Research Statement

Maria L. Pacheco

My long term goal as a researcher is to build language and information technologies that can be applied in real-world scenarios. Language "in the wild" is complex and ambiguous, and relies on a shared understanding of the world for its interpretation. Namely, a lot of the context needed to convey meaning is not explicit in the language used. As an example, we can look at two statements made by political opponents on the topic of immigration to the US (Ex. 1). These two statements express opinions on the same topic, and use very similar wording to communicate very different ideas. Knowledge of what liberals and conservatives value is key to grasp these contrasting views.

Example 1

Argument 1

If people are trying to flee the dangerous countries they are from, it is unjust to subject them to a grueling, long and demanding process to stay in the US.

Argument 2

Many legal immigrants to the US went through long and demanding procedures in order to gain their status. It is unjust to allow others to circumvent these rules.

Most current NLP methods represent language by learning word co-occurrence patterns from massive amounts of linguistic data. This representation can be very powerful, but it lacks the mechanism to represent real-world context. To address this challenge, we need to find a way to model the concepts and abstractions that allow us to characterize the information expressed in the text. For example, given that political opinions are grounded in morality, we could analyze the meaning of political statements by explicitly modeling the sentiments, beliefs and world views of their authors. In Example 1, Argument 1 is emphasizing *fairness* towards *asylum seekers*, which is a *liberal* talking point, while Argument 2 is emphasizing *fairness* towards *legal immigrants* within the US immigration system, which is a *conservative* talking point. Explicitly modeling and reasoning over these concepts lets us disambiguate the specific language used to express opinions, and allows us to create a model of the world that explains public discourse and human interactions.

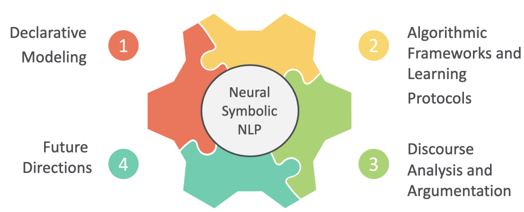

Figure 1: Neural-Symbolic Agenda

Reasoning about abstract concepts and grounding them in text requires us to combine the flexibility afforded by neural methods to identify patterns in large-scale raw data, with a principled way to reason over higher-level patterns. **My research is driven by the opportunities of combining the complementary strengths of neural and symbolic representations**. On the one hand, symbols have inherent explanatory power, and they can help us express domain knowledge and enforce consistency across different decisions. When analyzing public discourse, we could use symbols to explicitly represent conceptual frameworks studied in the social sciences, such as *framing dimensions* and *moral foundation theory*, and enforce consistency between them based on our understanding of the world. On the other hand, expressive distributed representations allow us to leverage the strengths of statistical language models to make sense of large amounts of linguistic data, and provide us with a shared high-dimensional space to ground abstract concepts in language and align different modalities. Following this observation, my work centers around four key challenges and opportunities in the space of neural-symbolic NLP, and covers all stages of the pipeline: modeling, learning, inference and real-world applications. My work in these areas lays the foundation to explore problems and applications with broader societal impact, especially those that can benefit from the integration of structured human knowledge, social context, and large scale language analysis, such as improving security against information warfare, analyzing public discourse at scale, studying social movements, and understanding the impact of policy decisions.

1 Declarative Modeling

Statistical approaches to NLP typically rely on an iterative process of collecting and annotating data, engineering features, specifying predictive models and analyzing errors. Knowledge is communicated as a set of labeled input-output pairs, and the focus is placed either on coming up with an informative set of features, or in devising learning models that capture the information in a latent high dimensional space. Alternatively, high-level modeling allows us to decompose the decision into smaller parts, and express knowledge in a structured way. Declarative modeling in particular, allows us to **shift our focus to what we want to achieve**, rather than how to achieve it. This is a key advantage when collaborating with people outside of machine learning and NLP, as it gives them an interface to contribute their domain expertise, and focus on the aspects, variables and interactions that they want to model.

As part of my dissertation, I developed DRaiL [1, 2], a declarative modeling framework that uses a combined neural-symbolic representation for modeling the interaction between multiple decisions in structured and relational domains. Unlike end-to-end neural networks that work directly over feature vectors, users can explicitly model high-level concepts by defining a set of relevant entities and relations. Then, dependencies between different aspects can be expressed using first-order logic rules. DRaiL's language allows us to quickly prototype relational models in a principled way, and study the interaction between representation, inference and learning. The neural component of DRaiL embeds entities and relations in a shared distributed space, allowing us to learn representations that are relation specific (e.g. in Example 1, Arg. 1 and Arg. 2 can be similar with respect to their moral foundation -*fairness*-, but different with respect to their ideological messaging). The symbolic component of DRaiL allows us to express consistency and constraints over different decisions. In turn, this can be exploited to align representations across different modalities and to combine multiple sources of indirect supervision. DRaiL has been successfully used to model various language domains, including debates and argumentation [2, 3], citation networks [4], scripts and narratives [5], and political messaging [6].

