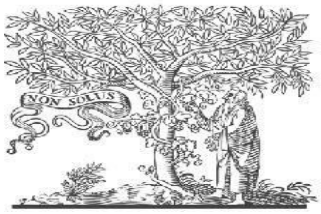
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

IJIEMR Transactions, online available on 1st Jan 2021. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12

DOI: 10.48047/IJIEMR/V09/I12/138

Title: AUTONOMOUS MACHINE LEARNING MODELING USING TASK ONTOLOGY
Volume 09, Issue 12, Pages: 813-817
Paper Authors

S. SINDHUJA, G. THIRUMALA, K. SHRAVANI, T.KANAKAIAH

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

AUTONOMOUS MACHINE LEARNING MODELING USING TASK ONTOLOGY

[1]S. SINDHUJA, [2]G. THIRUMALA, [3]K. SHRAVANI, [4]T.KANAKAIAH

[1,2,3,] B TECH Students, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

[4] Assistant Professor, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

ABSTRACT:

Over the last decades many machine learning experiments have been published, giving benefit to the scientific progress. In order to compare machine-learning experiment results with each other and collaborate positively, they need to be performed thoroughly on the same computing environment, using the same sample datasets and algorithm configurations. Besides this, practical experience shows that scientists and engineers tend to have large output data in their experiments, which is both difficult to analyze and archive properly without provenance metadata. In this paper, we classify typical problem solving steps for autonomous machine learning as tasks, and present a problem solving process. We propose the modelling method of an autonomous machine learning using processes of the task execution on machine learning such as workflow. The proposed task ontology-based machine learning model defines a task-based process grouping scheme of UML activities. And it will automatically generate and extend the machine learning models by transformation rules based on common elements and structures (relationships and processes between elements).

Keywords: UML diagram, dataset, scientific program, provenance.

1. INTRODUCTION

So far, we have seen a variety of publications on Semantic Web and Machine Learning (ML) topics, many of them contributing to the state of the art in their respective fields. However, in the last years we experienced a knowledge gap in the standardization of experiment results for scientific publications. This technological gap can be summed up by the following question: “How to achieve interoperability among machine-learning experiments over different system architectures?”. In particular, experimental results are often not delivered in a common machine-readable way, causing the information extraction and processing to be tricky and burdensome. Moreover, recurring issues regarding the experiment could benefit from the existence of a public vocabulary. Reviewers of publications on Machine Learning often need to investigate basic information on experiments conducted, e.g., which implementation of an algorithm was used, its configuration or choices for related hyper-parameters. On the other hand, autonomous machine learning is still in its infancy, and some techniques provide the ability to reduce the unnecessary tasks that are progressively refined to prepare the model and improve its accuracy. The tools of autonomous machine learning provide an optimal algorithm for machine learning tasks and functions to determine the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

hyper-parameter setting through self-analysis. The typical tools include Autosk learn, Auto-Weka, H2o Driverless AI and Google's Auto ML. In this paper, we describe a typical problem solving process for the machine learning as tasks, present their procedure, and propose the modelling method of autonomous machine learning for using task execution processes. The modelling method of autonomous machine learning based on the task ontology define a structure based grouping method of the UML(Unified Modelling Language activities and implement a function to automatically generate models based on common elements and structures.

2. LITERATURE SURVEY

The related works are discussed in twofold: task ontologies and machine learning ontologies

A Task ontologies

Ontology is defined in various fields depending on the field of applications. In the field of artificial intelligence, it is an explicit and formal specification of how objects and concepts described in the field of interest. In the Semantic Web, an ontology plays a very important role in processing, sharing, and reusing the knowledge for exchanging information between different databases. An ontology is also defined as an explicit description of concepts, attributes, constraints, and relationships between them on the domains. On the other hand, domain ontology can be defined as an 'explicit protocol for conceptualization' of the problem. Task ontology is defined as 'extracting and organizing the concepts and relations existing in the problem solving process domain-independent'. In particular, a task ontology is a specification of the concept structure for the task execution process. Thus, the core concept is the subject of processing and the procedure of processing for a problem solving. In general, a person becomes a subject in task ontology. However, in this paper, agents (programs) become subjects to perform the tasks. Expose ontology is ontology for machine learning experiments. It is used in open ML as a data structuring and data sharing (API) method. Machine Learning (ML) Schema is used to export all openML as linked open data. The DMOP ontology is explicitly designed to support data mining and machine learning. This covers the structure and parameters of predictive models, associated cost functions, and optimization strategies. OntoMD ontology provides a unique framework for data mining research.

