A Survey of Modeling Context in Recent NLP Work

Abigail Brooks-Ramirez
University of California, Berkeley
a.brooks@berkeley.edu

Abstract

In its essence, the goal of Natural Language Processing (NLP) is to process language at a human level. In achieving this goal, language must be understood as a socially constructed phenomena. NLP, as it stands today, often focuses on the content of text rather than the extra-linguistic context, which may be social, cultural, or political. Modeling and processing human contexts is a highly intersectional task, and this paper surveys the taxonomies and models within this field, as well as the applications of such work.

[1]

1 Introduction

Understanding language as a socially constructed entity is not a novel perspective (Wittgenstein, 2009), and social scientists have long argued that language is a fluid entity dependent on social and cultural context. In examining language in a computational sense, researchers inevitably begin to regard language as a type of data. The written and spoken word is cleaned, labeled, and tokenized into a form that can be processed by large language models (LLMs). It is a long-standing belief that this "datafication" of human processes inevitably abstracts away some human context (Scott, 1998). Further, the design choices for both data and LLMs is an inherently political one, and one that is grounded in an existing body of knowledge. In choosing labels, classifications, and applications, some perspectives may be silenced, while others may be magnified(Geoffrey C. Bowker, 1999).

However, this longstanding sociotechnical framework for understanding language is often underappreciated within the discourse of NLP, which often focuses on the content of data rather than the context (Bisk et al., 2020). A crucial axis for understanding language is *pragmatics* which is the ability to use language in context in order to achieve the desired communicative goal(Fried et al., 2023; Bisk et al., 2020). Semantic interpretation, the meaning directly encoded in the language, differs from pragmatic interpretation, which often depends on extra-linguistic information in order to be fully understood(Pujari et al., 2024). In recent years, NLP has made great progress along this semantic axis, and recent breakthroughs in generating convincingly human-like languages, especially in GPT-4 (OpenAI, 2024), have greatly progressed the field. However, the pragmatic progress has tailed behind, and it has been observed that large language models (LLMs) trained on massive datasets, containing a wide variety of contexts within the written tone, often assume the "average" voice of these datasets in the resulting text generation (Hovy and Yang, 2021). In practice, conversational LLMs often break down when confronted with unexpected human dialogue (Cercas Curry and Rieser, 2018) and while LLMs can sometimes account for cultural context, the cultural and linguistic context of low-resource languages performs significantly worse than dominant (and often Western) languages (AlKhamissi et al., 2024).

NLP research that incorporates social context offers a wide range of insights into both language itself and modeling. Supporting the notion of language as a social construct, recent work found that the mention of gender in the english language has decreased (Ted Underwood, 2018), providing insight into how changes in socio-politcal contexts appear in language. Additional work has introduced new taxonomies and models for representing such deeply sociotechnical information. Further work explores applications of contextual NLP in examining mental health, political information (and misinformation), and analyzing relationships within social networks.

[1]Word Count: 2,112

2 Approaches

2.1 Taxonomies & Data

A recent body of NLP research has worked to examine the issue of missing context in the form of taxonomies, which offer a system for organizing knowledge and metrics for evaluating text representations and models and some practitioners have extended this work to creating datasets.

World Scope (WS) is an approach to evaluating progress in NLP beyond just semantic progress (Bisk et al., 2020). WS describes five phases of progress, as enumerated below, which consider grounding, embodiment, and social interaction within language processing.

1. *WS1. Corpora and Representations (past)* – Text representations have evolved and inevitably grown more complex over time. Syntax trees have shifted into vector representations, and additional specialization has created contextualized pretrained representations. Further analysis of the corpora through methods such as Latent Dirichlet Allocation (Blei et al., 2003) and Latent Semantic Analysis(Scott Deerwester, 2001) allow researchers to further analyze the contents of the corpora.

2. *WS2. The Written Word (most of NLP as of 2020)* – With the advancement of web-crawls, the corpora in NLP has broadened to include unstructured, unlabeled, and multilingual data. As datasets grow larger, so too do the models trained on this data.

3. *WS3. The World of Sights and Sounds (Multimodal NLP)* – In order to truly understand language, speakers must be grounded in the same world, which contains physical heuristics and markers (i.e. "as light as a feather" requires physical context).

4. *WS4. Embodiment & Action* – To truly understand language, an agent must be capable of interacting with a physical or virtual world and transform language to action.

5. *WS5. The Social World* – WS5 aligns with the long-term goal of NLP: an agent must be able to use social context in order to utilize language is a manner that is indistinguishable from a human.