2 Algorithmic Frameworks and Learning Protocols

There are three main characteristics that contribute to successful language technologies: **their generalizability, their efficiency, and their transparency**. Different representations and learning paradigms have different strengths, and in most cases, improving one aspect comes at the expense of another. For instance, neural approaches offer high generalizability, but require a lot of resources, and are hard to interpret. Rule-based systems are transparent, but struggle to generalize. Probabilistic graphical models are easier to interpret, but require solving computationally intractable constrained optimization problems. Building on the thesis that neural-symbolic approaches offer the right balance to counteract this trade-off, I have studied algorithmic frameworks and learning protocols that emphasize these three aspects.

Efficient Neural-Symbolic Methods: To learn parameters in DRaiL, I proposed a deep structured prediction framework that combines expressive textual encoders, relational embeddings and constrained inference [2]. The objective functions used in DRaiL involve solving or approximating the MAP inference problem. While tractable solutions exist for traditional NLP tagging tasks, dealing with more complicated structures and arbitrary declarative constraints comes at a high computational cost. To solve this challenge, I explored the use of randomized inference for deep structured models composed of expressive neural encoders, where theoretical guarantees are weak or nonexistent. I obtained competitive results at a fraction of the cost for a set of tasks involving complicated discourse structures [3].

Learning with Explanations: Identifying the reasoning steps taken to arrive to a decision is important to understand and judge the quality of predictive models. In a lot of cases the specific properties that explain our decisions are not readily available to us. This is especially the case in end-to-end deep learning models, where numerous, complex computations obfuscate the way in which the model reasons. Neural-symbolic representations provide us with a way to explicitly model emerging properties in a given domain. In cases where we do not have explicit supervision, we can use background knowledge to encode reasonable behaviors for these properties, and rely on sources of weak supervision to initialize them. By making properties explicit, we can also exploit human judgments to obtain and incorporate feedback.

Following this rationale, my collaborators and I explored the use of symbolic explanations to model collective behaviors in conversations. Specifically, we looked at the task of identifying *collaborative conversations* [7]. Collaborative conversations are defined as meaningful interactions, where there are open and respectful exchanges of information. We used discrete latent variables to explicitly model behaviors such as "idea development", "balanced contributions" and "engagement", and scored them using expressive language encoders. We connected these behaviors to observed information using a symbolic reasoning framework. Our solution resulted in performance improvements, while providing a natural way to explain the final decision. My current work applies this protocol to other domains, and looks to incorporate human feedback to refine the explanations and interactively debug the underlying model.

3 Applications to Discourse Analysis

Most of the work in neural-symbolic NLP focuses on applications such as word math problems, visual reasoning and question-answering. These domains are a natural fit for NS representations, as they usually deal with symbolic inputs and outputs (e.g. knowledge graphs and mathematical symbols), or concepts that are very concrete (e.g. shapes and colors). While unexplored, I argue that **discourse analysis is an excellent fit for NS representations**. On the one hand, discourse has inherent structure. NS representations allow us to explicitly model the interactions between participants in a conversation, or between characters in a story, using typed relations. On the other hand, they give us a common language to reason across modalities. For example, we could exploit the principle of social homophily, stating that people with strong social ties are likely to hold similar views, to reason about statements made by two individuals. Finally, we could use symbols to model high-level conceptual frameworks that can support our analysis, and learn expressive representations for them. I will discuss two relevant scenarios.

Open-domain stance prediction: Stance prediction in debates is traditionally thought of as a classification task. In this setup, instead of predicting stances on specific topics (e.g. *emotional support animals*), we focused on debate claims expressed in text that are specific to each debate (e.g. *Medical documentation of disability should be the entry point for service dog access*). This results in a challenging textual inference problem, where we need to account for a wide range of positions with respect to nuanced claims. I proposed a NS framework to enrich textual representations with author information, and constrain them using the conversational structure. In addition to this, I modeled social interactions between participants and their ideological affiliations to characterize the types of arguments that different groups make. Some of this supporting inferences were observed (e.g. conversation turns), some were learned (e.g. social ties), and some were latent (e.g. ideological affiliation). By doing this, I improved performance with respect to deep language models, graph neural networks, and competitive probabilistic logic approaches. Further, the resulting model provided an explanation through explicit relations (e.g. the resulting social graph), as well as the distributed representations learned for each symbol [2].

Identifying morality frames in political messaging: Moral Foundations Theory provides a theoretical framework for analyzing expressions of moral values by introducing five themes that are observed across cultures (e.g. care/harm, fairness/cheating). Recent work in NLP and Computational Social Science looks at identifying moral foundations in political messaging. However, these themes can be too coarse grained to capture the differences in messaging between different political groups. My collaborators and I made the observation that when different political groups use the same moral foundation, their moral sentiment is directed at different targets. In Example 1, both arguments are using the *fairness* MF, but the targets vary (*refugees* vs. *legal immigrants*). To capture this, we introduced moral foundation frames, a symbolic structure that makes the moral roles of entities explicit. We proposed a NS framework to jointly embed text and symbols, and explicitly encourage consistency in the sentiment towards entities within a single political party. Our approach allowed us to predict and explain individual messages, as well as to aggregate and summarize the arguments made by the different political parties.

