February 2022
JDLA Office

Study Group ‘AI governance and its Evaluation’
Report on the Session #4 (Phase Ⅱ)

1. Introduction

The Japan Deep Learning Association establishes study groups as a forum for deepening knowledge and discussing domestic and international policy trends related to artificial intelligence (hereafter AI) and Deep Learning (hereafter DL). This study group, ‘AI Governance and its Evaluation,’ defines ‘governance’ as a system of management and evaluation by various actors and launched a study group in July 2020 to investigate what forms of governance are possible to help build trustworthy AI systems, and the phaseⅡ began in September 2021.

In the fourth meeting (Nov. 29, 2021), in the first half, Mr. Masaru Sogabe of GRID, Inc. spoke about “The current state of and countermeasures to data drift that effects AI stability”. In the second half, Mr. Yasukazu Hirata of DataRobot, Inc., which provides AI Cloud platforms spoke on “Using DataRobot MLOps for machine learning model governance.” There was also a presentation on “Moving towards a sustainable operation of an AI model” by Mr.Yuta Tatewaki of DataRobot, Inc.

This report is a reconstruction of these topics and a record of the discussion.

1. The current state of and countermeasures to data drift that effects AI stability

In the first half of the session, Mr. Sogabe of GRID, Inc. spoke on "The current state of and countermeasures to data drift that effects AI stability."

Issues affecting the progress of AI into an operational model

In recent years the number of companies utilizing AI and deploying operable indigenous AI has been increasing. In recent years, the main difference among these companies is the number that has advanced from AI’s development stage to the production stage and the corresponding issues which have arisen. Many of the development stage issues were related to a lack of knowledge about technical aspects of AI. However, in the deployment phase, there have been significant difficulties centered around maintaining AI performance and prediction accuracy. One of the causes of AI performance reduction over time is data drift.

Factors that cause data drift in the phase of AI operations

Data drift can be defined as “the tendency and nature of the input data to change from the time of the learning phase," this creates a concern that the prediction element of AI may not be able to keep up with the change. Below are three of the leading causes of data drift.

- Input data error (unexpected data is entered)
- Deterioration of sensor equipment
- Data pollution attacks on AI

Even should an anomaly identification process be added to the AI as a countermeasure against data drift, new data types will always be created, thus, the performance of the AI will eventually deteriorate. Therefore, the following three processes are necessary to mitigate this.

- Check if the appropriate data has been input into the AI
- Monitoring for AI performance degradation
- Following AI degradation, choosing the timing for a relearning phase for the AI

Technological solutions for data drift

Various technology-centric approaches have been tried to combat data drift. One of these approaches is to create an algorithm that detects input and output anomalies as well as data pollution attacks. The algorithm creates a linkage between the AI quality and the data extraction process to detect anomalies in the input data and thus stop the erroneous data from corrupting the AI.

Should such a process not be implemented, the erroneous data will be input into the AI, and a data engineer will be required to correct it. Having such a system in place will prevent poor prediction results, however. As in a real system having this process check every piece of input data is not practical. It would be more realistic for the AI to periodically analyze the data while using an anomaly threshold. Upon reaching the threshold, the AI's data engineers and data scientists would be alerted and the data would be examined by them.

In the case of GRID, Inc., "Proxy A Distance (PAD)", an algorithm that associates feature extraction with AI data quality works on both image recognition modelling and natural language analysis modelling to reduce prediction accuracy errors without the need for training data or annotations.

2. Using DataRobot MLOps as a governance for machine Learning model

In the second half, Mr. Hirata of DataRobot, Inc., which provides AI Cloud platforms spoke on using DataRobot MLOps as a model for Machine learning governance. There was also a presentation on "Moving towards a sustainable AI operations model" by Mr. Tatewaki.

Looking at the necessity of AI governance from domestic trends.

The leaders in AI for each industry also tend to be the major companies in each industry as well. As such, the effect of AI on them is significant. When considering the governance aspect of AI for a company, it is essential to consider its organizational structure. The presently devised three stages of AI maturity are listed below, with three being the most mature.

1. CoE type: A small number of departments consider the usages of AI, and management creates a separate CoE[1] or DX promotion team to take the lead in creating a case study for its use.
2. CoE support type: With the aim of supporting CoE, multiple departments consider the possible usages of AI and promote its use. Depending on a department's progress, a data scientist or DS may be introduced.
3. LoB[2] promotion type: A DS is assigned to every department, and AI is used company-wide. The CoE is integrated into the IT department to support the whole company.

