Recent Advances in Transferable Representation Learning

Muhao Chen[1,2], Kai-Wei Chang[1] and Dan Roth[2]
[1]Department of Computer and Information Science, UPenn
[2]Department of Computer Science, UCLA

1 Goal of Tutorial

This tutorial targets AI researchers and practitioners who are interested in applying deep learning techniques to cross-domain decision making tasks. These include tasks that involve multilingual and cross-lingual natural language processing, domain-specific knowledge, and different data modalities. This tutorial will provide audience with a holistic view of (i) a wide selection of representation learning methods for unlabeled text, multi-relational and multimedia data, (ii) techniques for aligning and transferring knowledge across multiple representations, with limited supervision, and (iii) a wide range of AI applications using these techniques in natural language understanding, knowledge bases, and computational biology. We will conclude the tutorial by outlining future research directions in this area.

Keywords: Representation learning, natural language processing, multi-relational data, knowledge transfer.

2 Tutorial Description

Many AI tasks require cross-domain decision making. For example, many NLP tasks involve predictions across multiple languages, in which different languages can be treated as different domains; in AI-aided biomedical study, the prediction of side effects of drugs is often in parallel to modeling the interactions of proteins and organisms. To support machine learning models to solve such cross-domain tasks, a requisite is to extract the characteristics and relations of data components in different domains, and capture their associations in a unified representation scheme. Towards such a demand, recent advances of representation learning often involve mapping unlabeled data of different domains into shared embedding spaces. In such a way, cross-domain knowledge transfer can be realized by vector collocation or transformations. Such transferable representations have seen successes in a range of AI applications involving cross-domain decision making. However, frontier research in this area faces two key challenges. One is to efficaciously extract features from specific domains with very few learning resources. The other is to precisely align and transfer knowledge with minimal supervision, since the alignment information that connects between different domains can often be insufficient and noisy.

In this tutorial, we will comprehensively review recent developments of transferable representation learning methods, with a focus on those for text, multi-relational and multimedia data. Beyond introducing the intra-domain embedding learning approaches, we will discuss various semi-supervised, weakly supervised, multi-view and self-supervised learning techniques to connect multiple domain-specific embedding representations. We will also compare retrofitting and joint learning processes for both intra-domain embedding learning and cross-domain alignment learning. In addition, we will discuss how obtained transferable representations can be utilized to address low-resource and label-less learning tasks. Participants will learn about recent trends and emerging challenges in this topic, representative tools and learning resources to obtain ready-to-use models, and how related models and techniques benefit real-world AI applications.

3 Outline of Tutorial Content

This tutorial presents a systematic overview of frontier approaches to transferable representation learning. We will begin with introducing the main research problems, and motivate this topic with several real-world applications. Then, we will introduce methods for learning embedding for each domain and approaches to align embeddings across domains. Specifically, we will discuss how comprehensive information such as linguistic contexts and probabilistic soft logic rules can be leveraged to improve the quality of the embedding, and how to align embeddings in different domains by strategies such as retrofitting, joint learning, and self-supervised learning. Moreover, we will exemplify the use of these embeddings in applications of various areas, and will outline emerging research challenges that may catalyze further investigation on this topic. The detailed contents are outlined below.

Motivation [30min]

We will define the main research problem and motivate the topic by presenting several AI applications requiring joint representation learning.

Retrofitting, Joint Learning and Self-supervised Alignment Learning of Embeddings [45min]

We will first provide the general overview of embedding learning methods for structured and unstructured data in different domains. On top of that, we will discuss about how domain-specific embedding spaces can be associated using retrofitting or joint learning methods. In addition, we will also cover recent self-supervised learning processes for cases where alignment information for knowledge transfer is not directly available (Pires, Schlinger, and Garrette 2019).

Transferable Representation Learning for Multilingual Natural Language Processing [45min]

We will discuss how transferable representations are incorporated into various multilingual NLP tasks. Taking dependency parsing (Ma et al. 2018; Ahmad et al. 2019; Tao, Peng, and Chang 2019) and entity linking systems (Tsai, Mayhew, and Roth 2016; Upadhyay, Gupta, and Roth 2018; Mayhew and Roth 2018) as examples, we will demonstrate how knowledge transfer allows NLP models trained on high-resource languages to be transferred to low-resource language tasks.

