10701 Spring 2018 Project

- TA contact: Shaojie Bai (shaojieb@andrew.cmu.edu)

1. Overview

The class project is an opportunity for you to explore an interesting problem of your choice in the context of a real-world data set. You can either choose one of the suggested projects we provided, or pick your own topic. Do not hesitate to discuss your project with TAs/instructors to get feedback on your ideas.

2. Group Size

$2 \leq \textbf{group size} \leq 3$

3. Requirement

- The core components of your implementation (if any) should not rely on any open-source code (e.g. GitHub). (In other words, you are free to use packages, etc., but you cannot simply duplicate other people's code for their project).
- In your deliverables (proposal, milestone report, final report, poster), you should clearly define and describe the problem. Cite your references. All reports must be in NIPS format.
- Plan your project accordingly. You shouldn't expect this to be a trivial assignment. Good projects usually start with a (perhaps small but) good/novel idea. Some examples of the topics are shown in the "Suggested Topics" section below.

4. Timeline

- Project Proposal due: Mar. 26
    - This should be a short description of the project, including: a title, andrew ids of the members, and a short description of your proposed project (about a page, excluding references). You **don't** need an abstract.
- Project Milestone Report Due: Apr 11
    - More detailed introduction of the project, review of related works, details of the proposed method, challenges you've encountered, next steps, and preliminary results if available. The milesstone report should have 4-5 pages (excluding references).
- Tentative Project Final Report due: May 02 (subject to change)
    - A full academic paper, including: problem definition and motivation, background and related work, details of the proposed method, experiments setup and results, conclusion and future work. 8 pages excluding references and appendix.

- Project Poster Presentations: Apr 30
    - Present your work to the peers, instructors, and other community members who will stop by. Location and time TBA.

5. Suggested Topics

> Feel free to choose other interesting topics as well :-)

Optimization's Untold Gift to Learning

In this project, we'll explore a recent line of work which studies the statistical biases introduced by different optimization algorithms. This has been of great interest wrt deep learning, where different optimization algorithms converge to different solutions, and each solution has different generalization properties. However, even in simpler learning settings, such as OLS, Logistic Regression, this question is unanswered. Recent works [1,2] have tried to answer this question in the context of underdetermined least squares and logistic regression for gradient descent. The goal of the project is two-fold; firstly, to replicate and understand the works of [1] and [2]. The second goal is to try to extend the results to a more general setting by drawing connections(empirically) between mirror descent[3] on likelihood functions and optimizing regularized likelihoods. Note that the second goal is a bit ambitious, hence, doing a good job on the first part will also be considered as a success.

TA: Adarsh

[1] The Implicit Bias of Gradient Descent on Separable Data (https://arxiv.org/abs/1710.10345)

[2] The Marginal Value of Adaptive Gradient Methods in Machine Learning (http://papers.nips.cc/paper/7003-the-marginal-value-of-adaptive-gradient-methods-in-machine-learning.pdf)

[3] Mirror Descent (https://blogs.princeton.edu/imabandit/2013/04/16/orf523-mirror-descent-part-iii/)

Testing for Differences in Gaussian Graphical Models

Graphical models have gained significant attention as a tool for discovering and visualizing dependencies among variables in multivariate data. Recently, attention has been drawn to situations where domain experts are interested in differences between the dependency structures of different populations. (e.g. differences between regulatory networks of different species, or differences between dependency networks of diseased versus healthy populations). In this work, we study and develop estimators which learn the differential network. For this, we begin by replicating the work of [1]. After that, the goal will be to compare and contrast(empirically) to other estimators such as [2,3].

