1 Introduction

Semantic Role Labeling (SRL) is an important component for natural language understanding (Palmer et al. 2013). It identifies the semantic role of words, abstracting away superficial variations caused by ways of expression. Semantic dependencies (upper part of Figure 1) embody such information, specifying the semantic role label actor of the argument it to the predicate eliminate. Semantic dependency provides useful information for tasks such as question answering (Yih et al. 2014), automatic summarization (Jin et al. 2020; Kumar and Raghuveer 2012), text mining (Poria et al. 2014), and beyond. On the syntax side, a similar notion of syntactic dependencies (lower part of Figure 1) arise to encode grammatical relations. Syntactic dependency specifies the grammatical relation modifier of the dependent word hardly to the head word eliminate.

The two types of dependencies correlate heavily with each other, creating many parallelisms between them. Figure 1 shows an example of such parallelism, where pigeon is both an experiencer argument and an object dependent to the verb predicate eliminate. In general, words that are object dependents of a verb predicate often be an experiencer argument of the predicate. The parallelism motivates a large body of literature enhancing machine learning-based semantic
parsers with syntactic dependencies. The literature includes works such as He et al. (2018) and Roth and Lapata (2016). While those works lead to performance improvements, most are satisfied with the score uplift and avoid a more fundamental problem: what underlines the correlation between syntactic and semantic dependencies. In our paper (Chen et al. 2022), we study the statistical property underlying the dependency correlation.

We interpreted the dependency correlation as a shift in semantic label distributions. The label distribution (Dozat and Manning 2018) models the distribution over semantic role labels for a dependency spanning the predicate and the argument. A non-relation label indicates no semantic dependency. We found that the label distribution changes significantly with the hop patterns of the shortest syntactic dependency path (SSDP) connecting the predicate and the argument. The hop pattern reflects the length of syntactic dependencies. The distribution shift corresponds to our changing expectations for semantic dependencies. For example, we would have a high expectation for a semantic dependency given a short syntactic dependency but a low expectation given a long syntactic dependency. We modeled the distribution shift using a mixture model-based semantic parser, which we will explain in the next section. Compared to previous syntax-aware semantic parsing methods, modeling the distribution shift improves performance in predicting short-distance semantic dependencies while retaining the performance advantage in long-distance dependencies. Modeling the distribution shift also provides a small but significant performance uplift compared to syntax-aware semantic parsing baselines, as well as an SRL system competitive with state-of-the-art methods.

2 Modeling Distribution Shift with Mixture Models

In this section, we present analyses supporting our interpretation and explain the motivation behind the usage of the mixture model. We interpreted the dependency correlation as a label distribution shift with SSDP hop patterns. SSDP, a salient feature for exploiting the dependency correlation, is the shortest path connecting the predicate and the argument in the syntactic dependency structure (Figure 1). Hop patterns \((\alpha, \beta)\) count the number of transitions going from the predicate to the argument. \(\alpha\) counts the dependent-to-head transition that goes in the opposite direction as syntactic dependencies, while \(\beta\) counts the head-to-dependent transition going in the same direction as syntactic dependencies. The count reflects the syntactic distance between the predicate and the argument. We refer readers to the paper for more technical details about SSDP and hop patterns.

The left figure in Figure 2 provides a glimpse of the label distribution shift. Each vertical
bar shows a label distribution for the corresponding hop pattern. We see that long hop patterns share a similar label distribution dominated by the non-relation label, whereas short hop patterns have disparate label distributions. The (0, 1) pattern has the most outstanding label distribution among all patterns. The non-relation dominating distribution for long patterns doubles down on an old belief that semantic dependencies co-occur mainly with short syntactic dependencies. Additionally, the figure suggests a variation in label distributions for short patterns.

The two statistical properties motivate the introduction of the mixture model (right diagram in Figure 2). The large number of patterns sharing a similar distribution requires clustering those patterns. The variation in label distributions requires the separate modeling of distributions with unique properties. The mixture model method meets the two requirements. It automatically learns a cluster assignment for hop patterns, as indicated by the dashed arrow in Figure 2. Meanwhile, it learns the label distribution for each cluster, modeling the variation.

