COURSE PROJECT[1]

10-708 PROBABILISTIC GRAPHICAL MODELS (SPRING 2021)
http://708.mlcourse.org

1 Overview

The course project for 10-708 *Probabilistic Graphical Models* provides a chance to explore a real-world application of graphical models. Successful completion of the project will require the following steps: (1) Careful identification of a task and dataset that permits modeling of multiple correlated output variables. (2) The establishing of a baseline approach with appropriate metrics. (3) Design of an experimental setup that involves the comparison of multiple approaches to either modeling, learning, or inference in a controlled setting. (4) Summarizing and reporting of your findings via the following milestones:

1. Team Formation [3-4 people] (§2.1)
2. Project Proposal [1-2 pages] (§2.2)
3. Midway Poster (§2.3)
    - Midway Poster PDF Submission
    - Midway Poster Presentation
4. Midway Executive Summary [3 pages] (§2.4)
5. Final Poster (§2.5)
    - Final Poster PDF Submission
    - Final Poster Presentation
6. Final Executive Summary [4 pages] (§2.6)

Below, we detail these milestones (§2), list ideas for datasets/tasks (§3), and provide examples of project ideas (§4). The grading breakdown for these components is as follows: Team Formation 5%, Project Proposal 15%, Midway Poster 10%, Midway Executive Summary 20%, Final Poster 20%, Final Executive Summary 30%.

[1]Compiled on Tuesday 18th May, 2021 at 20:37

2 Milestones

2.1 Team Formation

Each team will consist of 3-4 people. Teams must be specified in advance of the proposal deadline.

The deliverable for this milestone will be a correct **group submission** of a PDF containing the names of the team members to Gradescope. To receive full credit, the group must submit only one PDF associated to the usernames of all the members of the group. See Gradescope instructions for adding group members at the link below.

https://www.gradescope.com/help#help-center-item-student-group-members

(We **do not** want, for example, multiple PDFs submitted by the same group, or a single submission that leaves off one or more members of the group—both of these would be considered incorrect submissions.)

2.2 Project Proposal

The guidelines for the project proposal are as follows:

1. **Overview:** The proposal will describe the task, dataset, proposed methods, and related work. The proposal should be 1-2 pages (excluding your references/bibliography).
2. **Contents:** Your proposal should be organized as follows:
   - *Title box:* Project title and list of group members' names and Andrew emails.
   - *Section 1: Dataset and Task.* Describe the dataset you will use, a precise definition of the task (highlighting why it is a structured prediction problem), and what metric(s) you will use for evaluation.
   - *Section 2: Approach.* First describe a baseline approach that you will implement—this could, for example, be a simple reduction to classification. Next, describe your main idea: include a description of the model, inference, and learning methods you will use. You should include some comparison between two or more competing methods, i.e. by varying the model, inference, or learning in some way.
   - *Section 3: Related Work.* Provide a short literature survey of 5 or more relevant papers. You should include papers that are relevant in two senses: (1) describe other methods that have been used for your task and (2) describe other papers that use similar methods.
   - *Section 4: Expected Outcomes.* Describe the experiments you will run and the results you expect (or hope) to get at the end. (In the end, it's fine if your results do not match these expectations—this is research after all—but it's always good to articulate your hypothesis.)
   - *Section 5: Plan.* Identify how you will break up the work between each of the team members. (We strongly encourage you to consider pair programming or the like for the most important aspects of the implementation.) Identify what you plan to have completed by the Midway Poster deadline.

   (If the core contribution of your project will be theoretical, you should follow a similar organization to that described above replacing descriptions of methods/experiments with discussion of expected theoretical results.)
3. **Pre-existing Work:** Note that you are expected to implement the main methods described in Section 2 of your proposals from scratch. If you wish to rely on pre-existing open source code for your baseline model, upon which your main methods can be built, you must clearly identify it in Section 5 of your proposal (include a link to the code). (As an example of this, you might wish to use a pre-existing implementation of ResNet as a baseline model, but add some interesting inference layer atop it that you will implement from scratch.)