B. Machine learning ontologies

ML schema is a top-level ontology that provides classes, properties, and constraints for machine learning algorithms, datasets and experimentation suggested by the W3C(ML Schema community group). It can be easily extended and refined, and can be mapped to other domain ontologies developed in the field of machine learning and data mining. MEX vocabulary has been designed to solve the share of provenance information in a lightweight form. The extended PROV ontology provides a model for representing, capturing, and sharing provision information on the Web. This can enable the use of analytical data and code so that another person can reuse the results. The code and the mark-up language are written in a single file, and processed to create a document. A provenance meta information was proposed as a standard model of data management by W3C.

The provenance information is also "information about entities, activities, and people involved in producing a piece of data which can be used to form assessments about its quality, reliability or trustworthiness". As a standard query language, SPARQL is a query language similar to SQL and stored in Resource Description Framework(RDF) for queries on data.

3. RELATED STUDY

The collection of vocabulary is achieved by extracting words from a paper, a textbook or a machine learning library (API) tutorial, and then selecting keywords from the index and title of the textbook. The frequency of coincidence with the key word is calculated and labelled by a category item. Based on MEX vocabulary, it generates metadata about machine learning execution, algorithm environment, and execution result. The task ontology of the machine learning is modeling with the following items.

1. Project includes user information and task descriptions for the entire experiment.

2. Experimental information describes information about 'what, when, who, what ratings' the task is made'. A task is the procedure of machine learning that includes information such as data preparing, data pre-processing, preparation of learning data, machine learning model setting, a parameter setting of machine learning model, hyper parameter setting and code generation (implementation: python code).

3. Data describe on the characteristics (attributes), the type of data (training data sets, validation data sets, test data sets) and data format type (text, csv, image, json, xml …).

4. The ML framework describes on information such as its name, features, and so on. It describes the machine learning (machine Learning or deep Learning), model name, parameters, etc., and describes on the model evaluation algorithm and so on.

5. The algorithm describes information such as learning type, algorithm name, and parameters.

6. The hyper-parameter describes information about epochs, batch size, learning rate, etc. for the machine learning (deep learning).

7. The implementation manages the systematic information of users, the use date, and the use type about an applied model. It also stores and manages information about the generated models.

8. The generated machine learning code includes the prerequisites for computing information in the run configuration information for execution.

9. The execution environment describes information about the coding language, library, software information, and so on.

10. Performance describes evaluation methods and procedures for the applied model evaluation.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

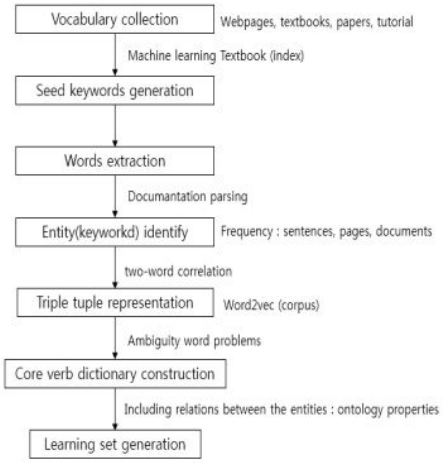


Fig.3.1. A vocabulary acquisition method of the machine learning.

4. PROPOSED SYSTEM

Machine learning knowledge is systematically described in terms, methods (function: executable script), execution environment, and procedures. This means that if a normal output is made through the preceding work, the next work is performed after confirming it. Based on this, it is possible to define a structure-based grouping scheme of activity and automatically generate a model based on common rules (functions) and structures (relations and processes between elements) based on conversion rules. When we design a machine learning workflow, we gradually create a concept graph. The concept graph is to a graph representation of the concept structure. The conceptual structure is expressed as a frame, and the reference object is expanded using the preorder expression of the keyword. This is an expression in which the keyword precedes the relevant machine learning word from the sentence extracted from the document.