Recent progress within the field of NLP would suggest that many NLP models exist within WS3. However, as this taxonomy describes, there is still progress to be made.

Another taxonomy, rather than evaluating models through the lens of chronological progress, outlines seven forms of context that are crucial to comprehensive language, and uses this lens to evaluate systematic failures in models (Hovy and Yang, 2021). The seven layers in the proposed taxonomy are 1) speaker and 2) receiver, 3) social relations, 4) context, 5) social norms, 6) culture and ideology, and 7) communicative goals. Examining the progress of NLP through any one of these lenses reveals breakdowns in the system, such as the tendency of hateful-content detection to often misclassify banter between friend groups (who may utilize slurs as a method of reclaiming previously harmful terms) as hateful content (Sap et al., 2019); examining this dialogue within the context of sender and receiver, and culture and ideology allows researchers to better understand the source of the breakdown.

Another taxonomy seeks to define social intelligence through a Social Intelligence Data Infrastructure, drawing upon previous work, and further applies this taxonomy to create a comprehensive data infrastructure, supported by a library of 480 NLP datasets (Li et al., 2024). Under this framework, social intelligence may be understood across three threads: *Cognitive Intelligence* is the ability to understand the mental state of others through both verbal and non-verbal queues. Cognitive intelligence is further decomposed into intents, beliefs, and emotions.

Situational Intelligence is awareness of a social context, which includes an understanding of the social event itself, social and moral norms, culture, and speaker information.

Behavioral Intelligence refers to the ability to communicate and act in a manner that successfully attains social goals.

The authors of this social intelligence taxonomy then annotated several NLP datasets according to the taxonomy, finding that the majority of NLP datasets focus on cognitive intelligence (64.2%), followed by behavioral (22.7%) and situational (3.8%). Only 9.4% of datasets spanned multiple social intelligence types (Li et al., 2024).

Grounding data within social context has found success when applied to specific tasks, such as the creation of datasets of tweets that were contextualized within political events and author sentiment (Pujari et al., 2024). Another effort cre-

Table 1: Comparison of four types of generalizable context modeling

Paper/Model	Approach	Data	Scoring Task	Accuracy
LINE	Graph Network Modeling	Language, social, citation networks	Word Analogy	73.79%
NTEL	Graph Network Modeling	Twitter users and mentions	Entity comparison	82.8%
CoSe-Co	Knowledge Graph (KG)	CommonsenseQA Dataset	Question Answering	78.15%
DiffuCOMET	Diffusion & KG	Narratives	Alignment	77.96%

ated a dataset to include broader context (including user interactions) in classifying hateful content (Yu et al., 2022). The work ultimately found that pre-trained and fine-tuned models performed better when the extra-linguistic context of the data was included, though the model results failed to reach a human-level performance.

2.2 Representations & Models

Much of the research on modeling social context is reliant on the notion of *homophily* which is often thought of as "like attracts like" (Yang et al., 2016). In social networks, closely linked individuals are likely to share interests, and in text, words with similar semantics often have similar representations. As a result, many state of the art methods for modeling context utilize graphs to represent relationships. A comparison of the methods discussed in this section is available in 1, which utilize the highest reported success metric for comparisons.

One of the seminal works in information network representation is LINE: Large-scale Information Network Embedding (Tang et al., 2015) which represents large networks through a low-dimensional vector space. The result is highly generalizable, functioning for both directed, undirected, weighted, and unweighted networks. Information networks are defined as *G = (V, E)* where *V* is the set of vertices, each of which is a data object, and *E* is the set of edges, such that each edge connects a data object and signifies a type of relationship between the two objects. In order to meaningfully preserve the network structure within a low-dimensional representation, LINE preserves both local and global networks. The local network is defined by first-order proximity between the vertices, which is the pairwise proximity between two vertices. However, with the understanding of homophily, LINE further preserves second-order proximity of two vertices. Mathematically, let $p_u = (w_{u,1}, ..., w_{u,|V|})$ denote the first-order proximity of *u* with all other vertices. Then the second-rder proximity between u and v is determined by the similarity between p_u and p_v. If no vertex links u and v then the second-order proximity of u and v is 0. The paper performs several tests and utilizes LINE to represent language, citation, and social networks, though the highest performing task was finding word analogies in a language network, a task introduced by (Mikolov et al., 2013). Given a word pair (a, b) and a word c, such as $(Paris, France)$ and $Beijing$, the output should be a word d with a similar relationship to c as the relationship between (a, b) (i.e. $China$; LINE, with second-order proximity, performed the task with 73.79% accuracy, which significantly outperformed other graph models at the time.

