Programme Area: Smart Systems and Heat

Project: WP1 Appliance Disaggregation

Title: Online learning and distributed learning

Abstract:

This deliverable describes the fifth task which is related to scalable inference. Specifically; it shows that the proposed inference algorithms of the previous deliverables work online. A full description of the scalable inference algorithms can be found in the previous deliverables and enclosed papers.

Context:

The High Frequency Appliance Disaggregation Analysis (HFADA) project builds upon work undertaken in the Smart Systems and Heat (SSH) programme delivered by the Energy Systems Catapult for the ETI, to refine intelligence and gain detailed smart home energy data. The project analysed in depth data from five homes that trialed the SSH programme's Home Energy Management System (HEMS) to identify which appliances are present within a building and when they are in operation. The main goal of the HFADA project was to detect human behaviour patterns in order to forecast the home energy needs of people in the future. In particular the project delivered a detailed set of data mining algorithms to help identify patterns of building occupancy and energy use within domestic homes from water, gas and electricity data.

Project: HFADA

HIGH FREQUENCY APPLIANCE DISAGGREGATION ANALYSIS

Online learning and distributed learning

Contents

1. History

Date	Issue	Details of Change
	Version 0.0	Initial Version. Authors: Saad Mohamad Hamid Bouchachia

2. Documents Referenced

Ref	Document	Title
1	Word document that describes the HEMS data.	Data collection and data format – ELECTRIC, WATER and HEMS-V1 MONITORING
2	Word document that describes the HEMS V1 Mongo data base structure.	HEMS V1 Mongo Data Base Structure
3	Deliverable 1	HFADA_Deliverable_Ver2
4	Deliverable 2	HFADA_Deliverable2_V0
	Deliverable 3	HFADA_Deliverable3_V0
5	Paper	Deep Online Hierarchical Unsupervised Learning for Pattern Mining from Utility Usage Data
6	Paper	Deep Online Hierarchical Dynamic Unsupervised Learning for Pattern Mining from Utility Usage Data

3. Glossary of Terms

Ref	Description
ETI	Energy Technologies Institute
HEMS	Home Energy Management System (also referred to as HEMS V1)
HFAD	High Frequency Appliance Detection
LDA	Latent Dirichlet Allocation
GLDA	Gaussian Latent Dirichlet Allocation
DBN	Deep Belief network
HMM	Hidden Markov Model
DBN-LDA-HMM	Our proposed Deep-Hierarchical-Dynamic model
NILM	Non-intrusive load monitoring

4. Executive Summary

This deliverable describes the fifth task which is related to scalable inference. Specifically; it shows that the proposed inference algorithms of the previous deliverables already work online which satisfies the requirements of deliverable 5. A full description of the scalable inference algorithms can be found in the previous deliverables and enclosed papers.

5. Introduction

As mentioned in the previous reports, the size of the available data is very huge (around 80Tb). Algorithms processing such large data must observe time and memory restrictions [1, 4]. In this project, we proposed complex hierarchical models to extract knowledge from the complex and big data which makes the time and memory restrictions even more challenging.

Probabilistic models with latent variables have grown into a backbone in many modern machine learning applications such as text analysis, computer vision, time series analysis, network modelling, and others. The main challenge in such models is to compute the posterior distribution over some hidden variables encoding hidden structure in the observed data. Generally, computing the posterior is intractable and approximation is required. Markov chain Monte Carlo (MCMC) sampling has been the dominant paradigm for posterior computation. It constructs a Markov chain on the hidden variables whose stationary distribution is the desired posterior. Hence, the approximation is based on sampling for a long time to (hopefully) collect samples from the posterior [11].

Recently, variational inference (VI) has become widely used as a deterministic alternative approach to MCMC sampling. In general, VI tends to be faster than MCMC which makes it more suitable for problems with large data sets. VI turns the inference problem to an optimization problem by positing a simpler family of distributions and finding the member of the family that is closest to the true posterior distribution [8]. Hence, the inference task boils down to an optimization problem of a non-convex objective function. This allows us to bring sophisticated tools from optimization literature to tackle the performance problems. Recently, stochastic optimisation has been applied to VI in order to cope with massive data [2]. While VI requires repeatedly iterating over the whole data set before updating the variational parameters (parameters of the variational objective), stochastic variational inference (SVI) updates the parameters every time a single data example is processed. Therefore, by the end of one pass through the dataset, the parameters will have been updated multiple times. Hence, the model parameters converge faster, while using less computational resources. The idea of SVI is to move the variational parameters at each iteration in the direction of a noisy estimate of the variational objective's natural gradient based on a couple of examples [2]. Following these stochastic gradients with certain conditions on the (decreasing) learning rate schedule, SVI provably converges to a local optimum [12].