Here, in creating an AI use case under structure 1, selecting a department with low-risk on AI usage is optimal.

Currently, AI is widely used in financial institutions throughout Japan. The scale varies between large and small, but predominantly the AIs are single business single model types. While large companies may use several AI models to perform a task, the single-use AI model remains in the PoC stage, so if a company wishes to become a leader in AI, it needs to use multiple AI models to promote job creation and as well as developing a stable pipeline to manage the job as the whole. Also, the governance idea should always take into account that AI accuracy will eventually deteriorate.

Realizing Stable AI Utilization

The factor that complicates the management of machine learning models is that while a regular system only requires code management, a machine learning model requires that the data and hyperparameters are also managed.

In small-scale single AI operations, all of the AI maintenance is likely to be done by the data scientists in charge of development. In this case, there is a high risk that maintaining the model’s accuracy and continuing its improvement will prove impossible.

In response to this, pipeline development methodology reduces subjectivity in AIs and allows for a continued focus on utilization. For pipeline development, both data scientists and AI engineers work together. Should a multi-model monitoring system become possible

[1] CoE（Center of Excellence）: A team or department to work on specific cross-sectoral issue.

[2] LoB（Line of Business）: Divisions that carries core business function to offer the companies products or services.

in an operating environment, it will also become possible to develop and use impartial AIs. By applying this kind of sustainable AI model development, AI governance can be truly examined for the first time.

Governance Development from a Practical Perspective

In recent years, the concept of MLOps has emerged. MLOps is a contraction of machine learning and operations made by DevOps. DevOps refers to development and operations staff working together, as opposed to MLOps[3], where data scientists and AI engineers maintain a close working relationship with maintenance providers.

Introducing MLOps to AI governance will make it possible to create an accurate business model. When MLOps is introduced into a conventional model's construction and deployment, it provides the infrastructure to allow the AI to detect problems such as data drift and apply a relearning process as necessary. It also allows for the performance of a system to be rated, maintenance performed, and committees set up.

To introduce MLOps, the project department needs to cooperate with the data science department. Of particular note is the importance of having an MLOps engineer in the data science department. MLOps engineers' knowledge of machine learning and IT makes them ideal for communicating with legal and compliance personnel. They can also communicate with project managers, data scientists and data engineers in model deployment approvals. This is an essential role for the pipeline development process.

The introduction of MLOps may change the roles of business managers, data scientists, and data engineers as follows.

- ➢ Business Manager：From creating ideas around individual issues to creating change at a business reform level.
- ➢ Data Scientist (DS)：Timely response to accuracy degradation will be included with their existing model construction and analysis role.
- ➢ AI Engineer：Adding operational metric thresholds to the traditional role of deployment and model-specific monitoring.

The following items should also be considered when implementing MLOps.

- ➢ Model Risk Rules：By aligning company policy and business practices, the amount of risk created by deploying a model can be evaluated through prepared processes and documents.
- ➢ Load Verification：The environment available meets the needs of the business.
- ➢ Data Quality and Compliance：Data quality checks including the consideration of

[3] DevOps and ML: refers to the cooperation of the development team and the operations team for the rapid deployment and quality control of machine learning models that decay over time.

personal data can be completed automatically.

This sort of machine learning governance is developed to reduce risk while considering the balance and cost of the system. It should consider the company’s focus and to what degree machine learning will be implemented. Domestically only a few companies have developed governance for machine learning models. However, it will be necessary for companies to establish governance for multiple models in the future.

Post MLOps Introduction Operations

There are two types of model deployment, one that is built with a dedicated server for prediction processing and one that spreads itself across a variety of terminals. After deploying the model, the following three things should be checked, processing speed, accuracy and normalcy. Of these, accuracy deterioration is the norm, but the following items should be viewed as the cause of this.

- Changes in characteristics： Refers to a change in the object to be predicted. For example, credit card fraud needs to redefine constantly as new kinds of fraud appear every day. This is called concept drift, and it is difficult to detect depending on data drift.
- Unknown input： Incorrect data is entered; this is easily detected.
- Signal disappearance： A change within the characteristics of the data can lead to a change in the features required for a prediction. This can sometimes be detected with data drift.

Concept drift is the unexpected change in the statistical characteristics of a predicted object over time. Concept drift can be divided into two subtypes; sudden concept drift, such as due to a natural disaster like the COVID-19 pandemic, and gradual concept drift, such as slow changes to economic statistics. While it is difficult to predict when sudden concept drift will occur, once it has occurred, it is easily detected.