   You are not permitted to use work you began prior to this course as your course project. If you wish to build on a task or dataset that you previously worked on (e.g. as a baseline), you should clearly identify which parts of the work were done ahead of time in Section 5.
4. **Submission:** Your project proposal PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

2.3 Midway Poster

Below are the guidelines for the midway poster milestone:

1. **Overview:** The midway poster offers each group a chance to present their progress halfway through the project's duration. You should think of the midway poster session as a more in-depth version of your project proposal. Students are required to attend the midway poster session.

2. **Contents:** A suggested organization for your midway poster is below. However, you are welcome to deviate from this organization, so long as you include all the relevant content.

   - *Title box:* Project title and list of group members' names and Andrew emails.
   - *Section 1: Dataset and Task.* Describe the dataset you will use, a precise definition of the task (highlighting why it is a structured prediction problem), and what metric(s) you will use for evaluation. You *must* include a few actual training examples (e.g. for a computer vision task this could be an actual image and its labels).
   - *Section 2: Approach.* First describe a baseline approach that you will implement—this could, for example, be a simple reduction to classification. Next, describe your main idea: include a description of the model, inference, and learning methods you will use. You should include some comparison between two or more competing methods, i.e. by varying the model, inference, or learning in some way. The methods section should be (nearly) in its final form. Be sure to graphically represent your approaches in some way (e.g. a neural net diagram, a graphical model, perhaps tied back to the training example). As appropriate you might want to include the mathematical details of your approach; this typically enables you to be more precise than a diagram, but is generally more difficult to describe.
   - *Section 3: Related Work.* Provide a short literature survey of 5 or more relevant papers. You should include papers that are relevant in two senses: (1) describe other methods that have been used for your task and (2) describe other papers that use similar methods. The goal here is to provide context to your audience and also to enable you to compare / contrast your method/-task with existing work. Again, showing a diagram of how existing methods compare to your methods is likely more effective than text.
   - *Section 4: Experiments.* The experiments section should include a precise description of the experiments that will be run. You are required to present results of a baseline model on your task of interest.

     Any other results you have may also be reported. If you have no results yet, that's fine, instead describe the results you expect (or hope) to get at the end. (In the end, it's fine if your results do not match these expectations—this is research after all—but it's always good to articulate your hypothesis.)

     Skeleton tables and plots are *required*. One of the cleanest ways to communicate your desired outcomes is to present us with skeleton tables / plots with blank spaces where you plan to fill in future numbers / plot elements.
   - *Section 5: Plan.* Identify how you will break up the work between each of the team members. (We strongly encourage you to consider pair programming or the like for the most important aspects of the implementation.) Identify the key milestones you need to hit prior to the final poster deadline. List the dates when you expect to complete each milestone and the team members who will be involved in each. In the event that your poster is not quite as polished as you would like,

list out how you plan to visually improve on the design / layout of the poster—we would rather see a mediocre poster with a clear plan for improvement, than a mediocre poster that is unaware of its own flaws.

3. **Poster Design:** Posters are a different medium than papers. They should convey a clear message in an engaging way. For the course project, you need to convey the problem at hand, the limitations of existing methods (briefly) and how you solved the problem, as well as some high level conclusions. Start by designing the poster on a sheet of paper, setup the entire organization, and then go on to create a poster. You should limit text as much as possible: do not have entire paragraphs on the poster. Similarly do not put blocks of equations without making sure they are essential to your work and clearly motivating them. Furthermore, make sure that the smallest font on the poster can be seen from two feet of distance. This also applies to graphs, please do not have small illegible labels on your graphs. Do not overcrowd the poster.

   We recommend a size of {40 inches x 30 inches} or {30 inches x 40 inches}.

4. **Poster Submission:** Your midway poster PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

5. **Poster Presentation:** The poster sessions will be hosted in Gather.Town. One member of your team should share screen so the poster is visible to everyone who visits. You should zoom in on the relevant portion that you are speaking about so that listeners will be able to see. **You will get 5 minutes to present your poster plus a couple of minutes for questions.** You are required to attend the midway poster sessions in its entirety—the purpose of this is for you to learn about other groups' course projects. Be sure to practice your presentation several times with your group to ensure that you can fit what you want to say in just 5 minutes.