Later work built upon LINE to create a Neural model for Tweet Entity Linking (NTEL) (Yang et al., 2016). NTEL leverages LINE to create a distributed representation of twitter users and combines this representation with embeddings of twitter mentions, which is the average of the embeddings of the words contained in the mention (which represented via a skip-gram model). The distributed representation is then combined with a feed-forward neural network that learns non-linear combinations of surface features. Evaluation then compared the gold standard entity the the output entity, with the best performing mode resulting in 82.8% accuracy.

Beyond social networks, there have been several studies that encode social and cultural norms within knowledge graphs (Xie et al., 2024). One study created CoSe-Co (Bansal et al., 2022), in order to tune a baseline knowledge graph to only include vertices and edges relevant to a given sentence, therefore reducing noise in the graph. The resulting highest accuracy, 78.15%, outperformed other models audited in the study.

Rather than modeling social context through graphs, another model, DiffuCOMET, utilized diffusion to model social norms that were necessary to comprehending written narratives (Gao et al., 2024). Diffusion models use a forward process to gradually corrupt real data, and are able to learn a process to de-noise data. DiffuCOMET uniquely utilizes two diffusion models to generate synthetic head entities, which are concated with a given context, and generate subsequent synthetic tail entities; this method creates a process for predicting social norms and relations that may be missing from the provided narrative. The model was evaluated on the alignment (average similarity of generated facts to gold facts) with the highest alignment of 77.96% which is comparable to the baseline models that utilize knowledge graphs (Ji et al., 2020).

3 Applications

Further research in NLP has explored potential applications of the contextual representations of text and extra-linguistic context discussed in Section 2, presenting more fine-tuned approaches with potential for social impact.

Several studies utilized social media text from users to model social context in order to predict suicide risk (Sawhney et al., 2021; Benton et al., 2017). Another study intersecting NLP with mental health examined text from reddit forums discussing mental health, creating deep contextual embeddings with topic modeling in order to create clusters of the most prevalent topics discussed by those suffering from mental health conditions (Kulkarni et al., 2021).

Socio-political context is often useful in analyzing the spread of information, and contextual models are frequently applied in this field. One study incorporated a novel framework, GOLDCOIN (Fan et al., 2024), into models in order to ground the model in privacy laws (a type of socio-politcal norm) in order to evaluate whether transmissions of information were permitted under the framework of contextual integrity (Nissenbaum, 2004; Malkin, 2022). Another study utilized semantic similarity and oppositeness in order to detect contentious twitter interactions, reasoning that higher social contention may correlate to untrue statements and potential misinformation (de Silva and Dou, 2021). A separate study included inference operators to provide context beyond textual data, largely by linking user entities through shared interests and assessing the reliability of article entities, in order to predict misinformation (Mehta et al., 2022). Similar work was done to analyze political tweets with a Graph Attention Network to generate contextualized node embeddings of authoring entities, referenced entities, issues, events and documents (Pujari and Goldwasser, 2021).

4 State of the Art & Conclusion

As seen in Section 2.2, graph-based models for representing contextual knowledge have consistently demonstrated strong performance. These models all outperformed the baseline performance of similar LLMs in their respective studies, thus demonstrating their improved accuracy. Though applications of these models are being applied in the physical world in disciplines such as information dissemination, politics, and mental health, these models still fail to satisfy the taxonomies presented in Section 2.1, namely World Scope (Bisk et al., 2020) and the social intelligence taxonomy (Hovy and Yang, 2021). Thus, in order to create socially responsible and robust systems, it becomes critical for future work in NLP to work towards the taxonomies in Section 2.1, and to continue considering the social implications of language.

References

Badr AlKhamissi, Muhammad ElNokrashy, Mai Alkhamissi, and Mona Diab. 2024. Investigating cultural alignment of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12404–12422, Bangkok, Thailand. Association for Computational Linguistics.

Rachit Bansal, Milan Aggarwal, Sumit Bhatia, Jivat Kaur, and Balaji Krishnamurthy. 2022. CoSe-co: Text conditioned generative CommonSense contextualizer. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1128–1143, Seattle, United States. Association for Computational Linguistics.

Adrian Benton, Margaret Mitchell, and Dirk Hovy. 2017. Multitask learning for mental health conditions with limited social media data. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 152–162, Valencia, Spain. Association for Computational Linguistics.

Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. 2020. Experience grounds language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, Online. Association for Computational Linguistics.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022.

Amanda Cercas Curry and Verena Rieser. 2018. #MeToo Alexa: How conversational systems respond to sexual harassment. In *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*, pages 7–14, New Orleans, Louisiana, USA. Association for Computational Linguistics.

Nisansa de Silva and Dejing Dou. 2021. Semantic oppositeness assisted deep contextual modeling for automatic rumor detection in social networks. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 405–415, Online. Association for Computational Linguistics.

Wei Fan, Haoran Li, Zheye Deng, Weiqi Wang, and Yangqiu Song. 2024. GoldCoin: Grounding large language models in privacy laws via contextual integrity theory. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 3321–3343, Miami, Florida, USA. Association for Computational Linguistics.

Daniel Fried, Nicholas Tomlin, Jennifer Hu, Roma Patel, and Aida Nematzadeh. 2023. Pragmatics in language grounding: Phenomena, tasks, and modeling approaches. *Preprint*, arXiv:2211.08371.

Silin Gao, Mete Ismayilzada, Mengjie Zhao, Hiromi Wakaki, Yuki Mitsufuji, and Antoine Bosselut. 2024. Diffucomet: Contextual commonsense knowledge diffusion. *Preprint*, arXiv:2402.17011.

Susan Leigh Star Geoffrey C. Bowker. 1999. *Sorting Things Out: Classification and Its Consequences*. MIT Press.

Dirk Hovy 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, Online. Association for Computational Linguistics.

Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Xiaoyan Zhu, and Minlie Huang. 2020. Language generation with multi-hop reasoning on commonsense knowledge graph. *Preprint*, arXiv:2009.11692.

Atharva Kulkarni, Amey Hengle, Pradnya Kulkarni, and Manisha Marathe. 2021. Cluster analysis of online mental health discourse using topic-infused deep contextualized representations. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 83–93, online. Association for Computational Linguistics.

Minzhi Li, Weiyan Shi, Caleb Ziems, and Diyi Yang. 2024. Social intelligence data infrastructure: Structuring the present and navigating the future. *Preprint*, arXiv:2403.14659.

Nathan Malkin. 2022. Contextual integrity, explained: A more usable privacy definition. *IEEE Security Privacy*.

Nikhil Mehta, Maria Leonor Pacheco, and Dan Goldwasser. 2022. Tackling fake news detection by continually improving social context representations using graph neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1363–1380, Dublin, Ireland. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *Preprint*, arXiv:1301.3781.

Helen Nissenbaum. 2004. Privacy as contextual integrity. *Washington Law Review*.

OpenAI. 2024. Gpt-4 technical report. *arXiv.org*.

Rajkumar Pujari and Dan Goldwasser. 2021. Understanding politics via contextualized discourse processing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1353–1367, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Rajkumar Pujari, Chengfei Wu, and Dan Goldwasser. 2024. "we demand justice!": Towards social context grounding of political texts. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 362–372, Miami, Florida, USA. Association for Computational Linguistics.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678, Florence, Italy. Association for Computational Linguistics.

Ramit Sawhney, Harshit Joshi, Lucie Flek, and Rajiv Ratn Shah. 2021. PHASE: Learning emotional phase-aware representations for suicide ideation detection on social media. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2415–2428, Online. Association for Computational Linguistics.

James C. Scott. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

George W. Furnas Thomas K. Landauer Richard Harshman Scott Deerwester, Susan T. Dumais. 2001. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, page 1067–1077. International World Wide Web Conferences Steering Committee.

Sabrina Lee Ted Underwood, David Bamman. 2018. The transformation of gender in english-language fiction. *Journal of Cultural Analytics 3*.

Ludwig Wittgenstein. 2009. *Philosophical investigations*. John Wiley Sons.

Yubo Xie, Zonghui Liu, Zongyang Ma, Fanyuan Meng, Yan Xiao, Fahui Miao, and Pearl Pu. 2024. Natural language processing with commonsense knowledge: A survey. *Preprint*, arXiv:2108.04674.

Yi Yang, Ming-Wei Chang, and Jacob Eisenstein. 2016. Toward socially-infused information extraction: Embedding authors, mentions, and entities. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1452–1461, Austin, Texas. Association for Computational Linguistics.

Xinchen Yu, Eduardo Blanco, and Lingzi Hong. 2022. Hate speech and counter speech detection: Conversational context does matter. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5918–5930, Seattle, United States. Association for Computational Linguistics.