How unknown data input is handled varies depending on the features of the AI. For numerical features, the output contents change depending on the types of algorithms present. For example, if the algorithm is a linear system, such as a regression or neural network type, it will extrapolate the data linearly. However, if it is a tree type, it will extrapolate the extreme values of the training data. Therefore, a tree type algorithm should be chosen in cases where predicted values cannot exceed a given point due to unknown inputs. In some cases where the prediction exceeds a defined endpoint value, it may be necessary to convert the system to use a fixed value. When the unknow input data is categories (categorial variables), how to deal with these can vary depending on the preprocessing mechanism. DataRobot treats these new categories as an "other category" in either a case of a one-hot or an ordinal encoding process. When the target category becomes strongly related to the prediction target, applying a re-learning process should be considered. If the inputs are from different forms of data expression, that should be taken into account in the preprocessing stage.

Rebuilding the Model

The following need to be considered when deciding whether or not to rebuild the model.

- Has concept drift occurred?
- Can a measured value of the prediction be made immediately?
- Has the accuracy deteriorated?
- Has data drift been observed?

The following are methods for rebuilding the model

- Study training period slides
- Delete all data prior to the concept drift beginning
- Give weight to the latest data
- Use an appropriate model for the conditions

The following are the two methods of rebuilding an old model into a new model. DataRobot can perform both of these automatically.

- Bulk replacement：Completely replacing the old model with a new model, only the environment is prepared for the replacement. The new model needs to be monitored for accuracy.
- Parallel operation replacement： New models are prepared and operated parallel to old models. After a trial period, the most accurate model will be adopted. This is, however, expensive as the data in each model needs to be stored and each model needs to be monitored.

3. Organizer’s summary of the main comments from the participants

In the 4th meeting, the development and management of tools as well as monitoring were discussed. The following questions were discussed based on the speeches provided. Mr. Mikio Ogawa of DataRobot Inc. also offered his insights on the topics.

Discussion on “The current state of and countermeasures to data drift that affects AI stability” by Mr. Sogabe

- Issues related to rebuilding a model.
  - ✓ Until 2018 Model development was being carried out via a trial-and-error process. During this, source code and data management began to be identified as an issue.

Managing these became more prevalent in AI ventures around the year 2020, but it was far from typical.

- Evaluating model accuracy improvements
  - ✓ No matter how a model is improved, the need to learn data distribution remains unchanged. Therefore, whether to improve the model itself or explanatory variables needs to be carefully considered. When considering features, R & D should consider the accuracy level needs of the business.
- Concept drift detection and relearning
  - ✓ Having fields with explanatory labels, such as demand forecast, makes detecting concept drift easier. However, should the field be an image recognition field, the correct answer needs to be defined in order to detect accuracy deterioration as the explanatory value is an XY coordinate. Monitoring the change quantification can quickly identify concept drift and prevent business losses. Incorporating daily monitoring into the business flow and performance monitoring is also preferable to longer monitoring periods. Monitoring business flow is a relatively new concept that has become an important point in the last year or so.
- Maintenance costs for managing multiple unique models
  - ✓ Naturally, the monthly DS cost, as well as maintenance and operational costs, are higher than that of a regular IT system. There are personnel limits to managing a large number of systems as well, but even when engineers are provided with the appropriate tools and work environment, excess staff may not be necessary. Some of the data scientist's workload can be reduced by systems that prevent fraudulent data from being input.

Discussion on "Using DataRobot MLOps as a model for Machine learning governance" by Mr. Hirata and Mr. Tatewaki

- Issues related to rebuilding a model.
  - ✓ In some circumstances, the DS can move jobs after creating the machine learning model, if not managed properly, this can affect the overall management of the system. However, DataRobot provides tools that allow non-engineers to manage the model automatically. DataRobot explains to non-engineers such as management how important model management is. Some departments will naturally have a higher AI literacy than other departments, so some areas may struggle with communication due to a lack of familiarity. In some circumstances, the premise of an AI predicted outcome might be difficult to understand. Hence, it is necessary for people in charge of an area utilizing an AI other than the engineer to have some knowledge of AI and IT.
  - ✓ Data catalogs and code management tools must be used during and after the

learning phase to record inputs such as learning data, learning code, and hyperparameters. In addition to recording inputs, wikis need to be kept in order to record input, output, and why the choice of a particular algorithm or engineering method was made. In addition to training engineers, other staff should be trained to perform audits and evaluations.

