Overview

To truly understand the narratives, motivations, and arguments put forth in written text, computers must traverse a linguistic landscape rife with complex phenomena (like the metaphor in this sentence). How can we encode all of the nuances of human language in a form that is both understandable and useful for machines? The recent success of neural networks—the backbone of deep learning—has forced researchers to rethink this question: words, sentences, and even documents can be represented by learned vectors of real-valued numbers.

For many traditional language-based tasks (e.g., translation, parsing, question answering), deep neural networks are state-of-the-art when given enough data. However, this reliance on big data has restricted researchers to linguistically-bland but omnipresent domains such as newswire text, which in turn limits both the types of tasks that can be tackled and the language diversity in training datasets for traditional tasks. Creative language—the sort found in novels, film, and comics—contains a wide range of syntactic variation (from Hemingway's staccato to the long, winding sentences of Faulkner) as well as higher-level discourse structures (e.g., narrative arcs of characters and relationships) that are absent from broadcast news transcripts or restaurant reviews. A better computational understanding of creative language will lead to more engaging conversational systems, novel multimodal architectures that intelligently combine text and images, and advances in the digital humanities; however, there have been few prior efforts to harness deep learning to this end. **My research focuses on designing deep neural architectures that can understand and generate creative language.**

Prior Work

Much of my work to date has explored the interplay between linguistic syntax and semantics in various deep architectures, primarily through the lens of question answering. In particular, I have designed neural networks for quiz bowl, a trivia game in which players are asked to identify entities (e.g., novels, paintings, battles) from paragraph-long descriptions (see Figure 1). While our models can answer questions about literature and art on par with human quiz bowl players, they have not actually "read" any novels or "seen" any paintings! Inspired by this observation, another thread of my research involves building deep models that understand creative source material by learning concepts from words or pixels; this is also the direction that I hope to pursue in the future.

Deep Learning for Question Answering: I collaborated with Richard Socher to create QANTA, one of the first deep learning models for question answering (Iyyer et al., 2014a). QANTA outperforms traditional question answering baselines on quiz bowl datasets; more impressively, it tied against a team of four former Jeopardy champions in its first exhibition match and, after an upgrade to the underlying neural network (Iyyer et al., 2015), soundly defeated Ken Jennings, Jeopardy's highest-earning player of all time. The network architectures we propose differ mainly in their sensitivity to syntax; we find that since most quiz bowl questions are entity-centric (e.g., "what hero slayed *Grendel*"), architectures that explictly encode syntactic structure and word order are overkill.

One event in this play is foreshadowed by a character's green pajamas. In another scene, a housewife leads several cafe patrons in a funeral procession for her dead cat. A character in this play claims that all cats are dogs because both have four paws, and that Socrates is a cat because he is dead, while attempting to explain syllogisms. In one scene in this play, a staircase's destruction prompts Mrs. Boeuf to leap out a window, after which Mr. Papillon orders Botard and the others to return to work. At this play's end, its protagonist is abandoned first by Dudard, then by his love interest Daisy, before screaming "I'm not capitulating!" For 10 points, identify this Eugene Ionesco play in which Berenger refuses to become one of the title animals.

Figure 1: A quiz bowl question about Eugene Ionesco's absurdist play *Rhinoceros*.

This work has had considerable research impact: QANTA won the **best demonstration award at NIPS 2015**, and the original paper has been cited over 100 times in the two years since its publication. As part of a question-answering workshop I co-organized at NAACL 2016, we set up a shared task that pitted different computerized quiz bowl systems against each other; the winning team essentially replaced pieces of QANTA with deeper network architectures. Finally, I was able to apply many of the techniques I learned while building QANTA to more complicated reasoning-based question answering tasks during internships at MetaMind (Kumar et al., 2016) and Microsoft Research (Iyyer et al., 2016c).

The experience of building and tuning QANTA led us to ask a broader question: to what degree is syntactic structure important for natural language processing? We experimented with two different classes of neural

architectures for QANTA: (1) *compositional* models that encode word order and syntactic structure, and (2) *unordered* models that treat each question as a bag-of-words. In what was a surprising result at the time, we found that for both quiz bowl and sentiment analysis, compositional models do not have any significant accuracy advantages over unordered models and are an order of magnitude slower to train. Other researchers have since found similar results on tasks such as textual entailment and sentence similarity (Wieting et al., 2016; Hill et al., 2016). Does this mean that syntax is not important for these tasks, or that syntax is important but existing neural architectures are inadequately performing composition? Our analysis reveals a third possibility: training datasets for these tasks do not include enough syntactically-interesting examples for a network to properly learn compositionality (Iyyer et al., 2015). Creative language falls at the opposite end of the spectrum, historically containing *too much* complexity for computational models to handle. The advent of deep learning presents the opportunity to revisit learning from creative domains.

Creative Language Understanding: In addition to increased linguistic variety, creative language also often contains subtext (e.g., unstated motivations or desires of characters); sometimes authors themselves have specific intents they want to convey. I have explored the latter phenomenon by building neural networks structured around syntactic parse trees to detect the political ideology of a sentence's author (**liberal** or **conservative**). Our data comes from a collection of politically-charged books by authors such as Al Franken and Ted Nugent, and it contains subtle ideological indicators like sarcasm and scare quotes that can be modeled by compositional network architectures (e.g., liberals are more likely than conservatives to talk about **climate change**, but small modifications to this phrase can dramatically change its ideological bias, as in **so-called "climate change"**). As liberals and conservatives tend to use very similar vocabularies to discuss political issues, these small syntactic variations contain substantial predictive power; our syntactically-aware neural network backs up this intuition by outperforming unordered baselines (Iyyer et al., 2014b).

