Augmented Ontology by Handshaking with Machine Learning

Marie Kim, Hyunjoong Kang, Soonhyun Kwon, YongJoon Lee, Kwihoon Kim, Cheol Sik Pyo
KSB Convergence Research Department, ETRI
218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, KOREA Korea
mariekim, kanghj, kwonshzzang, yjl,kwihooi,cspyo @etri.re.kr

***Abstract*— Artificial intelligence products are already around us and will be emerging dramatically a lot in near future. Artificial intelligence is all about data analysis. When it comes to data analysis, there are two representative techniques: machine learning and semantic technology. They stand on the other side from where to begin analysis. Simply speaking, machine learning is based on the data while semantic technology relies on human domain knowledge (human learning). What if collected data are insufficient to reflect whole phenomenon? This is a limitation of machine learning. What if circumstance changes a lot as time goes by? Manual rule updating by experts is not a good solution in that circumstance. Based on these observations, we investigate two approaches and find a good solution which maximizes the advantages of both techniques and mitigates the limitations of them. This paper suggests a novel integration idea to compensate each technology with the other: that is semantic filtering. This paper includes a toy semantic modelling and a machine learning algorithm implementation to realize the proposed concept, semantic filtering.**

***Keywords*—data analysis, semantic filtering, machine learning, semantic technology, Internet of Things**

I. Introduction

The artificial intelligence market is estimated to grow from USD 419.7 Million in 2014 to USD 5.05 Billion by 2020], according to market forecaster. After Google's *Alphago* competition, artificial intelligence (AI) is on the centre of the spotlight from all around world. Some people are excited about upcoming hot smart devices and services, meanwhile, the others are afraid of horrible our future which will be brought by intelligent robots. Nevertheless, at the core of artificial intelligence, there is (actionable) data analysis.

Roughly, we can classify data analysis techniques into two categories: data-driven and rule-based. Traditionally, many smart devices and services use rule-based situation recognition and provide services. Basically, rules in rule-based data analysis are built by human experts based on domain knowledge. This domain knowledge is the result of long-time human experiences (human learning). Therefore, we can define this approach as a human learning approach. The knowledge acquired by human for a long time is transferred into a set of rules, and then these rules are used for situation recognition and inference. Semantic technology [1] is a representative technique for this approach. That is, semantic modeling is done by human experts and then inference can be performed based both on new data and already-built semantic models (a set of rules in semantic language).

Nowadays, IT worlds are more focusing on data-driven approach due to internet of things (IoT), big data, high-performance processors, and high-accuracy algorithms. Deep learning is on the front line of artificial neural network resurgence and it has been achieving pretty good results especially in image recognition and natural language processing [2].

But two technologies have their own strong points and weak points due to their starting point: data-driven and rule-based. Regarding machine learning, there needs assumption on the data that is collected data reflects whole phenomena we are interested in. If this assumption is not satisfied, analysis itself becomes meaningless. Regarding semantic technology, semantic rules should fit the real phenomena consistently. If circumstance changes dynamically and drastically, then rules may be obsolete. It may not be accurate for the upcoming circumstances anymore.

Therefore, we need a way to overcome these weak points of each technology for dynamic real environment application. In this paper, we propose a novel way to integrate two technologies via semantic filtering.

Section II describes the definitions and current status of two technologies. Section III describes a problem statement and describes how semantic filtering can handle the problem. Section IV summarizes the idea and mentions on further works.

II. Related works

A. Machine learning (ML)

Machine learning is a type of data analysis method which finds the optimal models and algorithms based on data. From data, it finds models & algorithms which explain patterns of the data and predict requested information. There are a lot of different machine learning methods such as decision tree, naïve Bayes, apriori, KNN, K-means, SVM, etc. [3]. Nowadays, the most loved one is a deep neural network approach such as CNN, RNN, DBM, etc. [4].

The strong points of machine learning are as followings: extract data features without human intervention, flexible adaptation on new data pattern (ex. online learning), high accuracy especially in computer vision and natural language processing and, generate new artifacts by using extracted features (Google's DeepDream). The weak points of it are as followings: noise-sensitive, balanced data assumption, insufficient exploitation of already acquired domain knowledge, etc.

B. Semantic Technology

Semantic technology is a different type of data analysis method which uses semantic rules designed by domain experts. Semantic rules are modelled by experts based on past experiences. When new data is coming in, semantic technology inferences new knowledge based both on new data and built-in semantic rules. Semantic Web is the prominent outcome of semantic technology [2]. Nowadays, with the proliferation of internet of things, semantic technology gets high attention as a heterogeneity resolver in internet of things world [5].