TA: Adarsh

[1] Testing for Differences in Gaussian Graphical Models: Applications to Brain Connectivity (https://arxiv.org/pdf/1512.08643.pdf)

[2] Song Liu, Taiji Suzuki, and Masashi Sugiyama. Support consistency of direct sparse-change learning in markov networks

[3] Sihai Dave Zhao, T Tony Cai, and Hongzhe Li. Direct estimation of differential networks. Biometrika,

Movie QA

Watching a movie and understanding the story behind is a relaxing activity for humans, but it extremely difficult for machines. Creating an AI agent that can answer multiple choice questions from movie segments consisting on video+subtitles is an ongoing challenge [1] (see the MovieQA challenge). Current state-of-the-art solutions mostly rely on different types of attention models [2] and memory networks [3]. However, we can see from the leaderboard of this challenge that the question answering accuracy obtained on video+subtitles is very close to the accuracy of methods that use the subtitles only. This suggests that the current methods are not good at leveraging well both sources of information.

Can we design a model that is better at combining information from video and subtitles, and can answer better multiple choice questions about movie segments?

TA: Otilia

[1] Tapaswi, Makarand, et al. "Movieqa: Understanding stories in movies through question-answering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[2] Nam, Hyeonseob, Jung-Woo Ha, and Jeonghee Kim. "Dual attention networks for multimodal reasoning and matching." arXiv preprint arXiv:1611.00471 (2016).

[3] Na, Seil, et al. "A Read-Write Memory Network for Movie Story Understanding." arXiv preprint arXiv:1709.09345 (2017).

Solving algebra problems with Machine Learning

Machine Learning has an immense potential in improving the way education works. One direction is to create AI agents that can solve math problems, and then explain the solution to a human student. There are ongoing efforts at this task (see for example Project Euclid). However it is often the case that models that can solve problems well, cannot explain them well, and models that can explain the sequence steps to reach the solution, can only do so for very simple operations (e.g. addition, subtraction, multiplication) [2, 3], or are not guaranteed to reach the right answer [4].

Can we design a model that is the best of both worlds, and is better both at solving problems and explaining the solution?

TA: Otilia, Dimitris

[1] Kushman, Nate, et al. "Learning to automatically solve algebra word problems." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. 2014.

[2] Hosseini, Mohammad Javad, et al. "Learning to solve arithmetic word problems with verb categorization." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014.

[3] Roy, Subhro, and Dan Roth. "Solving general arithmetic word problems." arXiv preprint arXiv:1608.01413 (2016).

[4] Ling, Wang, et al. "Program induction by rationale generation: Learning to solve and explain algebraic word problems." arXiv preprint arXiv:1705.04146 (2017).

Understanding how the brain represents meaning

The way the human brain represents the meaning of words is still an open question in neuroscience. However, machine learning is now offering us new tools to approach this problem. In [1], the authors design a model that predicts how the brain responds (using functional magnetic resonance imaging (fMRI)) when a person is reading a noun. This model not only can predict well the brain response, but analyzing its parameters allows us to investigate which parts of the brain respond to different concepts (e.g. which brain regions respond to concepts related to eating, or related to movement etc.). The data used in this experiment is available here.

Can you design other interpretable models that can do these predictions better? What other interesting questions about the brain can you answer using your model?

There are several directions that can be attempted:

- Consider using different word embeddings, that still allow for the interpretation of the results.
- [1] train a different model per human participant. You may consider training a single model for all participants, or train different models for different participants together (e.g. in a multi-task learning [2] fashion).
- Consider other machine learning models that can predict better the fMRI response, while still allowing us to interpret the results.

TA: Otilia

[1] Mitchell, Tom M., et al. "Predicting human brain activity associated with the meanings of nouns." science 320.5880 (2008): 1191-1195.

[2] Zhang, Yu, and Qiang Yang. "A survey on multi-task learning." arXiv preprint arXiv:1707.08114 (2017).

Text Summarization

We are exposed to sources to stay informed about our society everyday - from newspaper articles and Facebook posts to academic papers. What if we are able to generate reasonable summaries on long texts, and still be able to deliver coherent and understandable contents. In order to achieve such a task, our machines will need to understand the text, extract features that characterize the text, and to generate a summary. Currently, there are interesting research projects that utilizes the Restricted Boltzmann Machine to extract meaningful features from texts, while others approach the problem using Sequence to Sequence Learning. This task emphasizes on both features extraction techniques and text generation, projects can be focused on thinking about different ways to build a language model, and not restricted to using deep learning techniques. Other relating projects include image captioning if you are interested in language modeling in general.