2.4 Midway Executive Summary

Below are the guidelines for the midway executive summary milestone.

1. **Overview:** The midway executive summary offers each group a chance to present their progress halfway through the project's duration. The midway executive summary should be 3 pages (excluding your references/bibliography).
2. **Contents:** A suggested organization for your midway executive summary is below. However, you are welcome to deviate from this organization, so long as you include all the relevant content.
    - *Title box:* Project title and list of group members' names and Andrew emails.
    - *Section 1: Introduction* Concise overview of the entire executive summary. The introduction should in length be similar to that of a long abstract. It should highlight your motivation, proposed method, and expected results.
    - *Section 2: Dataset and Task.* Detailed description of the task, dataset, and metric(s) for evaluation. This section should be nearly in its final form.
    - *Section 3: Approach.* Description of both (a) your baseline approach and (b) the main method. This section should be nearly in its final form.
    - *Section 4: Related Work.* A short literature survey of 5 or more relevant papers. This section should be nearly in its final form.
    - *Section 5: Experiments.* Precise description of the experiments you will run, and any results if applicable. You are required to present results of a baseline model on your task of interest. You must include skeleton tables/plots. You should also include prose with references to the skeleton tables/plots describing what they will contain.
    - *Section 6: Plan.* Timeline of remaining milestones with dates and who is responsible for each milestone.
3. **Executive Summary Design:** The executive summary should stand-alone. That is, it should not be considered as an appendix to the poster.

    The key point of this milestone is to communicate a clear motivation and goals to the reader. If you received constructive feedback on your proposal direction from the course staff, this is your chance to course correct and present any changes to your original proposal. As with any research endeavor, we expect that a carefully planned project is much more likely to succeed. If this is the first time you find yourself following instructions to "Write the Paper First", please read Jason Eisner's advice page on the topic here.
4. **Executive Summary Submission:** Your midway executive summary PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

2.5 Final Poster

Below are the guidelines for the final poster milestones:

1. **Overview:** The final poster should be the natural extension of your midway poster. It should include all the components detailing your work as well as your now completed results. Students are required to attend the final poster session.
2. Instructors will stop by your poster during the session. **You will get 5 minutes to present your poster plus a couple of minutes for questions.** We will be looking for the following:
   (a) Motivation: Is the goal of the project clear? Are the task, dataset, and metrics clearly described? Is there a clear understanding of related work?
   (b) Details of the proposed methods: Which methods are applied? Are they clearly explained? Is the choice of methods technically sound? How are they adapted to this problem? Is the main approach novel?
   (c) Results: Are the current results reasonable? Are the results explained well? Were the errors analyzed in order to understand why each method behaves as it does?
   (d) Presentation: Were poster/explanations concise? Were poster/explanations clear? Did Q&A demonstrate understanding?
3. **Poster Submission:** Your final poster PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.
4. **Poster Presentation:** We will arrange the timing of the poster session so that you have a chance to present your poster to your classmates as well as view their posters.

2.6 Final Executive Summary

Below are the guidelines for the final executive summary milestone.

1. **Overview:** The final executive summary gives each group the opportunity to fully describe their course project work. The midway executive summary should be 4 pages (excluding your references/bibliography).
2. **Contents:** A suggested organization for your final executive summary is below. However, you are welcome to deviate from this organization, so long as you include all the relevant content.
    - *Title box:* Project title and list of group members' names and Andrew emails.
    - *Section 1: Introduction* Concise overview of the entire executive summary including motivation, proposed method, and results.
    - *Section 2: Dataset and Task.* Detailed description of the task, dataset, and metric(s) for evaluation.
    - *Section 3: Methods.* Description of both (a) your baseline approach and (b) the main method.
    - *Section 4: Related Work.* A short literature survey of 5 or more relevant papers.
    - *Section 5: Experiments.* Full description of your experimental design, results, and analyses.
    - *Section 6: Conclusion.* Summary of your main findings and considerations for future work.
3. **Executive Summary Design:** The executive summary should stand-alone. That is, it should not be considered as an appendix to the final poster. Rather, it should offer the prose narrative that your visual poster lacks. You should not hesitate from including figures and visuals in the executive summary if they enhance the reader's experience. We expect the best of these executive summaries to be ready for submission as short (i.e. 4-page) conference papers to top-tier machine learning conferences.
4. **Executive Summary Submission:** Your midway executive summary PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