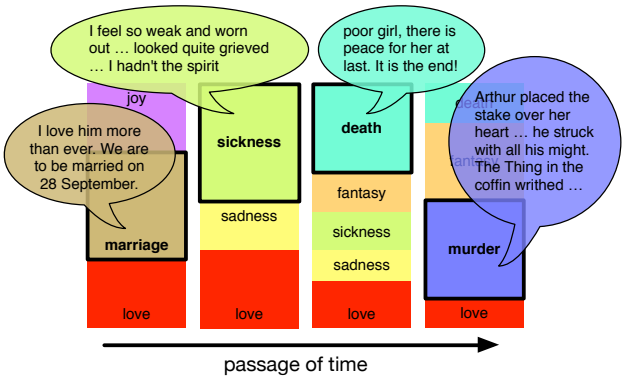

Figure 2: The dynamic relationship between Lucy and Arthur in Bram Stoker's *Dracula*, which starts with love and ends with Arthur killing the vampiric Lucy.

Moving beyond sentence-level authorial intent into the realm of fictional narratives, I designed an interpretable deep autoencoder that understands how relationships evolve between characters in novels (Iyyer et al., 2016a) . Unlike previous work, which requires plot summaries annotated with "friend" or "enemy" labels, our model takes an input an unlabeled corpus of novels and learns both a set of relationship states (e.g., sadness, love, murder) and an ordering of these states for each relationship, as shown in Figure 2. In this work, which won the **best paper award at NAACL 2016**, I propose a novel combination of deep learning and dictionary learning that significantly improves over topic models, which are traditional tools of digital humanities researchers.

Current & Future Work

I have two ongoing projects that seek to advance creative language understanding: (1) generating language in different styles, and (2) modeling how text and images interact in comic book narratives. My long-term vision is to leverage models trained on creative text to improve conversational agents.

I am currently experimenting with a model that can disentangle style and content in language, analogous to a recently-developed computer vision method (Gatys et al., 2015) that allows users to add artistic flavor into everyday photos. While "style" is less well-defined for language than it is for images, we find that by constraining linguistic "style" to certain syntactic elements (e.g., part-of-speech tags, parse trees), we can modify a sentence's style without appreciably changing its content. As an example, the model takes the original sentence **he had nowhere else to go** as input; when I tell it to make the sentence longer and from the first person point-of-view, it generates **and now i knew that a part of him had nowhere to go**. The next step is to learn what separates one author's style from another's; imagine an application that can rewrite any user input into Shakespearean verse, or more practically one that can anonymize an input text by removing all traces of the author's style.

While the sentence style project takes inspiration from methods in computer vision, I am also interested in explicitly integrating deep architectures for vision and language. Recent neural models for image captioning can tell you what is happening in a given image (Xu et al., 2015); can we move beyond simply describing an image with text to understanding how text and images work together to tell a complete story (as in Figure 3)? To answer this question, we collect a large dataset of "Golden Age" comic books (from 1930-1954) and design tasks that test a model's understanding of narrative structure (e.g., predict what text belongs to a particular speech balloon given the previous n panels as context) (Iyyer et al., 2016b). This dataset contains over 1.2 million panels (making it the largest publicly-available dataset of human-created images), and we hope that the research community finds it useful not only as a source of visual narratives but also for transfer learning purposes (can features learned on comics help performance on natural images and vice versa?).

In the long term, I am interested in incorporating creative language into both the language processing and generation aspects of conversational agents. Services such as Amazon's Alexa and Apple's Siri, while limited to a small set of domains, function as user interfaces for millions of people every day. Expanding the types of conversations these systems can have is a rapidly-growing research topic, and neural architectures offer a way to increase conversational acuity by taking advantage of large existing datasets of human interactions (e.g., Twitter). However, current neural conversational systems routinely spit out short, commonplace responses such as "I don't know" at inappropriate times; moreover, most of them do not have mechanisms for predicting the inner moods or beliefs of their users. Integrating networks that can understand fictional relationships and narratives into these agents will hopefully lead to more natural and meaningul conversations (Danescu-Niculescu-Mizil et al., 2011). A complementary approach is to develop text generation models that learn how to mimic the speaking style of their conversational partner (or others, real or fictional). Finally, grounding agents in the visual world will allow for more contextual utterances (e.g., HAL 9000 complimenting Dave on his drawings in *2001: A Space Odyssey*). These ideas are necessarily inter-disciplinary, and to implement them I hope to collaborate with researchers in digital humanities, computational social science, and computer vision.

Figure 3: The dialogue in the left panel gives us enough context to understand the ensuing explosion.

References

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the World Wide Web Conference*.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv*.

Felix Hill, Kyunghyun Cho, and Anna Korhonen. 2016. Learning distributed representations of sentences from unlabelled data. In *Conference of the North American Chapter of the Association for Computational Linguistics*.

Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III. 2014a. A neural network for factoid question answering over paragraphs. In *Proceedings of Empirical Methods in Natural Language Processing*.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014b. Political ideology detection using recursive neural networks. In *Proceedings of the Association for Computational Linguistics*.

Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the Association for Computational Linguistics*.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016a. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Conference of the North American Chapter of the Association for Computational Linguistics*.

Mohit Iyyer, Varun Manjunatha, Anupam Guha, Yogarshi Vyas, Jordan Boyd-Graber, Hal Daumé III, and Larry Davis. 2016b. The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. *arXiv preprint arXiv:1611.05118*.

Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. 2016c. Answering complicated question intents expressed in decomposed question sequences. *arXiv preprint arXiv:1611.01242*.

Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, Ishaan Gulrajani James Bradbury, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing.

John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2016. Towards universal paraphrastic sentence embeddings. *Proceedings of the International Conference on Learning Representations*.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the International Conference of Machine Learning*.