TA: Lam

[1] Text Summarization Using Unsupervised Deep Learning - Mahmood Yousefi-Azar and Len Hamey

[2] Extractive Summarization using Deep Learning - Sukriti Verma and Vagisha Nidhi

[3] Sequence to Sequence Learning with Neural Networks - Ilya Sutskever, Oriol Vinyals, Quoc V. Le

Information Retrieval using Probabilistic Graphical Models

Even though we are exposed to information sources online, one concern with using the information from online is the quality of such content and the credibility of user-contributed content. Previously, such a credibility analysis task is merely based on assumptions about structures of relational databases or features of text classification. However, can we leverage the interactions and dynamics of online communities? Currently, there are projects focusing on using models based on Conditional Random Fields to incorporate partial expert knowledge for semi-supervised learning. Other project ideas include building generative models based on Hidden Markov Model and Latent Dirichlet Allocation to capture user expertise, and identify expert users and credible content sources.

TA: Lam

[1] Probabilistic Graphical Models for Credibility Analysis in Evolving Online Communities - Subhabrata Mukherjee

[2] Latent Credibility Analysis Jeff Pasternack, Dan Roth

[3] Information Credibility: A Probabilistic Graphical Model for Identifying Credible Influenza Posts on Social Media - Qiaozhen Guo, Wei (Wayne) Huang, Kai Huang , and Xiao Liu

[4] Social Computing for Personalization and Credible Information Mining using Probabilistic Graphical Models - Jun Zou

Music Generation

Compared to text, music is a domain where many more combinations are possible. For instance, a piano has 88 keys, which means there could be up to 2^{88} different combinations of keys to press. Moreover, music has some interesting properties , such as melody and harmony, rests and all chords, which makes composing music a very interesting and yet challenging task. Music generation requires generative modeling. Some prior methodologies include the usage of Restricted Boltzmann Machine (RBM) and Recurrent Neural Networks (RNNs), but you are welcomed to introduce and explore other ways (e.g. GAN, VAE, etc.)! Moreover, besides using existing datasets (warning: some of which may be too small for generation tasks), you can also prepare and process your own music dataset from midi files, and train a model on a specific music style (e.g. jazz, classic, etc.).

TA: Shaojie

[1] Deep Jazz http://www.asimovinstitute.org/analyzing-deep-learning-tools-music/

[2] Allan Huang, Raymond Wu. "Deep Learning for Music"

[3] Vasanth Kalingeri, Srikanth Grandhe. "Music Generation Using Deep Learning" https://arxiv.org/abs/1612.04928

[4] Jean-Pierre Briot, Gaetan Hadjeres, Francois Pachet. "Deep Learning Techniques for Music Generation--- A Survey" https://arxiv.org/pdf/1709.01620

Image Captioning

The problem is about generating captions for real-world images. Given an image, the task is to generate a caption which briefly explains the image. It is a multi-modal problem as it involves language and vision data. Most of the proposed solutions follow a seq-2-seq neural model trained over datasets with a lot of image-captions datasets such as CLEVR and MS-COCO.

TA: Satya, Dimitris

[1] Vinyals, Oriol, et al. "Show and tell: A neural image caption generator." *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015

[2] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. (2015, June). Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning* (pp. 2048-2057).

[3] Rennie, Steven J., et al. "Self-critical sequence training for image captioning." *arXiv preprint arXiv:1612.00563* (2016).

Natural Language Understanding

The problem is to generate representations of natural language which can be interpreted by machines for analysis. This is very similar to semantic parsing in which we create Machine Readable Language (MRL) from user utterances. The solutions so far uses supervised deep learning models trained over datasets with samples of the form "(utterance, logical form)", such as ATIS and GeoQuery, These datasets have around 6-10k samples which are not sufficient to train a deep network and makes this problem very challenging.