3 Dataset and Task Ideas

Natural Language Processing

- Morphological Reinflection: CoNLL-SIGMORPHON 2018 Shared Task
- Part-of-Speech Tagging: Penn Treebank, Twitter POS Dataset, Universal Dependencies
- Constituency Parsing: Penn Treebank, Universal Dependencies
- Dependency Parsing: Universal Dependencies
- Named Entity Recognition: CoNLL-2003 Shared Task, WNUT 2016 Shared Task on Twitter NER, WNUT 2017 Shared Task on Novel/ Emerging NER, CALCS 2018 Shared Task on Code-switched NER
- Coreference Resolution: CoNLL 2012 Shared Task
- Semantic Role Labeling: CoNLL 2005 Shared Task
- Semantic Parsing: AMR (Abstract Meaning Representation) Parsing, SQL Query Parsing
- Event Extraction: ACE 2005 Dataset, MUC-3/4 Datasets
- Joint Event-Entity Extraction: ACE 2005 Dataset
- Event Schema Induction: MUC-3/4 Datasets, InScript Dataset
- Joint Intent Classification and Slot Filling: ATIS Dataset
- Dialog State Tracking: DSTC-2 Dataset, MultiWOZ 2.0 Dataset
- Dialog Response Generation: MultiWOZ 2.0 Dataset for task-oriented dialog, PersonaChat Dataset for chit-chat dialog
- Data-to-Text Generation: WebNLG Challenge 2017, E2E NLG Challenge
- Abstractive Summarization: CNN/ Dailymail Dataset, NEWSROOM Dataset
- Long-form Question Answering: ELI5 Dataset
- Procedural Text Generation: Now You're Cooking Dataset for recipe generation
- Story Generation: WritingPrompts Dataset
- Machine Translation: English-French Dataset of Canadian Parliamentary Proceedings, EuroParl Dataset, WikiMatrix with 1620 language pairs

Computer Vision

- Image Generation: CelebA, LSUN-Bedroom Places2 wikiArt
- Mesh Generation: ShapeNet, ModelNet
- Point Cloud Generation: ShapeNet, ModelNet
- 2D-to-3D Object Generation: ShapeNet
- Image Instance Segmentation: COCO, Cityscapes
- Stereo Depth Estimation: KITTI, NYU Depth-v2

- Pose Estimation: MPII Human Pose
- Image Super Resolution: ImageNet, COCO
- Image Captioning: COCO
- Video Captioning: Dense-Captioning Events in Video
- Visual Question Answering: VQA-v2
- Medical Image Analysis: fastMRI ultrasound nerve segmentation CheXpert

Signal Processing

- Phoneme recognition: Timit speech corpus
- Piano music transcription: MAPs dataset
- Stock market prediction: Historical daily prices and volumes of all U.S. stocks and ETFs (Kaggle)
- Speech recognition: LibriSpeech
- Speech recognition: Spoken Wikipedia Corpora
- Speech recognition: CommonVoice dataset
- Medical intervention prediction from waveform data: MIMIC-III Waveform Database

Databases

- Knowledge base completion (e.g. populating Wikipedia infoboxes): Wikipedia article text and infoboxes
- Knowledge base completion (e.g. predicting relations between entities): Dbpedia

Healthcare

- Electronic health records: MIMIC-III and MIMIC-IV
- Medical time series: eICU
- Chest X-rays: CheXpert and MIMIC-CXR

Computational Biology

- Gene identification: Drosophila Melanogaster Genome
- Microarray cancer datasets for unsupervised clustering: prostate cancer dataset (Dhanasekaran et al., 2001), lung cancer dataset (Garber et al., 2001), lymphoma cancer dataset (Alizadeh et al., 2000)