SVI is inherently serial and requires the model parameters to fit in the memory of a single processor. Authors in [13] present an inference algorithm, where the data is divided across several workers and each of them performs VI updates in parallel. However, at each iteration, the workers are synchronised to combine their obtained parameters. Such synchronisation limits the scalability and decreases the speed of the update to that of the slowest worker. To avoid bulk synchronisation, authors in [10] adopt asynchronous distributed update based on few (mini-batched or single) data points acquired from distributed sources. The update steps are then aggregated to form the global update.

In this project, we developed online inference versions of the proposed complex hierarchical models. The main concept is inspired form the development of SVI from VI. In deliverable 2, GLDA is derived from the family of model presented in [2]. This is done by specifying the distributions from the exponential family, then developing the corresponding update equations. More details can be found in the next section. In deliverable 3, the proposed model can also be cast as a member of the family of models discussed in [2]. However, some structure in the model can be exploited to allow better inference than when applying SVI. Hence, inspired form SVI [2] and [9], we develop a novel online inference algorithm. More details can be found in the next section.

6. Online Models and Inference

In the following, we present an overview of the models in each deliverable and their proposed online inference algorithms.

6.1 Online Inference for GLDA

As explained in Deliverable 2, Gaussian Latent Dirichlet Allocation (GLDA) is based on hierarchical Bayesian mixture model. More precisely, this model is a member of the family of graphical models proposed by [2] where observations (input data), global hidden variables, local hidden variables, and some fixed parameters are brought together to define the structure of the model. Under some assumptions, in addition to those indicated in [2], we end up with a Gaussian version of Latent Dirichlet Allocation (GDLA) where the observations (input data) are continuous. In particular, we assume that the hidden local variables are conditionally independent. Hence, the observations can be treated as a bag of words. This approach has drawn inspiration from the success that LDA has achieved in the domain of text modelling. GLDA only differs from LDA in the distribution over the input data, whereas in GLDA, it is Gaussian distribution while LDA uses Multinomial distribution. Since the distribution over the global components of GLDA must be the conjugate distribution of the one over the input, we end up with Normal-inverse-Wishart rather than the Dirichlet distribution as in LDA. Changing these two distributions lead to new update equations for the inference algorithm of GLDA. More details can be found in the enclosed paper [5].

The purpose of the inference algorithm is to approximate an intractable posterior representing the energy consumption patterns. Variational inference (VI) approximates this posterior by positing a family of simple distributions (proxy) and find the member of the family that is closest to the posterior (closeness is measured with KL divergence). The resulting optimization problem is equivalent to maximizing the evidence lower bound (ELBO) [5]. To find the targeted approximate distribution, VI optimizes ELBO with respect to variational parameters defining the simple distribution approximating the real intractable posterior. However, VI requires going through the whole data at each optimisation iteration. Therefore, rather than analysing all the data to compute the update of each iteration, stochastic optimization can be used. Assuming that the data samples are uniformly randomly selected from the dataset, an unbiased noisy natural gradient can be computed based on a single data point. More details can be found in the enclosed paper [5].

6.2 Online Inference for LDA-HMM

In deliverable 3, The proposed approach is a three-module architecture composed of Deep Belief Network (DBN), a hierarchical Bayesian mixture model based on Latent Dirichlet Allocation (LDA) and a Dynamic Bayesian Network model based on Bayesian Hidden Markov Model (HMM). Hence, we call it DBN-LDA-HMM. DBN is pre-trained and used to extract informative features from the raw input. LDA-HMM is then applied on the new feature space. This model was motivated by the success and the efficiency of the bag-of-features approach, adopted by topic modelling, in solving general high-level problems. The temporal ordering power of HMM is used to correlate the activity at high-level. Thus, DBN constructs appliance-specific features which are used by the hierarchical Bayesian mixture model to construct components (topic)-specific features. Mixtures of these components form the residents' energy consumption patterns. The dynamic modelling part exploits the temporal regularity in the human behaviour leading to better performance and allowing forecasting energy demand.

We propose an online inference algorithm, inspired by [2], which exploits the HMM-part structure allowing scalable inference to cope with the massive amount of energy consumption data. The developed update equations involve terms closely related to LDA and other related to HMM. It can be easily shown that LDA-HMM is a member of the family of graphical models proposed in [2] where observations, global hidden variables, local hidden variables, and fixed parameters are involved. The global hidden variables include appliance-related (low-level) variables, patterns-related (high-level) variables and dynamic-related variables. The local hidden variables include HMM ”state” selection variables and LDA ”topic” selection variables. The state variables are distributed according to Multinomial distribution governed by the global dynamic parameters. They select the patterns generating the topic selection variables which are also Multinomial distribution. Note that the observations are the discrete output of the DL algorithm. Hence, SVI for LDAHMM can be derived following similar but more complicated steps as LDA in [2] and HMM in [9]. However, for simplification, we develop tailored SVI to LDA-HMM.