The strong points of semantic technology are as followings: exploit built-in human knowledge for data analysis, interpret data semantically and holistically. The weak points are as followings: inflexible rule management for upcoming new data patterns, strong human intervention.

C. New Move towards Integrating

There are roughly two research activities on combining two technologies, but these activities are at the starting point. AKSW(Agile Knowledge Engineering and Semantic Web) group[6] is developing methods, tools and applications for adaptive Knowledge Engineering in the context of the Semantic Web. They are using a Semantic Web Framework and compensate its limitation with supervised machine learning. In the same vein, there some papers on inductive learning for Semantic Web [7]. On the other hand, there is another activity regarding two technologies convergence. EIS(Enterprise Information Systems)[8] is considering semantic web knowledge as a machine learning input data. Therefore, they try to find a good feature representation of entities in Knowledge Graphs.

III. INTEGRATE TWO APPROACHES IN IOT WORLD

IoT will change everything in our life. Unconnected physical things are becoming connected to the internet and so revolutionary services will use those things for their information world and actionable world. In this IoT world, accurate recognition about surrounding circumstance is crucial to provide best-fit actionable services.

Until now, *data-driven* and *rule-based* approaches are not discussed together sufficiently. Intrinsically, they are complementary for each other. *Data-driven* approach is more agile towards dynamic data changes but does not utilize human expert knowledge sufficiently. And it analyzes the raw phenomena data which also includes noise data, and not-noise but-undesirable action/operational data. On the contrary, *rule-based* approach has more holistic view for data analysis. Human experts have learnt related domain knowledge through their careers and build semantic model (rules) based on those learnt experiences. But, rules are not sufficiently flexible for dynamic changing environment.

Based on these comparative analyses, we propose *semantic filtering* as a complementary instrument for machine learning and semantic technology. 1) *Semantic filter* extracts sound rules from machine learning-generate rules by using semantic information, 2) *Semantic filter* extends ontology with extracted rules in step 1) dynamically, and therefore ontology can be more adaptive towards dynamic changing environment.

In this paper, we describe a *device auto-configuration service scenario* for a smart home service as an example of a *semantic filtering*.

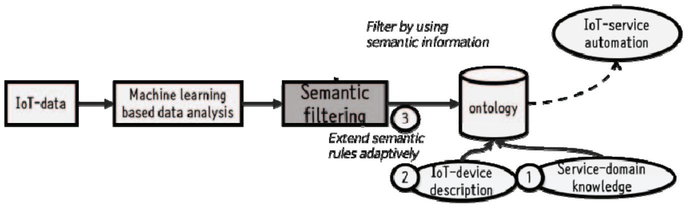


Figure 1. Complementary data analysis flow

A. *Data Collection Step: Internet of Things and Semantic Technology*

In figure 1, IoT data are flowed into machine learning. In a device auto-configuration scenario, we assume 6 devices 8 services [Figure 2].

Id	1	2	3	4	5	6	7	8
deviceName	DVDplay er	TV_audio	TV_player	TV_display	projector	Vacuums	Washing Machine	Wifi_speaker

Figure 2. Example device configuration

IoT device descriptions may be like Figure 3. It is just a simple example. Regarding real implementation with real devices, it needs to be more complex.

```
DVDplayer::                      Wifispeaker::
    deviceID = 1                     deviceID = 8
    serviceType = "player"           serviceType = 'sound"
    networkInfo = "WiFi"             networkInfo = "WiFi"
    Owner = abc                      Owner = cde
    valueType = boolean              valueType = boolean
```

Figure 3. IoT device description example

In the scenario, IoT-data is about service-on/off status. If ***a***-service and ***b***-service is on at the same time, [***a,b***] can be a data instance. Following is a simple example dataset in *Python*.

```
def loadDataSet():
    return [ [6], [7], [1,2,4,6], [2,3,4,7], [1,5,8], [1,2,4], [2,3,4], [1,5,6,8],
        [1,2,4], [2,3,4,7], [1,5,8], [1,2,4,7], [2,3,4], [1,5,8], [1,5,8], [1,5,8],
        [1,5,8] ,[6,8],[6,7,8],[6,7,8],[6,7,8],[6,7,8],[6,7,8],[6,7,8] ]
```

Figure 4. Dataset example

And, by using domain knowledge and IoT device description schema, we design a simple semantic model. This model is to be used for a *semantic filtering*.