4 Example Projects

4.1 Projects from Past Offerings of 10-708

- 10-708 in Spring 2020 Project Suggestions
- 10-708 in Spring 2019 Completed Projects
- 10-708 in Spring 2019 Project Suggestions
- 10-708 in Spring 2017 Completed Projects
- 10-708 in Spring 2016 Project Suggestions

4.2 Point Cloud Generation (Computer Vision)

Overview: Studying the effect of independence assumptions on point cloud generation

Main Motivation: Good generative models for point cloud data would allow for the easy creation of larger datasets for machine learning on 3D data as a whole as well as opening the door for the creative generation of objects for video games, 3D printing, or sculpture. Current models for point cloud generation either treat each point as separately generated from a hierarchical distribution or conditionally generated on its nearest neighbors. Only recently have graphical model approaches been embraced in this field. The goal of this project is to investigate how different generative models (and therefore independence assumptions) affect the generation of realistic point clouds.

Task: Point cloud generation

Dataset: ShapeNet

Baseline Models: Point Cloud GAN, PointFlow

Assumptions to Compare: Mean-field approximation, points independent given class, points locally dependent, mixture of distributions conditioned on part of object, etc.

Evaluation Metrics: Chamfer Distance, Distance to Face (D2F), Coverage

References: https://arxiv.org/abs/1810.05795
https://arxiv.org/abs/1906.12320
https://arxiv.org/abs/1712.07262

4.3 Speech Recognition (Signal Processing)

Overview: Pushing ASR performance by using SCRFs for feature combination.

Main Motivation: Segmental Conditional Random Fields enable the integration of new sources of information on top of state-of-the-art baselines by using features based on acousticsl, such as template matching, duration models, phoneme detections, Poisson process models and deep neural features like phoneme detectors. The flexibility of the approach opens the possibility of building new systems based on information integration across numerous sources.

Task: Speech recognition

Dataset: Wall Street Journal(WSJ), Broadcast News

Baseline Models: HMM with postprocessing(WSJ), IBM Attila decoder (Broadcast) (Chen et al., 2006)

Feature combinations to compare : Gains achieved by using heuristic features based on template matching and word detectors, and gains by adding neural features.

References: https://www.nyu.edu/projects/bowman/icassp2011.pdf

4.4 Topic Modeling:

Goal Study the application of modern advances in topic modeling to expression microarray data, a key clustering problem that arises in bioinformatics.

Main Motivation Recent work in topic modeling has introduced a variety of new models that treat words not as discrete, but as continuous vectors (aka. word embeddings) (Das et al., 2015; Batmanghelich et al., 2016; Xun et al., 2017; Moody, 2016; Jin et al., 2018; Li et al., 2016; Xu et al., 2018; Bunk and Krestel, 2018). Simultaneously, there have been a variety of applications of more traditional topic modeling approaches (i.e. those that treat words as discrete) to emerging problems in bioinformatics where the data is not text, but biological data (see Liu et al. (2016) for a survey of such applications). The bioinformatics community introduced latent process decomposition (LPD) which is quite similar to LDA, but with the word multinomial distribution replaced by a Gaussian (Rogers et al., 2005; Masada et al., 2009). Here we propose to explore modern advances that support continuous valued word vectors and their application to the microarray clustering tasks explored in Rogers et al. (2005) and Masada et al. (2009).

Task Unsupervised clustering of microarray dataset for cancer.

Dataset Prostate cancer dataset (Dhanasekaran et al., 2001), lung cancer dataset (Garber et al., 2001), or lymphoma cancer dataset (Alizadeh et al., 2000)

Baseline Model None (since the methods below are so distant)

Decoding Techniques to compare : The distilled Wasserstein method of Xu et al. (2018) and latent process decomposition (LPD)

Evaluation Metrics Marginal likelihood; Clustering metrics using the ground truth labels

4.5 Uncertainty Estimation (Computer Vision)

Goal Estimating the underlying uncertainty in images and deep learning models

Main Motivation Deep-learning-based algorithms have achieved great success in computer vision, but the deployment of such algorithms in real world has been hindered by the lack of reliable quantification of uncertainty in the images and models. Estimating such uncertainty can help humans understand what the models have learned and when the outputs might not be reliable, allowing a better and safer use of learning-based methods in practice.