- Rating the implementation of model accuracy improvements
  - ✓ Concept drift needs to be decided on based on knowledge and experience because there is no way to detect it. When re-learning is undertaken, measures, such as increasing the weighting on the data post the concept drift, need to be undertaken. Decisions about data drift should be made after examining the log. There are many cases where relearning with data from one or two years ago has been performed.
  - ✓ There are limits to how much accuracy can be improved on via experience. Allowing more time won't guarantee an improvement. In reality, accuracy degradation can occur in one to two weeks in a fast case and one to two months in a slow instance. So, these issues need to be dealt with quickly to avoid degradation. Accuracy improvement obtained outside of a deployment environment can be superficial and may not eventuate to the expected degree. Individual companies need to set their own judgement criteria and timeframes for accuracy improvement.
  - ✓ Besides the simplification of the structure, increasing intrinsic and fundamental data is also crucial. For example, during the COVID-19 pandemic, the sales of stores in the Marunouchi areas decreased. This was not due to location but rather a decrease in foot traffic. To prevent concept drift identifying the critical data as the flow of people rather than location is necessary.
- The impact of concept drift
  - ✓ Depending on the type of concept drift, there can be various effects on a business. The effect is negligible if relearning is possible by simply shifting the data period. However, a large-scale environmental change such as the COVID-19 pandemic likely requires human intervention to correct the data.
- Detecting common changes among different models with data similarities.
  - ✓ While DataRobot has clients nationwide, it remains unclear whether the various models used by clients share any commonalities. As such, it remains necessary to monitor them individually for data drift.
- When using multiple AIs in a single business process, data management departments should carry the following comprehensive controls over as well as MLOps.
  - ✓ For data construction, a "Datamart" style of system needs to be established to manage versions of features. This is sometimes called a feature store. By

grouping and managing features with are unique and those which are common across multiple models redundancy can be prevented and efficiency increased.

- Collaboration between second-tier leaders for data management and project management.
  - ✓ The DS assigned to the CoE needs to have the ability to work in and grasp the situation across all of the departments in the organization, whereas a DS assigned to an individual department does not need this level of skill. However, a DS assigned to a department needs to understand that department's role and expertise.
  - ✓ In the previous generation of systems, data analytics was performed for the core systems and information system infrastructure separately. However, machine learning works differently. Previous systems only analyzed data offline, machine learning can work with customers offline and in real-time. Presently Japanese companies are only providing services with information generated in real-time and cannot link real-time data with historically analyzed data.

    [Chatbot example]
    - Japan： While chat services are possible, services with the analysis of past data that forms the basis of replying to customers are presently unavailable.
    - Amazon： Can respond to questions by using customers' past purchase history.
- Should business risks and social risks be grouped together in MLOps?
  - ✓ Business risk needs to be assessed with an element of social risk included, such as gender and race. This ensures a fair and unbiased risk assessment of the model is performed. As with IT governance, it is essential to have a member from the compliance committee or business division as part of the review committee in order to formulate policy and manage the company's governance needs.
- Maintenance costs for managing multiple unique models.
  - ✓ Japanese data scientists tend to place a degree of importance on creating new models from scratch. However, there is no actual need for this to be the case. As maintenance management tools for models expand, the demand for creating automated maintenance protocols will increase over programming techniques. These new tools are also expected to reduce maintenance cost, which has otherwise been growing.
- Overseas issues with joint public-private partnerships across multiple organizations for managing data drift.
  - ✓ International cases of this are unknown.
  - ✓ Any company needs to keep corporate secrets about which features promote efficacy. For example, AI models for monitoring employee turnover are generally

regarded as ineffective. This is due to each company having individual reasons for staff resignation. These reasons can also vary by country. Unnecessary features in a model can also create additional costs in a specialized system and be nothing more than a distraction.

The discussion of AI Governance domestically and internationally will continue through this study group.

Written by Akihiko Yoshida

Translated by David Shield

＜The 4th Session of the Study Group＞

Date/Time: Monday, November 29th, 15:00-17:00（On Zoom）

Contents：

・Topic 1：”The current state of and countermeasures to data drift that effects AI stability” provided by Mr. Masaru Sogabe (GRID, Inc.)

・Topic 2：”Using DataRobot MLOps as a model for Machine learning governance” provided by Yasukazu Hirada (DataRobot, Inc.)

“Moving towards a sustainable AI operations model” provided by Mr. Yuta Tatewaki (DataRobot, Inc.)

・Questions and Discussion