Joint Representation Learning on Multi-relational Data [60min]

We will present recent research efforts in a number of tasks that apply joint representation learning on multi-relational data. From the methodology perspective, we will discuss how entities can be aligned based on multi-view learning on diverse entity profiles in different modalities (Chen et al. 2018; Sun, Hu, and Li 2017; Zhang et al. 2019; Vulić et al. 2016), and how rank-based boosting techniques (Freund et al. 2004) can be deployed to integrate knowledge from multiple and possibly inconsistent views. From the application perspective, we will investigate representative systems that utilize transferable multi-relational embeddings to address knowledge base integration (Chen et al. 2017), entity typing (Hao et al. 2019), protein-protein interaction prediction (Chen et al. 2019b; 2019a) and polypharmacy side effect identification (Zitnik, Agrawal, and Leskovec 2018; Xuan et al. 2019). Moreover, we will discuss how the obtained relational embeddings can be utilized to address label-less learning tasks (Lee et al. 2018).

Conclusion and Future Research Directions [30min]

Transferable representation learning impacts on a wide spectrum of data-driven and knowledge-driven AI tasks. We will conclude the tutorial by presenting some challenges and potential research topics in designing transferable representation learning models for data with complex relational properties and diverse modalities.

4 History, Expected Attendance and Prerequisite Knowledge

This tutorial has not been presented elsewhere. Based on the level of interest in this topic, we expect around 50∼70 participants. No specific background knowledge is assumed of the audience.

5 Tutorial Instructors

Muhao Chen is currently a postdoctoral fellow in Department of Computer and Information Science, UPenn. He received his Ph.D. in Computer Science from UCLA in 2019, where he was supported by the Dissertation Fellowship to work on his thesis research about multi-relational representation learning and knowledge acquisition. Muhao has worked on various topics in machine learning and natural language processing, including relational learning, sequence modeling, lexical semantics and graph representation learning. His recent research also applies machine learning techniques to computational biology. He has published over 30 papers in leading venues of these areas. Additional information is available at `http://muhaochen.github.io`.

Kai-Wei Chang is an assistant professor in the Department of Computer Science at the University of California Los Angeles. His research interests include designing robust machine learning methods for large and complex data and building language processing models for social good applications. Chang has published broadly in machine leanring, natural language processing, and artificial intelligence conferences. His awards include the EMNLP Best Long Paper Award (2017), the KDD Best Paper Award (2010), and the Okawa Research Grant Award (2018). Kai-Wei has given tutorials at NAACL, FAT, AAAI, EMNLP on various research topics. Additional information is available at `http://kwchang.net`.

Dan Roth is the Eduardo D. Glandt Distinguished Professor at the Department of Computer and Information Science, University of Pennsylvania, and a Fellow of the AAAS, ACM, AAAI, and the ACL. In 2017 Roth was awarded the John McCarthy Award, the highest award the AI community gives to mid-career AI researchers. Roth was recognized for major conceptual and theoretical advances in the modeling of natural language understanding, machine learning, and reasoning. Roth has published broadly in machine learning, natural language processing, knowledge representation and reasoning, and learning theory, and has developed advanced machine learning based tools for natural language applications that are being used widely. Roth has given tutorials on these and other topics in all ACL and AAAI major conferences. Until February 2017 Roth was the Editor-in-Chief of the Journal of Artificial Intelligence Research (JAIR). He was the program chair of AAAI'11, ACL'03 and CoNLL'02, and serves regularly as an area chair and senior program committee member in the major conferences in his research areas. Prof. Roth received his B.A Summa cum laude in Mathematics from the Technion, Israel, and his Ph.D. in Computer Science from Harvard University in 1995. Additional information is available at `http://www.cis.upenn.edu/~danroth/`.

The presenters of this tutorial have given the following tutorials at leading international conferences and venues:

- Kai-Wei Chang:

- EMNLP '19: Conference on Empirical Methods in Natural Language Processing. A tutorial on Bias and Fairness in Natural Language Processing.
- ACM FAT*'18: ACM Conference on Fairness, Accountability, and Transparency. A tutorial on Quantifying and Reducing Gender Stereotypes in Word Embeddings.
- TAAI '17: Conference on Technologies and Applications of Artificial Intelligence. A tutorial on Structured Predictions: Practical Advancements and Applications in Natural Language Processing.
- AAAI'16: The AAAI Conference on Artificial Intelligence. A tutorial on Learning and Inference in Structured Prediction Models.
- NAACL'15, The North American Conference of the Association on Computational Linguistics. A tutorial on Hands-on Learning to Search for Structured Prediction.