We confirm the label distribution shift using a mutual information analysis. In the analysis, we compute the mutual information for models aware of hop pattern information and models unaware of such information, and define the gap in mutual information value as *mutual information gain*. The left heatmap in Figure 3 shows the mutual information gain for each hop pattern. We see that long patterns have a near-zero information gain, whereas short patterns have various information gains. The (0, 1) pattern has the highest information gain.

The mixture model-based semantic parser learns a cluster assignment agreeing with the mu-
Fig. 3 Illustration of the mutual information analysis and learned cluster assignments. The left heatmap shows the mutual information gain value for hop patterns. The right table shows the cluster assignment extracted from a model. Hop patterns assigned to the same cluster are color-painted in the same color.

We see that the model assigns one cluster to long patterns sharing the non-relation dominating distribution, and assigns different clusters to short patterns. The model also assigns a unique cluster to the (0, 1) pattern, the pattern with the most outstanding label distribution and with the highest information gain. The parallelism between the label distribution visualization, the mutual information analysis, and the learned cluster assignment supports our interpretation that semantic label distributions shift with hop patterns.

3 Discussion

The abundance of parallelism between the two types of dependencies indicates a strong correlation between syntactic and semantic dependencies. The parallelism suggests that we can extract a large chunk of semantic dependencies using solely syntactic dependencies. It builds up a high expectation that syntactic dependencies provide plenty of information for extracting semantic dependencies.

Despite the strong correlation, the impact of syntactic information had decreased over time. Syntactic information changes from being indispensable (Punyakanok et al. 2008; Gildea and Palmer 2002) to being optional (He et al. 2017) for high-performance semantic parsers. The phenomenon is especially vocal in the neural era when the extra syntactic information provides a moderate boost or even does harm to model performance (Sachan et al. 2021).
between the expectation and the reality necessitates the study of the dependency correlation. More specifically, the statistical property underlines the dependency correlation. A deeper understanding of the statistical property would help us better utilize the correlation.

Our interpretation is a generalization over a widely-adopted co-occurrence bias. The bias suggests that semantic dependencies co-occur mainly with short syntactic dependencies and are unlikely to co-occur with long dependencies. We showed that the long hop patterns have zero mutual information gain and share a label distribution dominated by the non-relation label. Previous methods adopt this bias, focusing on the short dependencies rather than the long ones. He et al. (2018, 2019) discriminates short and long syntactic dependencies by heuristics, pruning semantic dependencies that co-occur with long syntactic dependencies. Cai and Lapata (2019) employs an auxiliary task predicting whether the word is a direct descendent or ascendant of predicates, implicitly discriminating between one-hop and multi-hop dependencies. Shi et al. (2020) encodes the dependency correlation via a joint syntactic-semantic label scheme designed specifically to handle semantic dependencies co-occurring with short syntactic dependencies.

Our mixture model method provides better interpretability than neural methods for utilizing the correlation. The mixture model clusters hop patterns with similar label distributions and models disparate distributions for the clusters. The cluster assignment provides insight for studying the statistical property of the correlation. In contrast, the neural methods encode the dependency correlation implicitly, offering little insight into its statistical properties. Roth and Lapata (2016) encode SSDPs as a continuous embedding using a Long-Short Term Memory model. They found the method learned to cluster semantic dependencies co-occurring with specific linguistic phenomena. However, the captured phenomena are fragmental and limited to a few syntactic constructions. This is because Neural encoders, such as the LSTM model, project SSDPs in a high-dimensional space. The complex structure of the space renders Euclidean-based clustering analysis less effective.

A better understanding of the correlation would provide insights for better utilization. Neural methods are notorious for their trial and error practice that tries all possible ways to model the correlation and reports only the successful one. On the one hand, the score uplift confirms the benefit of utilizing syntactic information in semantic dependency parsing. On the other hand, the practice reveals little nature about the correlation. In contrast, a better understanding of the correlation can guide the design of neural models. Our approach is an example of such a practice where the distribution shift motivates the introduction of the mixture model. Usage of the mixture model leads to an improvement in scores and, more importantly, confirms the distribution shift as a statistical property underlying the correlation.
However, it is to be noted that our interpretation is only a small step towards a deeper study of the correlation. The interpretation used only the hop pattern, a feature reflecting the length of syntactic dependencies. Many other factors, such as the syntactic dependency label, also play an important role in the correlation. A more comprehensive analysis is needed for a better understanding of the correlation.

