Initial Plan

Final Year Project - CM3203 - 40 Credits

Generating Differentially Private Datasets Using Deep Learning (in collaboration with: Office for National Statistics)

Author:
Sam Wincott

Supervisor:
Dr. George Theodorakopoulos

February 3, 2020

Contents

1 Project Description **2**
1.1 Abstract 2
1.2 Description 2

2 Project Aims and Objectives **3**
2.1 Aim 3
2.2 Objectives 3

3 Ethical Approval **3**

4 Work Plan **3**
4.1 Milestones 4
4.2 Weekly Plan 4

1 Project Description

1.1 Abstract

Government organisations, businesses, academia, members of the public and other decision-making bodies require access to a wide variety of administrative and survey data to make informed and accurate decisions. However, the collecting bodies are often unable to share sensitive data without risking breaking the confidentiality and consent checks required to obtain this data.

Therefore, researchers have proposed many methods for generating synthetic data to replace the raw data for the purposes of processing and analysis. A good synthetic dataset has two properties: it is representative of the original data and it provides strong guarantees about privacy.

This project is in collaboration with the Office for National Statistics, and it involves the application of the concept of differential privacy for the generation of synthetic data using deep learning techniques. It is a condition by the ONS that the code from the project be open and available on GitHub.

1.2 Description

Previous privacy-aware data sharing methods primarily focus on obscuring or redacting a dataset, whereas this project sets out to develop a method to produce an entirely new dataset with the same statistical characteristics as the original. The accuracy of the statistical characteristics is key to the utility of the generated dataset. This project sets out to create datasets that are useful in traditional statistical analysis, and can be used to train deep learning models.

Differential privacy is based on the notion of adding noise to query results in order to protect privacy (Dwork & Roth 2014). Many techniques have been proposed to protect privacy of a dataset while still benefiting from the statistical characteristics, methods include k-anonymity (Sweeney 2002), t-closeness (Li et al. 2007), and l-diversity (Machanavajjhala et al. 2007). Differential privacy has been accepted as a rigorous standard and is widely used.

To generate the synthetic data, the initial plan for this project is to use Generative Adversarial Networks (GANs) introduced in Goodfellow et al. (2014). GANs have been shown to be a powerful method of generating synthetic data from real data, however they do not provide any privacy guarantees.

2 Project Aims and Objectives

2.1 Aim

Effective statistical representation and privacy preservation are the two topmost priorities of the data generation in this project. The aim is to produce a system that can, given a dataset, generate a differentially private synthetic version.

2.2 Objectives

Here are the objectives:

- Research generative deep learning models
  - The initial plan is to research and use GANs. However, time permitting, I will explore other generative methods such as Variational Autoencoders (Kingma & Welling 2019)
- Create deep learning model to produce synthetic datasets
- Analyse how representative the synthetic dataset is of the original
  - This will be part of an iterative process of updating the code and tuning parameters to improve accuracy
- Analyse how well privacy has been preserved in the new dataset
  - This will also be part of an iterative process, tweaking the code to achieve the desired privacy

3 Ethical Approval

Having discussed this with my supervisor, for this project I will use a public dataset. Because I am using a public dataset I will not need ethical approval, as it is stated in the School of Computer Science and Informatics Ethics Procedures and Guidelines that publicly and lawfully available information is not subject to the ethical review process.

4 Work Plan

Outlined below is a list of milestones detailing points of significant progress in the project, and a weekly plan to expand upon these milestones.

4.1 Milestones

1. Submit initial plan
2. Generate first synthetic dataset
3. Generate first privacy preserving synthetic dataset
4. Finalise the code
5. Finish final report
6. Submit final report and code

4.2 Weekly Plan

Week 1 (27/01/2020)
- Write initial plan

Week 2 (03/02/2020)
- **Milestone** - Submit initial plan
- Review current work in the field
- Read and expand knowledge on GANs and differential privacy

Week 3 (10/02/2020)
- Obtain dataset to use for project
- Experiment with various machine learning frameworks in Python (Tensorflow, PyTorch, Keras)
- Start implementing code for this project

Week 4 (17/02/2020)
- **Milestone** - Generate first synthetic dataset

Week 5 (24/02/2020)
- **First review meeting**
- Analyse how closely the synthetic dataset represents the original
- Tweak code and tune parameters based on the analysis

Week 6 (02/03/2020)
- Begin incorporating differential privacy techniques to data generation

Week 7 (09/03/2020)
- **Milestone** - Generate first privacy preserving synthetic dataset

Weeks 8, 9, and 10 (16/03/2020) - (05/04/2020)
- Analyse the privacy preservation of the synthetic dataset
- Tweak code and tune parameters to increase accuracy of representation and privacy preservation

Weeks 11 and 12 (06/04/2020) - (19/04/2020)
- **Milestone** - Finalise the code
- First draft of report

Week 13 (20/04/2020)
- **Second review meeting**
- Expand and improve upon initial draft of report

Week 14 (27/04/2020)
- **Milestone** - Finish the report
- Check the report, fix minor errors and missing details

Week 15 (03/05/2020)
- **Milestone** - Submit project report and code
- Publish code to GitHub

References

Dwork, C. & Roth, A. (2014), 'The algorithmic foundations of differential privacy', *Foundations and Trends® in Theoretical Computer Science* **9**(3–4), 211–407.
URL: *http://dx.doi.org/10.1561/0400000042*

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014), Generative adversarial nets, *in* Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence & K. Q. Weinberger, eds, 'Advances in Neural Information Processing Systems 27', Curran Associates, Inc., pp. 2672–2680.
URL: *http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf*

Kingma, D. P. & Welling, M. (2019), 'An introduction to variational autoencoders', *Foundations and Trends® in Machine Learning* **12**(4), 307–392.
URL: *http://dx.doi.org/10.1561/2200000056*

Li, N., Li, T. & Venkatasubramanian, S. (2007), 't-closeness: Privacy beyond k-anonymity and l-diversity', *2007 IEEE 23rd International Conference on Data Engineering* .

Machanavajjhala, A., Kifer, D., Gehrke, J. & Venkitasubramaniam, M. (2007), 'L-diversity: Privacy beyond k-anonymity', *ACM Trans. Knowl. Discov. Data* **1**(1), 3–es.
URL: *https://doi.org/10.1145/1217299.1217302*

Sweeney, L. (2002), 'k-anonymity: A model for protecting privacy', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10**(05), 557–570.