TA: Satya

[1] Fan, Xing, et al. "Transfer learning for neural semantic parsing." *arXiv preprint arXiv:1706.04326* (2017).

[2] Grefenstette, E., Blunsom, P., de Freitas, N., & Hermann, K. M. (2014). A deep architecture for semantic parsing. *arXiv preprint arXiv:1404.7296*

[3] Aghaebrahimian, Ahmad, and Filip Jurcícek. "Machine learning for semantic parsing in review." Proceedings of 7th Language and Technology Conference, Poznan, Poland. 2015.

Neural Networks for Multi-view Learning across Images and Text

The problem of image/scene understanding is an important and challenging one. Often images are accompanied with descriptions that describe them. This is an important in image search. Many multi-view learning approaches have been proposed that extract features for both sentences and images, and map them to the same semantic embedding space. These methods are used to address multiple tasks such as retrieving the sentences given the query image, retrieving the images given the query sentences, generating captions that describe image scenes, etc.

Problem 1: The first proposed problem is to link objects in the images to appropriate mentions in the captions. We reason about which particular object each noun/pronoun in the captions is referring to in the image. This could potentially allow us to jointly model the textual and visual information to disambiguate the coreference resolution problem within and across images and texts. Towards this goal, one could explore deep-learning or structure prediction models that exploit features computed from text and RGB-D imagery to reason about the class of the 3D objects, the scene type, as well as to align the nouns/pronouns with the referred visual objects.

[1] C. Kong et. al. What are you talking about? Text-to-Image Coreference. CVPR 2014.

Problem 2: The first proposed problem is to recognize what appears in images while incorporating knowledge of spatial relationships and interactions between objects and some background knowledge (knowledge of how the world works - e.g. books are placed on a table - usually not under it). Another challenge here is in generating a description that is not only relevant but also grammatically correct, thereby, requiring a model for language. In this project, one could explore integrating recursive deep learning methods for image understanding either with existing language models or other neural networks that learn a language model.

TA: Sreena

[1] H. Fao et. al. From Captions to Visual Concepts and Back.

[2] R. Kiros et. al. Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models.

[3] A. Karpathy and Li Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions.

[4] http://googleresearch.blogspot.com/2014/11/a-picture-is-worth-thousand-coherent.html

[5] J. Donahue et. al. Long-term Recurrent Convolutional Networks for Visual Recognition and Description.

[6] J. Mao et. al. Explain Images with Multimodal Recurrent Neural Networks

[7] http://blogs.technet.com/b/machinelearning/archive/2014/11/18/rapid-progress-in-automatic-image-captioning.aspx

[8] http://mscoco.org/dataset/#download

Semantic Segmentation for Images

Semantic segmentation associates one of the pre-defined class labels to each pixel of an image. The input image is divided into the regions, which correspond to the objects of the scene or stuff. To perform a semantic segmentation of an image is to infer the semantic label for every pixel. Using simple semantic labels, the pixels in the image have been explained, each one generated by some unknown model for the category label. If such a segmentation can be achieved, then the image can be catalogued for image search, used for navigation, or any number of other tasks which require basic semantic understanding of arbitrary scenes. A wide range of machine learning techniques, including convolutional neural network, graphical models, and spectral methods etc., have been extensively employed in this interesting task. In this task, you need to investigate existing methods/models, evaluation metrics, public dataset for supervised semantic segmentation tasks, and then propose your solution for image semantic segmentation, and evaluate it on standard datasets.

TA: Sreena

[1] Fully Convolutional Networks for Semantic Segmentation. CVPR 2015

[2] Semantic Segmentation using Regions and Parts. CVPR 2012

[3] Recurrent Convolutional Neural Networks for Scene Labeling. ICML 2014

[4] Dilated Convolutions. ICLR 2016