Acknowledgement

The author would like to thank his supervisor, Yusuke Miyao, for the support and insightful suggestions in studying the dependency correlation. The author also wants to express gratitude to his labmates for valuable discussions.

References

Cai, R. and Lapata, M. (2019). “Syntax-aware Semantic Role Labeling without Parsing.” Transactions of the Association for Computational Linguistics, 7, pp. 343–356.

Chen, J., He, X., and Miyao, Y. (2022). “Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models.” In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 7959–7969, Dublin, Ireland. Association for Computational Linguistics.

Dozat, T. and Manning, C. D. (2018). “Simpler but More Accurate Semantic Dependency Parsing.” In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 484–490, Melbourne, Australia. Association for Computational Linguistics.

Gildea, D. and Palmer, M. (2002). “The Necessity of Parsing for Predicate Argument Recognition.” In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 239–246, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

He, L., Lee, K., Lewis, M., and Zettlemoyer, L. (2017). “Deep Semantic Role Labeling: What Works and What’s Next.” In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 473–483, Vancouver, Canada. Association for Computational Linguistics.

He, S., Li, Z., and Zhao, H. (2019). “Syntax-aware Multilingual Semantic Role Labeling.” In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and
He, S., Li, Z., Zhao, H., and Bai, H. (2018). “Syntax for Semantic Role Labeling, To Be, Or Not To Be.” In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2061–2071, Melbourne, Australia. Association for Computational Linguistics.

Jin, H., Wang, T., and Wan, X. (2020). “SemSUM: Semantic Dependency Guided Neural Abstractive Summarization.” In The 34th AAAI Conference on Artificial Intelligence, AAAI 2020, The 32nd Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The 10th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020, pp. 8026–8033. AAAI Press.

Kumar, V. R. and Raghuvieer, K. (2012). “Dependency Driven Semantic Approach to Product Features Extraction and Summarization Using Customer Reviews.” In Meghanathan, N., Nagamalai, D., and Chaki, N. (Eds.), Advances in Computing and Information Technology - Proceedings of the 2nd International Conference on Advances in Computing and Information Technology (ACITY) July 13–15, 2012, Chennai, India - Volume 3, Vol. 178 of Advances in Intelligent Systems and Computing, pp. 225–238. Springer.

Palmer, M., Titov, I., and Wu, S. (2013). “Semantic Role Labeling.” In NAACL HLT 2013 Tutorial Abstracts, pp. 10–12, Atlanta, Georgia. Association for Computational Linguistics.

Punyakanok, V., Roth, D., and Yih, W.-t. (2008). “The Importance of Syntactic Parsing and Inference in Semantic Role Labeling.” Computational Linguistics, 34 (2), pp. 257–287.

Roth, M. and Lapata, M. (2016). “Neural Semantic Role Labeling with Dependency Path Embeddings.” In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1192–1202, Berlin, Germany. Association for Computational Linguistics.

Sachan, D., Zhang, Y., Qi, P., and Hamilton, W. L. (2021). “Do Syntax Trees Help Pre-trained Transformers Extract Information?” In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pp. 2647–2661, Online. Association for Computational Linguistics.
Shi, T., Malioutov, I., and Irsoy, O. (2020). “Semantic Role Labeling as Syntactic Dependency Parsing.” In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7551–7571, Online. Association for Computational Linguistics.

Yih, W.-t., He, X., and Meek, C. (2014). “Semantic Parsing for Single-Relation Question Answering.” In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 643–648, Baltimore, Maryland. Association for Computational Linguistics.

**Junjie Chen**: Junjie Chen is currently pursuing his PhD degree at the University of Tokyo under the supervision of Yusuke Miyao. He has a broad interest across most areas of Natural Language Processing. Currently, he is actively researching about unsupervised parsing.