Large-scale sophisticated linguistic monitoring

Lisa S. Pearl, Gregory Scontras, Sameer Singh
University of California, Irvine

1 Sophisticated linguistic monitoring

A key means of monitoring current and evolving events is through their linguistic signatures in text-based information sources (e.g, social media like Twitter, Facebook, and Reddit). Given the vast quantities of available data, reliable automated analysis is vital for assisting human intelligence in detecting credible threats. This is especially true given both (i) the complexity of information communicated via language, and (ii) the complexity of the human communication process.

Currently, automated analysis software capable of operating over large-scale datasets in real time is limited in its ability to both (i) reliably detect sophisticated non-linguistic information communicated via the linguistic signal, and (ii) reliably decipher the true meaning underlying a particular linguistic signal. As one example, consider subtle linguistic cues that communicate bias: one post may describe a group as "*freedom fighters*" while another describes them as "*terrorists*". The connotation of the first expression communicates something quite different from the second, indicating the writer's bias for or against the group, which itself may indicate sympathies or group affiliations. The emerging field of **computational sociolinguistics** (Nguyen, Doğruöz, Rosé, & de Jong, 2016) tackles the automatic detection of this kind of non-linguistic information from the linguistic signal (e.g., *connotations*: Rashkin, Singh, & Choi, 2015; *identity*: Pearl & Steyvers, 2012; Pearl, Lu, & Haghighi, 2017; *mental state*: Pearl & Steyvers, 2010, 2013; Pearl & Enverga, 2015; *perspectives*: Hardisty, Boyd-Graber, & Resnik, 2010; Card, Boydstun, Gross, Resnik, & Smith, 2015).

As another example, consider the complex reasoning process people use to understand language in context: if someone posts "*oh yeah, i just want to murder that guy*", there are several possible interpretations. First, this post may be an example of hyperbole or exaggeration, with the writer negatively disposed towards the person in question but without plans to actually murder him. Second, this post may be an example of sarcasm or irony, with the writer positively inclined towards the person in question but using this expression to make a rhetorical point. Third, this post is literal truth and so is a legitimate death threat that should be monitored. Recent approaches in the field of **computational pragmatics** draw on shared context and human processes of social reasoning (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013; Goodman & Frank, 2016) to identify the true interpretation behind a particular linguistic expression (*hyperbole*: Kao, Wu, Bergen, & Goodman, 2014; *irony*: Kao & Goodman 2015; *resolving ambiguity*: Savinelli, Scontras, & Pearl, 2017; *politeness*: Yoon, Tessler, Goodman, & Frank, 2016; *metaphor*: Kao, Bergen, & Goodman, 2014; *humor*: Kao, Levy, & Goodman, 2015).

We see three key challenges for creating sophisticated linguistic monitoring systems that incorporate the insights from computational sociolinguistics and computational pragmatics, and which operate accurately on large-scale datasets in real time:

1. **The multilingual challenge**: Current techniques for sophisticated language understanding typically focus on English data. Can we adapt current methods to perform well across multiple languages?

2. **The symbolic-to-deep-learning challenge**: State-of-the-art machine learning techniques with high accuracy typically involve deep learning methods operating over distributed representations. Many approaches in both computational sociolinguistics and computational pragmatics currently rely on symbolic methods that may not yield accuracy as high for larger datasets. Can we connect existing high-performing symbolic models to deep learning models that perform with high accuracy on other automated linguistic tasks?

3. **The big-data-deployment challenge**: Current techniques for sophisticated language understanding typically operate accurately on small-scale data only. Can we adapt current methods or create new methods that scale?

Addressing these challenges will require insights from linguistics, psychology, natural language processing, artificial intelligence, and computer science. Below we describe recent relevant advances in computational sociolinguistics, computational pragmatics, and deep learning. We then offer suggestions for new interdisciplinary lines of investigation and the resources needed to make significant progress on these challenges.

2 Computational approaches to language understanding

2.1 Computational sociolinguistics

Human language is social by nature—we communicate to others not just about the content of our words (i.e., the linguistic information), but also about aspects of our group and individual identity (i.e., the non-linguistic socially relevant information).

As one example, writer attitude can leave linguistic markers in text that encode a variety of non-linguistic information relevant for group affiliation. Consider the sentence "*The terrorists destroyed the hospital*"; it indicates attitude components such as the writer's perspective about the event being described (negative), the perceived effect of the event (something negative happened to the hospital), the writer's perspective on the event participants (dislike of the subject of the sentence, *terrorists*), and the writer's estimation of the value of the event participants (the hospital is/was valuable). These markers are typically encoded in the words and phrases writers choose to convey their message, with much of the attitude in the sentence above conveyed through the choice of the noun *terrorists* and the verb *destroyed*. Recent techniques like *connotation frames* (Rashkin et al., 2015) organize these rich dimensions of meaning into a formal representation that unifies disparate approaches to nuanced sentiment, perspective, bias, and frame semantics. This formal representation can then be coupled with state-of-the-art machine learning techniques to yield accurate extraction of subtle attitude information.