References: https://arxiv.org/abs/1703.04977
https://openaccess.thecvf.com/content_CVPR_2020/html/Carvalho_Scalable_Uncertainty_for_Computer_Vision_With_Functional_Variational_Inference_CVPR_2020_paper.html

4.6 Abstractive Summarization (NLP)

Goal Studying the effect of different decoding techniques on abstractive summarization

Main Motivation Despite significant advances, abstractive summarization models still suffer from the problem of producing repetitive phrases. Many studies attribute this degeneration problem to the use of likelihood as the decoding objective, which compromises diversity in its aim to optimize for output with high probability. A recent study on text generation abilities of neural language models by Holtzman et al. (2019) has shown that sampling-based decoding techniques can be used to mitigate this problem to some extent. Moreover the choice of sampling technique significantly affects diversity of generated text. They also propose a new decoding technique called nucleus sampling and show that it outperforms all other decoding methods such as beam search, sampling with temperature and top-k sampling. The aim of this project is to compare and contrast the performance of several decoding techniques on abstractive summarization and evaluate whether sampling-based techniques do better on this less open-ended task.

Task Abstractive Summarization

Dataset CNN/ Dailymail

Baseline Model Pointer-generator networks (See et al., 2017)

Decoding Techniques to compare Beam search, Top-k sampling, Sampling with temperature, Nucleus sampling

Evaluation Metrics Automatic evaluation (ROUGE-1,2,L), Potential human evaluation for diversity

References

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., and Yu, X. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503.

Batmanghelich, K., Saeedi, A., Narasimhan, K., and Gershman, S. (2016). Nonparametric Spherical Topic Modeling with Word Embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 537–542, Berlin, Germany. Association for Computational Linguistics.

Bunk, S. and Krestel, R. (2018). WELDA: Enhancing Topic Models by Incorporating Local Word Context. pages 293–302.

Chen, S. F., Kingsbury, B., Mangu, L., Povey, D., Saon, G., Soltau, H., and Zweig, G. (2006). Advances in speech transcription at ibm under the darpa ears program. *Trans. Audio, Speech and Lang. Proc.*, 14(5):1596–1608.

Das, R., Zaheer, M., and Dyer, C. (2015). Gaussian LDA for Topic Models with Word Embeddings. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 795–804, Beijing, China. Association for Computational Linguistics.

Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature*, 412(6849):822.

Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., Van De Rijn, M., Rosen, G. D., Perou, C. M., and Whyte, R. I. (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences*, 98(24):13784–13789.

Holtzman, A., Buys, J., Forbes, M., and Choi, Y. (2019). The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Jin, M., Luo, X., Zhu, H., and Zhuo, H. H. (2018). Combining Deep Learning and Topic Modeling for Review Understanding in Context-Aware Recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1605–1614, New Orleans, Louisiana. Association for Computational Linguistics.

Li, S., Chua, T.-S., Zhu, J., and Miao, C. (2016). Generative Topic Embedding: a Continuous Representation of Documents. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 666–675, Berlin, Germany. Association for Computational Linguistics.

Liu, L., Tang, L., Dong, W., Yao, S., and Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, 5(1):1608.

Masada, T., Hamada, T., Shibata, Y., and Oguri, K. (2009). Bayesian multi-topic microarray analysis with hyperparameter reestimation. In *International Conference on Advanced Data Mining and Applications*, pages 253–264. Springer.

Moody, C. E. (2016). Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec. *arXiv:1605.02019 [cs]*. arXiv: 1605.02019.

Rogers, S., Girolami, M., Campbell, C., and Breitling, R. (2005). The latent process decomposition of cDNA microarray data sets. *IEEE/ACM transactions on computational biology and bioinformatics*, 2(2):143–156.

See, A., Liu, P. J., and Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Xu, H., Wang, W., Liu, W., and Carin, L. (2018). Distilled Wasserstein Learning for Word Embedding and Topic Modeling. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 1716–1725. Curran Associates, Inc.

Xun, G., Li, Y., Zhao, W. X., Gao, J., and Zhang, A. (2017). A Correlated Topic Model Using Word Embeddings. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4207–4213.