• Dan Roth:

- Data Science Summer Institute (DSSI) 2007, 2008, 2010, 2011, 2012. A tutorial on Machine Learning in Natural Language Processing.
- ACL'18, The Conference of the Association on Computational Linguistics. A tutorial on Multi-lingual Entity Discovery and Linking.
- EACL'17, The European Conference of the Association of Computational Linguistics; A tutorial on Integer Linear Programming Formulations in Natural Language Processing.
- AAAI'16, The Conference of the Association for the Advancement of Artificial Intelligence; A tutorial on Structured Prediction.
- ACL'14, The International Conference of the Association on Computational Linguistics. A tutorial on Wikification and Entity Linking.
- AAAI'13, The AAAI Conference on Artificial Intelligence. Information Trustworthiness.
- COLING'12, The International Conference on Computational Linguistics. A Tutorial on Temporal Information Extraction and Shallow Temporal Reasoning.
- NAACL'12, The North American Conference of the Association on Computational Linguistics. A Tutorial on Constrained Conditional Models: Structured Predictions in NLP.
- NAACL'10, The North American Conference of the Association on Computational Linguistics. A Tutorial on Integer Linear Programming Methods in NLP.
- EACL'09, The European Conference of the Association on Computational Linguistics. A Tutorial on Constrained Conditional Models.
- ACL'07, The International Conference of the Association on Computational Linguistics. A Tutorial on Textual Entailment.

References

Ahmad, W.; Zhang, Z.; Ma, X.; Hovy, E.; Chang, K.-W.; and Peng, N. 2019. On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing. In *NAACL*, 2440–2452.

Chen, M.; Tian, Y.; Yang, M.; and Zaniolo, C. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *IJCAI*.

Chen, M.; Tian, Y.; Chang, K.-W.; Skiena, S.; and Zaniolo, C. 2018. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. *IJCAI*.

Chen, M.; Ju, C.; Zhou, G.; Chen, X.; Zhang, T.; Chang, K.-W.; Zaniolo, C.; and Wang, W. 2019a. Multifaceted protein-protein interaction prediction based on siamese residual rcnn. *Bioinformatics* 35(14):i305–i314.

Chen, X.; Chen, M.; Shi, W.; Sun, Y.; and Zaniolo, C. 2019b. Embedding uncertain knowledge graphs. In *AAAI*.

Freund, Y.; Iyer, R.; Schapire, R. E.; and Singer, Y. 2004. RankBoost: An efficient boosting algorithm for combining preferences. *JMLR* 4(6):933–969.

Hao, J.; Chen, M.; Yu, W.; Sun, Y.; and Wang, W. 2019. Universal representationlearning of knowledge bases by jointly embedding instances and ontological concepts. In *KDD*.

Lee, C.-W.; Fang, W.; Yeh, C.-K.; and Frank Wang, Y.-C. 2018. Multi-label zero-shot learning with structured knowledge graphs. In *CVPR*, 1576–1585.

Ma, X.; Hu, Z.; Liu, J.; Peng, N.; Neubig, G.; and Hovy, E. 2018. Stack-pointer networks for dependency parsing. In *ACL*, 1403–1414.

Mayhew, S., and Roth, D. 2018. Talen: Tool for annotation of low-resource entities. In *ACL, demo*, 80–86.

Pires, T.; Schlinger, E.; and Garrette, D. 2019. How multilingual is multilingual bert? In *ACL*.

Sun, Z.; Hu, W.; and Li, C. 2017. Cross-lingual entity alignment via joint attribute-preserving embedding. In *ISWC*, 628–644.

Tao, M.; Peng, N.; and Chang, K.-W. 2019. Target language-aware constrained inference for cross-lingual dependency parsing. In *EMNLP*.

Tsai, C.-T.; Mayhew, S.; and Roth, D. 2016. Cross-lingual named entity recognition via wikification. In *CoNLL*, 219–228.

Upadhyay, S.; Gupta, N.; and Roth, D. 2018. Joint multilingual supervision for cross-lingual entity linking. In *EMNLP*, 2486–2495.

Vulić, I.; Kiela, D.; Clark, S.; and Moens, M.-F. 2016. Multi-modal representations for improved bilingual lexicon learning. In *ACL*, 188–194.

Xuan, P.; Cao, Y.; Zhang, T.; Wang, X.; Pan, S.; and Shen, T. 2019. Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics*.

Zhang, Q.; Sun, Z.; Hu, W.; Chen, M.; Guo, L.; and Qu, Y. 2019. Multi-view knowledge graph embedding for entity alignment. In *IJCAI*.

Zitnik, M.; Agrawal, M.; and Leskovec, J. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34(13):i457–i466.