INFO-F-530 - Computer Science Seminar

Seminar 1
Data-driven approach to the Air Traffic Flow Management problem

January 23th 2019 presented by Luigi De Giovanni

Université Libre de Bruxelles
Master in Computer Science

Hakim Boulahya

1 Outline

The seminar presents and discuss the Airflow Traffic Flow Management problem: multiple flights have to be travel from various origins to different destinations. The objective is to minimize congestion based on different criteria of the air traffic flow such as the user preferences or sector capacity limitation. This enables avoidance of conflicting trajectories with minimal impact on the preferred ones, and other criteria based on the context of the fleet.

During the seminiar the airspace capacity problem is discussed in detail. Each sector are either assigned a fixed capacity or dynamic one. For example a dynamic sector can either be closed or have more flights during peek hours. Such conditions depend heavily on external variables such as number of flights requesting to cross the sector, the number of controllers assigned to a sector, restrictions of aircraft speed controls and other variables. The goal is to resolve the flights trajectories to avoid congestion by respecting the airspace capcity restrictions. To reach this goal there exists different strategies: delaying plane departures or arrivals, deviates flights trajectories or controlling the speed of the planes. The solution is a set of feasible and optimized strategies to avoid congestion.

The presentation is divide in two major parts. The first part discuss the classical approach to handle such problem. It describes a classic approach available in the litterature using Integer Programming (IP) models. In this part the presenter introduced the different variables to take into consideration, the conditions to respect, the cost of changes and the objective function that is minimizing the delays. The second part discuss the data-driven approach. The goal of this approach is to learn from an IP model based on

trajectory selection. That is, the learning algorithm starts from a set of trajectories and if and new trajectory is required, it will build the new trajectories based on the existing ones available in various data repositories. The talk concludes with some perspectives and the future objectives such as studying the performance of the data-driven approach and comparing it with the classical approaches.

2 Major points

- Multiple constraints, variables and rules have to be take into consideration.It is very complex to take into consideration every constraints, especially when some of them are user defined and intentionnally unprecised.
- Using classical approaches is often not realistically applicable to the real world due to the user requests that makes the IP model structure very complex. The model that uses data repositories to select trajectories is more accurate into taking account of user preferences.

3 Relevant field

The presented approaches make uses of Integer Programming to resolve the problem, which makes it relevant to Mathematical Optimization. It is also relevant to Machine Learning because the data-driven approach make uses of clustering algorithms, a supervised learning approach to data analysis.

4 Similar works

In 2008 Bertsimas et al. provided an integer optimization approach for the Air Traffic Flow Management problem [ATFM2], the same method presented during the seminar. In [ATFM1], a multiagent approach using reinforcement learning, a machine learning technique to solve the reduce congestion by taking into consideration various constraints such as climate conditions, resource allocation and other various constraints, provide another method to resolve the ATFM problem. Another interesting paper is [ATFM3], that presents a stochastic integer programming approach to the problem.

5 Relevant questions

- Integer Programming model was presented during the talk with the objective to minimize the congestion. Is focusing on this minimization objective sufficient to resolve the ATFM problem at scale ?
- The data-driven approach is highly dependent of the data repositories used for trajectories selection. How are the data repository validated ? What is the impact of wrong datasets on the efficiency ?

6 Appreciation and critics

Solving optimization algorithms using data analytics is a subject that interests me a lot. Having an Integer Programming approach to a minimization objective problem provided

me a new approach on how to use data to solve optimization problems. The description of the mathematical problem was well introduced and explain by the speaker as well as the results.

References

[ATFM1] Adrian K. Agogino and Kagan Tumer. A multiagent approach to managing air traffic flow. *Autonomous Agents and Multi-Agent Systems*, 24(1):1–25, January 2012.

[ATFM2] Dimitris Bertsimas, Guglielmo Lulli, and Amedeo Odoni. The Air Traffic Flow Management Problem: An Integer Optimization Approach. In Andrea Lodi, Alessandro Panconesi, and Giovanni Rinaldi, editors, *Integer Programming and Combinatorial Optimization*, volume 5035, pages 34–46. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[ATFM3] Avijit Mukherjee and Mark Hansen. A dynamic rerouting model for air traffic flow management. *Transportation Research Part B: Methodological*, 43(1):159–171, January 2009.