As another example, writer identity can often be communicated via linguistic style, encoded as a *writeprint*, which is a set of weighted linguistic features (Iqbal, Binsalleeh, Fung, & Debbabi, 2010; Pearl & Steyvers, 2012). The intuition behind writeprints is that certain components of a writer's linguistic usage are unconscious and do not change from document to document, serving as a linguistic fingerprint for that individual. These quantifiable components then become an individual identity marker, allowing for automatic, accurate attribution of a given text to a specific author when writeprints are combined with state-of-the-art machine learning methods—even when writers actively seek to mask their identity (Pearl & Steyvers, 2012). Similarly, *mindprints* (Pearl & Steyvers, 2010, 2013; Pearl & Enverga, 2015) represent the linguistic markers of mental states that are often consciously perceived as the tone of a document (e.g., emotions such as anger, attitudes such as confidence, and intentions such as persuasion). As with writeprints, these quantifiable linguistic components become a signal of the writer's underlying mental state, with automatic and accurate identification possible when combined with state-of-the-art machine learning methods.

These advances in computational sociolinguistics have provided excellent proof-of-concept models for the automatic recognition of subtle non-linguistic information from the linguistic signal. Yet, work remains to show that these approaches can (i) scale, (ii) perform as well across languages, and (iii) perform as well on data targeted for the intelligence community's needs.

2.2 Computational pragmatics

Natural language affords a rich mode of information transfer—beyond the already-dense linguistic code, speakers and listeners enrich the literal interpretations of their messages with so-called *pragmatic* meaning. For example, "*Could you pass the salt?*" is rarely an information-seeking question, but rather a polite request for action. In the fields of natural language processing and artificial intelligence, this pragmatic reasoning that humans effortlessly and often unconsciously deploy poses serious problems when it comes to correctly interpreting naturalistic utterances.

For accurate real-time threat assessment, success may often hinge on the ability to correctly interpret language as humans do.

Recent advances in computer science, in the form of simulation-based probabilistic programs, have paved the way for implementable models of sophisticated language use and understanding. Rather than merely describing a pragmatic reasoning process, these models articulate and implement one, deriving both qualitative and quantitative predictions of human behavior. These predictions consistently prove correct, demonstrating the viability and value of the framework (see Goodman & Frank, 2016 for an accessible summary of applications).

This framework for modeling probabilistic language understanding offers crucial insight into the heretofore off-limits land of pragmatic meaning; we are now closer than ever to *naturalistic* natural language processing. Still, much work remains. Despite inhabiting the same broad modeling framework, each phenomenon is treated with its own application-specific model, most of which have focused on English data. To deploy these methods at scale across languages, the field can look to recent advances in computational linguistics using deep learning techniques.

3 Deep learning in computational linguistics

Deep learning models have recently become useful tools for the analysis of language (Manning, 2016). Mikolov, Sutskever, Chen, Corrado, and Dean (2013) initiated this line of research by establishing that dense vector representations of words, called *word embeddings*, encode various lexical and semantic structures within the vector semantics, and so can be used very effectively in predictive language modeling. Using these compact, continuous representations for language has provided impressive empirical results for a number of natural language tasks (*fine-grained sentiment in structured representations*: Socher et al., 2013, *machine translation*: Cho et al., 2014; Sutskever, Vinyals, & Le, 2014, *summarization*: Chopra, Auli, & Rush, 2016, *dependency parsing*: Chen & Manning, 2014). However, this deep learning revolution for computational linguistics is still in its infancy, with significant potential for growth in representing different aspects of linguistic knowledge.

The primary idea underlying deep learning models is the sharing that occurs within the dense vector representations. This sharing can theoretically provide an exponential representational advantage, and in practice, offers improved learning systems. Unfortunately, these models rely on large amounts of annotated data to learn these representations, and cannot harness the explicit structured linguistic knowledge that language scientists have developed. Because of this, higher-level language processing tasks have not seen the same dramatic error-rate reductions as signal processing oriented tasks (e.g., vision and speech transcription). Moreover, deep learning models are *opaque* and act essentially as black-box functions: the dense representations are incapable of providing the linguistic insight that is crucial for computational natural language analysis.

Because of this, we see two crucial challenges when applying deep learning techniques to sophisticated language understanding:

1. **Incorporating current symbolic insights**: We need to translate current insights in computational sociolinguistics and computational pragmatics into deep learning models, and have these models perform as well as (or better than) existing symbolic models that handle small-scale data. Several recent approaches have made significant headway in this direction by using semantic relations to inform word-level (Faruqui et al., 2015) and phrase-level (Rocktaschel, Singh, & Riedel, 2015) vector representations.