Similar to GLDA, our ultimate purpose is to compute the posterior distribution over the hidden variables or some of them. By doing so, we can get insight into the energy consumption behaviour and lifes[1]tyle of the residents. However, it can be clearly seen that computing such posterior is intractable and approximation is needed. In contrast ot GLDA where the mean-field variational family, which is the commonly used and simplest approximation where each hidden variable is independent and governed by its own parameter, is used. We propose partial mean-field variational distributions by retaining the dynamic structure of the HMM-part of the model because inference for those variables is tractable using the well-known Forward-backward algorithm. More details can be found in the enclosed paper [7].

7. Scaling-up the Models

In order to cope with the real-time aspect efficiently, the proposed algorithms GLDA and DBN-LDA-HMM have implemented on a distributed platform using ps-lite[2] framework. The goal of the parameter server to coordinate distributed machine learning applications. As shown in Figure 1 below, ps-lite generalises the Master-Slave architecture by enabling server nodes, servers nodes and a schedule node.

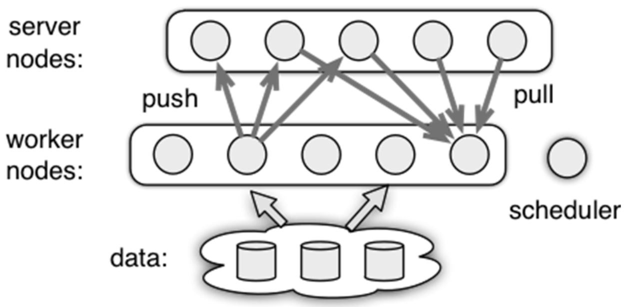

Figure 1: ps-lite framework

Worker nodes are responsible for the computations, the server nodes maintain the overall model. Worker nodes communicate with the server nodes via *push* and *pull*. A worker node pushes its partial results to the servers and pulls the up-to-date model from the servers. On the other hand, the scheduler node monitors the aliveness of other nodes.

The details of the distributed implementation of the proposed algorithms are mainly based on the synchronous variational inference algorithm proposed in [10].

[1] https://github.com/dmlc/ps-lite/blob/master/docs/overview.md

[2] ps: stands for "parameter server"

8. Conclusion

In this report, we have briefly explained how the proposed can cope with high velocity and big volumes of data by enabling online and distributed learning. Overall, the present work is original and most of it has been submitted for publication.

9. References

[1] https://www.proteus-bigdata.com/project-description/scalable-online-machine-learning/

[2] Hoffman, Matthew D., et al. "Stochastic variational inference." *Journal of Machine Learning Research 14.1 (2013): 1303-1347.*

[3] Mohamad, Saad, Abdelhamid Bouchachia, and Moamar Sayed-Mouchaweh. "A Bi-Criteria Active Learning Algorithm for Dynamic Data Streams." *IEEE Transactions on Neural Networks and Learning Systems .2016.*

[4] Gama, João, et al. "A survey on concept drift adaptation." *ACM Computing Surveys (CSUR) 46.4 (2014): 44.*

[5] Saad Mohamad, Abdelhamid Bouchachia "Online Gaussian LDA for unsupervised pattern mining from utility usage data". Submitted *to ECML-PKDD, 2018*

[6] Saad Mohamad, Damla Arifoglu, Chemseddine Mansouri, and Abdelhamid Bouchachia "Deep online hierarchical unsupervised learning for pattern mining from utility usage data". Submitted to *UKCI, 2018*

[7] Saad Mohamad, Damla Arifoglu, Chemseddine Mansouri, and Abdelhamid Bouchachia "Deep Online Hierarchical Dynamic Unsupervised Learning for Pattern Mining from Utility Usage Data". Submitted to *PAMI, 2018*

[8] Martin J. Wainwright and Michael I. Jordan. "Graphical models, exponential families, and variational inference." *Foundations and Trends® in Machine Learning* 1.1–2 (2008): 1-305.

[9] M. Johnson and A. Willsky, "Stochastic variational inference for bayesian time series models," in *International Conference on Machine Learning*, 2014, pp. 1854–1862

[10] Saad Mohamad, Abdelhamid Bouchachia, and Moamar Sayed-Mouchaweh. "Asynchronous Stochastic Variational Inference." *arXiv preprint arXiv:1801.04289* (2018).

[11] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. 2003. An introduction to MCMC for machine learning. *Machine learning* 50, 1-2 (2003), 5–43.

[12] Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics* (1951), 400–407.

[13] Neiswanger, Willie, Chong Wang, and Eric Xing. "Embarrassingly Parallel Variational Inference in Non-Conjugate Models." *arXiv preprint arXiv:1510.04163*(2015).