Autonomous machine learning modeling is the work for standardization and abstraction to the core of the components base on the meta information of the machine learning. The model consists of the task and process and saves as the method library(API). It defines into small units for modular and systematization of its components. The defined components redefine as a UML-based metamodel for the consistency, traceability, reusability, and implementation-ready between tasks and the results. So the core class of the UML-based meta-model consist of tasks and processes. The Knowledge of the autonomous machine learning also describes a small task unit based on the MEX vocabulary. Fig.5 depicts a part of the knowledge of the object detection using the "YOLO" of the deep learning algorithm. The machine learning pipeline for object detection consists of data import, decision of attribute selection or schema, selection of learning model, construction of learning model, hyper-parameter setting, model training, measurement of model performance, and so on. In this way, the knowledge representation of the project unit is written as '.json' files using the mapping rules based on the machine learning schema and the vocabulary, and convert it into a UML-based meta-model. This model makes that objectives, optimizers, metrics, and layers in the Keras API are meta-model for deep learning.

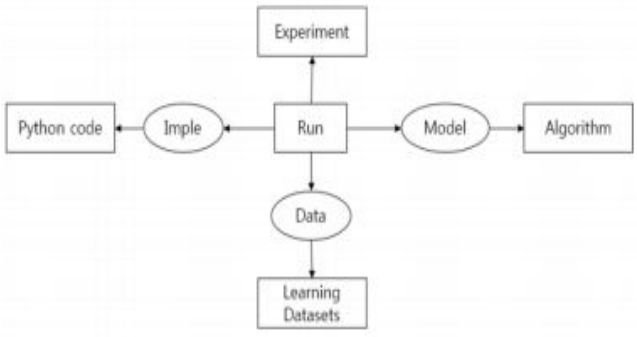


Fig.4.1. A part of conceptual graph.

5. CONCLUSION

In this paper, we extracted important keywords for constructing ontology from papers and textbooks about machine learning. Moreover we designed a task ontology based on the MEX vocabulary. We also studied workflow for the autonomous machine learning model. The proposed method is applicable for automatic workflow according to designated autonomous level. Therefore, the non-experts are capable of doing complex tasks using the proposed method and can easily implement the machine learning model in a specific application.

REFERENCES

[1] C. Perez, The Deep Learning A.I. Playbook : Strategy for Disruptive Artificial Intelligence,I.M.:Intution Machine, 2017

[2] V. Joaquin, S. Larisa, "Expos´e: An Ontology for Data Mining Experiments", Third Generation Data Mining Workshop at ECML PKDD 2010.

[3] I. Mitsuru, S. Kazuhisa, K. Osamu, M. Riichiro, "Task ontology: Ontology for building conceptual problem solving models", In proceeding of ECAI98 Workshop on Applications of Ontologies and Problem-Solving model, pp.126-133, ECA, 1998.

[4] A. F. Martins, R. A. F. De, "Models for Representing Task Ontologies", Proceeding of the 3rd Workshops on Ontologies and their Application, 2008.

[5] S. Kanjana, S. Maleerat, "Ontology Knowledge-Based Framework for Machine Learning Concept", iiWAS '16 Proceedings of the 18th International Conference on Information Integration and Webbased Applications and Services, pp. 50-53, 2016.

[6] P. Gustavo Correa, E. Diego, L. Agnieszka, P. Panče, S. Larisa, S. Tommaso, V. Joaquin, Z. Hamid, "ML Schema: Exposing the Semantics of Machine Learning with Schemas and Ontologies", ICML 2018 - RML Workshop.

[7] G. Eason, B. Noble, and I. N. Sneddon, MEX Interfaces: Automating Machine Learning Metadata Generation, https://www.researchgate.net/publication/305143958, 2016.

[8] E. Diego, N. M. Pablo, M. Diego, C. D. Julio, Z. Amrapali, L. Jens, "MEX Vocabulary: A Lightweight Interchange Format for Machine Learning Experiments", SEMANTiCS 2016 Proceedings of the 12th International Conference on Semantic Systems, pp. 17- 24, 2016.