INFO-F-530 - Computer Science Seminar

Seminar 2
Real-Time Data Mining

March 29th 2019 presented by João Gama

Université Libre de Bruxelles
Master in Computer Science

Hakim Boulahya

1 Outline

The talk discuss one of the major problem of our century: the exponential growth of data with a focus of real-time consumption of those. This problem is often known as Big Data and the objective of the talk was the discuss an important part of Big Data: building decision models based on real-time data analysis.

Such a problem a usually approach in a static manner, that is generating a model based on an existing finite data set. One of the emphasize of the talk is that this learning methodology is not adapted to the world problems: static solution can hardly be used overtime for dynamic models.

To discuss the problem of real time decision model implementation, the presenter introduced a use case: a network of sensors to monitor electrical power supply. The methodology proposed to build the model is by using the Only Divisive-Agglomerative Clustering (ODAC), to maintain a continuous cluster structure from real-time data. This structure is build dynamically by using different operations: split and merge. Those operations enable the cluster to be expanded in two ways when changes are required. Respectively for the two operations: more details in the cluster by splitting nodes or merging substructures. Splits or merges of the cluster applied in a probabilistic maneer using Hoeffding inequality, a mathematical bounds based on confidence level and observed data streams.

Finally, the presenter conclude the talk by discussing the available tool for data streams analysis and a overview of a generic model for online adaptive learning algorithms.

2 Major points

- Importance of dynamic models and learning from real-time data streams for real-word application. A major point was that problem can change overtime, and static models have to be rebuild from scratch making the previous ones obselete.
- Data streams: important part of the problem solution is to be able to learn from the data and were they come from.
- Clustering time series: methodology used to solve the problem to build a structured model based on the data.
- Hoeffding bound and the parameter configurations: the operations to build the clustering structures are mainly based on the Hoeffding bound, and all the complexity of the algorithm is based on the comparison performs with the bounds.
- Algorithms presented take into consideration the resource limitations (memory, time), which is important to solve the applications that the speaker discuss.

3 Releveant field

This talk is mainly relevant to Artificial Intelligence field, specifically Machine Learning algorithms. (Big) Data analysis is the main subject of the talk, with an emphasize on real-time data streams.

4 Similar work

The work [MINING1] discuss the classification of automatic classification of faults in transmission lines using data mining, a practical problem of the use case presented by Dr. Gama. In [MINING2] a thorough survey on clustering algorithms, the method presented during the seminar regarding the unsupervided learning to build the model based on the data stream, is highliy relevant to the walk and provide an in-depth classification of the state-of-the-art algorithms available in the litterature.

5 Relevant questions

- Has the clustering time series data stream method to build the dynamic model has been implemented in real-world applications ? How does it compare with the existing solutions (that is statically generated models) in term of efficiency and effectiveness ?
- Does other mathematical structures, such as Neural Networks, have been used to resolve the same problem ? If yes, does it compete well with clustering structures ? What are the benefits of clustering structures ?

6 Appreciation and critics

Big Data is a really interesting subject. The seminar discuss a highly complex problem, that is real-time data analysis. It really helped having a first grasp of the problem that

can be resolved using those methodology and provided a good overview of the problem. I would be very interested into understanding in more details how to solve those problems, and also having real-world example and datasets.

References

[MINING1] Yomara Pires, Jefferson Morais, Claudomir Cardoso, and Aldebaro Klautau. Data Mining Applied to the Electric Power Industry: Classification of Short-Circuit Faults in Transmission Lines. In Nadia Nedjah, Luiza de Macedo Mourelle, and Janusz Kacprzyk, editors, *Innovative Applications in Data Mining*, Studies in Computational Intelligence, pages 107–122. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[MINING2] Jonathan A. Silva, Elaine R. Faria, Rodrigo C. Barros, Eduardo R. Hruschka, André C. P. L. F. de Carvalho, and João Gama. Data stream clustering: A survey. *ACM Computing Surveys*, 46(1):1–31, October 2013.

