and models in social NLP to facilitate follow-up research; however, even if properly anonymized, data might contain clues to users' identity, and adversary can perform training data extraction attacks to recover personally identifiable information such as names and phone numbers by querying large pretrained language model (Carlini et al., 2020).

(6) Does Social NLP speak English? Most of today's research mainly focuses on 10 to 20 high-resource languages with a special focus on English, though there are thousands of languages and dialects with billions of speakers in the world. Language, dialect and the culture behind largely influences the comprehension of social NLP. For instance, Blodgett et al. (2016) found that existing language identification and dependency parsing tools on African-American Vernacular English text demonstrated very poor performances compared to on Standard English text. As NLP is now applied to everyday interaction globally, meaningful and impactful technologies will have to thoroughly model these social factors to avoid hegemonic approaches assuming all conversations follow Western culture and norms.

Acknowledgement I would like to thank Dirk Hovy for early discussion on related topics and the anonymous reviewers for their valuable feedback.

References

Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.

Bernstein, Michael S., Margaret Levi, David Magnus, Betsy Rajala, Debra Satz, and Charla Waeiss. 2021. Esr: Ethics and society review of artificial intelligence research.

Blodgett, Su Lin, Lisa Green, and Brendan O'Connor. 2016. Demographic dialectal variation in social media: A case study of African-American English. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1119–1130, Austin, Texas. Association for Computational Linguistics.

Bowman, Samuel and George Dahl. 2021. What will it take to fix benchmarking in natural language understanding? In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4843–4855.

Brown, Penelope and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*, volume 4. Cambridge university press.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. 2020. Extracting training data from large language models. *CoRR*, abs/2012.07805.

Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41:87–100.

Flek, Lucie. 2020. Returning the n to nlp: Towards contextually personalized classification models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7828–7838.

Grice, Herbert P. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Halliday, Michael Alexander Kirkwood and Christian MIM Matthiessen. 2013. *Halliday's introduction to functional grammar*. Routledge.

Hovy, Dirk and Diyi Yang. 2021. The importance of modeling social factors of language: Theory and practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602.

Kiela, Douwe, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, Pratik Ringshia, et al. 2021. Dynabench: Rethinking benchmarking in nlp. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4110–4124.

Lefcourt, Herbert M. 2001. *Humor: The psychology of living buoyantly*. Springer Science & Business Media.

Moravec, Hans. 1988. *Mind children: The future of robot and human intelligence*. Harvard University Press.

Nguyen, Dong, Laura Rosseel, and Jack Grieve. 2021. On learning and representing social meaning in nlp: a sociolinguistic perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 603–612.

Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2020. Data and its (dis)contents: A survey of dataset development and use in machine learning research.

Pinker, Steven. 2003. *The language instinct: How the mind creates language*. Penguin UK.

Vincent, James. 2021. Openai's latest breakthrough is astonishingly powerful, but still fighting its flaws. In *The Verge*.

Wittgenstein, Ludwig. 2010. *Philosophical investigations*. John Wiley & Sons.