4 Future Research Directions

My vision for future research emphasizes the role of humans in language technologies. While the goal of machine learning is to communicating knowledge from humans to machines, most current methods approach this through labeled examples. The bandwidth of interaction between humans and AI systems can be much greater. My work in neural-symbolic NLP investigates ways to characterize knowledge using intermediate concepts that can be shared across many learning problems. Approaching learning as a form of knowledge communication is a step towards bridging the gap between human experience and statistical modeling. My background and expertise put me in a unique position to collaborate effectively with researchers in human-computer interaction, cognitive science and the social sciences to pursue high impact, real-world applications that can benefit from the integration of structured human knowledge and large-scale language analysis.

Knowledge Intensive NLP: Most existing NLP technologies have been trained on readily available textual resources like news articles and documents scrapped from the web. Specialized resources like technical specifications, legal documents, medical and scientific literature, and historical texts are of great value to practical, high impact applications, but remain largely unexplored in NLP research. In contrast to news and web text, specialized documents are highly structured, use specialized vocabulary to describe concepts and procedures, and require domain expertise for their interpretation.

My collaborators and I studied the task of extracting information from technical documentation to support automated attack discovery in networking protocols [8]. One of the main challenges that we faced was the lack of relevant annotated resources. We also encountered ambiguities and unspecified behaviors in the text, which made end-to-end extraction infeasible. NS methods offer a lot of opportunities to deal with these challenges, as they can learn representations from large amounts of unlabeled information, while providing an interface to model expert domain knowledge and encode expectations, potentially allowing us to make sense of specialized data without extensive supervision.

Learning from Interaction: Human-in-the-loop approaches amplify the role of human experts in the process of learning and refining machine learning models. Most current human-in-the-loop protocols work directly over the space of inputs and their labels, for instance, by soliciting examples from people to augment the training data, or by disabling specific input features. While straightforward, this low-level representation does not take advantage of people's abilities to model concepts and higher-level abstractions. NS representations offer great opportunities to explain the state of machine learning models and encode human expertise beyond labeling. This is a promising direction for two reasons: 1) labeling is tedious, repetitive work, and previous studies have shown that uses prefer richer interaction protocols, and 2) higher-level abstractions have the potential to generalize to more scenarios, having a stronger impact in the model performance than adding or modifying a handful of training examples.

Multi-Modal Language Grounding: Language alone is not enough to convey all the information needed to communicate effectively. The ambiguities inherent to written and spoken words are complemented by their surrounding context and our shared understanding of the world. Context in the real world comes in multiple forms and modalities, such as auditory, visual, formal, spatial and social. While there has been increased interest in the NLP community to link concepts expressed in language to the external world, the majority of current studies focus on connecting language with vision. NS representations offer a common language to reason in a way that is agnostic of the underlying modality. Reasoning across multiple modalities is currently unexplored, particularly when incorporating abstract dimensions like the social world. I have the tools to explore NS representations to represent and reason about abstract concepts, and ground them across diverse domains and modalities.

[1] X. Zhang*, **M.L. Pacheco***, C. Li, and D. Goldwasser. Introducing DRAIL – a step towards declarative deep relational learning. In *EMNLP Workshop on Structured Prediction for NLP*, 2016.

[2] **M.L. Pacheco** and D. Goldwasser. Modeling content and context with deep relational learning. *Transactions of the Association for Computational Linguistics*, 2021.

[3] M. Widmoser*, **M.L. Pacheco***, J. Honorio, and D. Goldwasser. Randomized deep structured prediction for discourse-level processing. In *European Chapter of the Association for Computational Linguistics*, 2021.

[4] **M.L. Pacheco**, I. Dalal, and D. Goldwasser. Leveraging representation and inference through deep relational learning. In *NeurIPS Workshop on Relational Representation Learning*, 2018.

[5] I-T. Lee, **M.L. Pacheco**, and D. Goldwasser. Modeling human mental states with an entity-based narrative graph. In *North American Chapter of the Association for Computational Linguistics*, 2021.

[6] S. Roy, **M.L. Pacheco**, and D. Goldwasser. Identifying morality frames in political tweets using relational learning. In *Empirical Methods in Natural Language Processing*, 2021.

[7] A. Jain, **M.L. Pacheco**, S. Lancette, M. Goindani, and D. Goldwasser. Identifying collaborative conversations using latent discourse behaviors. In *Special Interest Group on Discourse and Dialogue*, 2020.

[8] **M.L. Pacheco**, M. von Hippel, B. Weintraub, D. Goldwasser, and C. Nita-Rotaru. Automated attack synthesis by extracting finite state machines from protocol specification protocols. In *IEEE Symposium on Security and Privacy (to appear)*, 2022.