INFO-F-530 - Computer Science Seminar

Seminar 3
Artificial Intelligence and Data Science in Earth Observation

February 19th 2019 presented by Xiaoxiang Zhu

Université Libre de Bruxelles
Master in Computer Science

Hakim Boulahya

Remark *This keynote has been held during Big Data From Space (BiDS 2019) [BIDS1] in Munich, Germany. I attended this conference as a member of the company I am working for and it has been agreed by Professor Markowitch that one of the invited talks of the conference can be presented as part of this course.*

1 Outline

The talk discuss and describes different Artificial Intelligence methods that make use of Earth Observation (EO) data. The talk is divided in two main subjects: deep learning in remote sensing and geoscientific applications. With deep learning in remote sensing the speaker emphasize on the fact that data classification is only a small part of remote sensing. One major discussion made is the importance of data fusion, due to the disparity of earth observation data generated by different satellites in different format.

The main content of the keynote is the presentation and comparison of different algorithms for image analysis, such as object location detection using Sentinel satellite imagery. Convolutional Neural Networks for time series data analysis to detect changes in image and remote sensing imagery analysis are also compared and discussed.

The last part of the talk is presenting different practical implementation of geoscientific applications such as: zone nuage coverage, car instance segmententation, global classification (building heigths, settlement types, etc) and better undestanding global change process of urbanization. Experimental results of the deep learning techniques used such problems are also presented.

Finally, the talk conclude by explaining the methodologies used to build the dataset. Mainly the importance of labelling sattelite imagery and how the importance of semantic and metadata play an important role in implementing deep learning algorithms for geoscientific applications.

2 Major points

- Importance of labelling of sattelite imagery: petabytes of data, but lack of (significant) metadata which lead to a lack of sufficient training data.
- Proliferation of Earth Observation data encourage the implementation of various deep learning algorithms.
- AI in EO comprise various methods such as data mining, data fusion and deep learning.
- Data fusion between EO and non-EO data provide results that enables a better understanding of current problems and open the door to new researches.

3 Relevant field

The talk is obviously linked to Artificial Intelligence field where the major points discuss the use of Big Data specifically in Earth Observation domain, providing different Machine Learning techniques making use of neural networks to demonstrate results using satellite data imagery.

4 Similar work

Various interesting papers in the same topic of the keynote such as deep learning from Sentinel-1 SAR imagery for classification or benchmark of a Convolutional Neural Network (CNN) for Sentinel-2 images can be found in the proceedings of the conference [BIDS3]. One interesting paper that join the main subject of the talk is the road passability estimation using Deep Neural Networks [BIDS2]. Finally, one might be interested in the fusion of remote sensing data, about which Zhang provides a thorough survey in [BIDS4].

5 Relevant questions

- During the keynote, the issue of unlabelled data was raised. What are the main differences between implementing unsupervised and supervised learning algorithms ?
- How optimal were the results provided by the CNN algorithms for the various geoscientific applications presented ? Is the success rate of the various detection sufficient enough ? How those results compare when data fusion with non-Earth Observation data, such as social media data, is taken into consideration ?

6 Appreciation and critics

Machine Learning algorithms is a trending field and one that I am involved professionally. The talk provided me a good overview of various geoscientific applications implemented using Deep Learning algorithms that makes uses of satellites imagery to analyse and provide solution to the application problems. Nonetheless, the presenter introduced the subject correctly but lacked a bit of a more thorough inspection of remote-sensing techniques and mainly focused on presenting the results for various applications.

References

[BIDS1] Big Data from Space 2019. https://www.bigdatafromspace2019.org, 2019. [Online; accessed 14 June 2019].

[BIDS2] Anastasia Moumtzidou, Marios Bakratsas, Stelios Andreadis, Ilias Gialampoukidis, Stefanos Vrochidis, and Ioannis Kompatsiaris. Road Passability Estimation Using Deep Neural Networks and Satellite Image Patches. February 2019.

[BIDS3] Publications Office of the European Union. Proceedings of 2019 Big Data from Space (BiDS'19) : 19-21 February 2019, Munich (Germany)., February 2019.

[BIDS4] Jixian Zhang. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1(1):5–24, March 2010.