2. **Eliciting interpretable, actionable insights**: When deep learning models perform well on sophisticated language understanding, we must identify the underlying structural, symbolic representations encoded in these models. This will allow further insights into human sociolinguistic and pragmatic reasoning that aid subsequent linguistic analysis. Recent work in *model-agnostic explanations* (Ribeiro, Singh, & Guestrin, 2016) provides tools for simple insights into deep learning models that can be pursued for applications in computational linguistics, such as sophisticated language understanding.

4 Going large-scale with deep learning across languages

We believe a fruitful way forward is to combine the insights from current models of sophisticated language understanding with state-of-the-art deep learning techniques that have proven to provide the best performance on large-scale data across a variety of languages. Concrete immediate goals include

1. **Multilingual performance**: Proof-of-concept testing of current symbolic computational sociolinguistic & computational pragmatic models on realistic data from multiple languages relevant to the intelligence community.

2. **Deep learning implementation**: Proof-of-concept transfer of current symbolic models to deep learning implementations that capture current insights and/or provide innovations in sophisticated language understanding.

3. **Large-scale deployment**: Creation of prototypes of sophisticated language understanding that operate sufficiently rapidly over large-scale data from multiple languages.

We note that in order to accomplish these goals and attune the results to the intelligence community's needs, datasets of linguistic messages (in English and other languages) communicating information of interest to the intelligence community will need to be available/accessible.

References

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., & Smith, N. A. (2015). The Media Frames Corpus: Annotations of Frames Across Issues. In *Acl (2)* (pp. 438–444).

Chen, D., & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. In *Emnlp* (pp. 740–750).

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of EMNLP*.

Chopra, S., Auli, M., & Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. *Proceedings of NAACL*, 93–98.

Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., & Smith, N. A. (2015). Retrofitting word vectors to semantic lexicons. In *Proceedings of NAACL*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, *5*(1), 173–184.

Hardisty, E., Boyd-Graber, J., & Resnik, P. (2010). Modeling Perspective using Adaptor Grammars. In *Proceedings of Empirical Methods in Natural Language Processing* (pp. 284–292). Boston, MA: ACL-EMNLP.

Iqbal, F., Binsalleeh, H., Fung, B. C., & Debbabi, M. (2010). Mining writeprints from anonymous e-mails for forensic investigation. *Digital Investigation*, *7*(1), 56–64.

Kao, J. T., Bergen, L., & Goodman, N. (2014). Formalizing the Pragmatics of Metaphor Understanding. In *Proceedings of the 36th annual meeting of the Cognitive Science Society*.

Kao, J. T., Levy, R., & Goodman, N. D. (2015). A Computational Model of Linguistic Humor in Puns. *Cognitive science*.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Manning, C. D. (2016). Computational linguistics and deep learning. *Computational Linguistics*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Nguyen, D., Doğruöz, A. S., Rosé, C. P., & de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*.

Pearl, L., & Enverga, I. (2015). Can you read my mindprint? Automatically identifying mental states from language text using deeper linguistic features. *Interaction Studies*, *15*(3), 359–387.

Pearl, L., Lu, K., & Haghighi, A. (2017). The character in the letter: Epistolary attribution in Samuel Richardson's *Clarissa*. *Digital Scholarship in the Humanities*, *32*(2), 355–376.

Pearl, L., & Steyvers, M. (2010). Identifying Emotions, Intentions, & Attitudes in Text Using a Game with a Purpose. In *Proceedings of NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text.* Los Angeles, CA: NAACL.

Pearl, L., & Steyvers, M. (2012). Detecting Authorship Deception: A Supervised Machine Learning Approach Using Author Writeprints. *Literary and Linguistic Computing*, *27*(2), 183–196.

Pearl, L., & Steyvers, M. (2013). "C'mon – You Should Read This": Automatic Identification of Tone from Language Text. *International Journal of Computational Linguistics*, *4*(1), 12–30.

Rashkin, H., Singh, S., & Choi, Y. (2015). Connotation frames: A data-driven investigation. *arXiv preprint, arXiv:1506.02739*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. In *Knowledge discovery and data mining (KDD).*

Rocktaschel, T., Singh, S., & Riedel, S. (2015). Injecting logical background knowledge into embeddings for relation extraction. In *Proceedings of NAACL.*

Savinelli, K., Scontras, G., & Pearl, L. (2017). Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior. In *Proceedings of the 39th annual meeting of the Cognitive Science Society.* London, UK: Cognitive Science Society.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., ... others (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Empirical methods in natural language processing (EMNLP)* (Vol. 1631, p. 1642).

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104–3112).

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the 38th annual meeting of the Cognitive Science Society.*