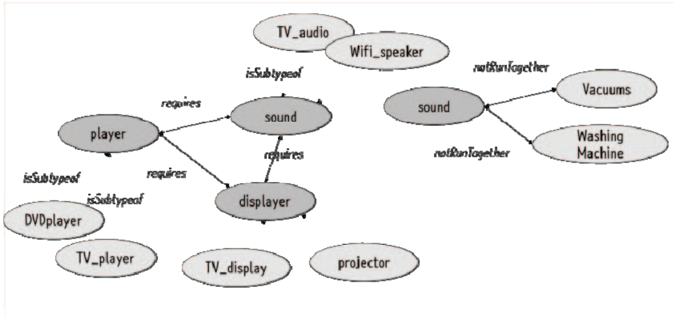


Figure 5. Semantic model example for an auto-device configuration

B. Adaptive Rules Learning Step: Machine Learning and Semantic Technology

- Machine learning analyzes IoT-data and generates association rules.
- *Semantic filter* extracts sound rules from ML-generate rules. To do this, it uses semantic rules which are designed by human experts [Figure 5].
- Extracted rules are added to ontology for later uses

This step is performed iteratively to adapt ontology for dynamically changing IoT environment.

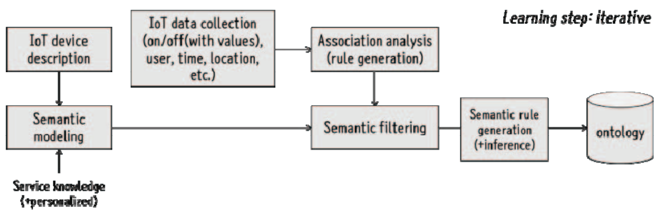


Figure 6. Rule learning step

We use *apriori* algorithm to extract association rules among multiple devices. A particle of a code is in Figure 7.

```
dataSet = loadDataSet()
L, suppData = apriori(dataSet, minSupport=0.2)
rules = generateRules(L, suppData, minConf=0.6)
```

Figure 7. Example code to extract association rules using *apriori*

Figure 8 is the resulting association rules by *apriori*. In this test, we use *support* value 0.2 and *confidence* value 0.6.

```
frozenset([5]) --> frozenset([1]) conf: 1.0
frozenset([1]) --> frozenset([5]) conf: 0.636363636364
frozenset([6]) --> frozenset([8]) conf: 0.8
frozenset([5]) --> frozenset([8]) conf: 1.0
frozenset([7]) --> frozenset([8]) conf: 0.6
frozenset([1]) --> frozenset([8]) conf: 0.636363636364
frozenset([7]) --> frozenset([6]) conf: 0.6
frozenset([6]) --> frozenset([7]) conf: 0.6
frozenset([4]) --> frozenset([2]) conf: 1.0
frozenset([2]) --> frozenset([4]) conf: 1.0
frozenset([5]) --> frozenset([8, 1]) conf: 1.0
frozenset([1]) --> frozenset([8, 5]) conf: 0.636363636364
frozenset([7]) --> frozenset([8, 6]) conf: 0.6
frozenset([6]) --> frozenset([8, 7]) conf: 0.6
```

Figure 8. Example association rules extracted by *apriori*

By applying *semantic filtering* [Figure 4], we get the filtered resulting rules in Figure 9.

```
frozenset([5]) --> frozenset([1]) conf: 1.0
frozenset([1]) --> frozenset([5]) conf: 0.636363636364
frozenset([5]) --> frozenset([8]) conf: 1.0
frozenset([1]) --> frozenset([8]) conf: 0.636363636364
frozenset([7]) --> frozenset([6]) conf: 0.6
frozenset([6]) --> frozenset([7]) conf: 0.6
frozenset([4]) --> frozenset([2]) conf: 1.0
frozenset([2]) --> frozenset([4]) conf: 1.0
frozenset([5]) --> frozenset([8, 1]) conf: 1.0
frozenset([1]) --> frozenset([8, 5]) conf: 0.636363636364
```

Figure 9. Filtered association rules by *semantic filtering*

As a result of this step, we could add two rules to ontology: ① DVDplayer(1), projector(5), and Wifi_speaker(8) run together ② TV_audio(2) and TV_display run together.

C. Intelligent Service Provision Step: Semantic Technology and Internet of Things

When a service initiation is triggered by a user, ontology is referred to find best-match device configuration for the service. By using IoT, selected devices are activated and get ready for the service. In a real situation, synchronization among related devices may be a difficult but crucial task.

Figure 10. Service provision step

If a user turns on the *DVDplayer* as a service triggering, then a device auto-configuration service turns on both *projector* and *Wifi_speaker* together and get them ready for a user's next action.

D. Overall Service Provision Lifecycle

For adaptive ontology evolution, a rule learning step iterates continuously therefore a data collection step should be performed continuously. A service provision step is performed when it is triggered by a user. Figure 11 shows the overall service lifecycle.

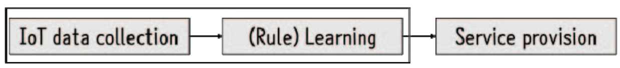


Figure 11. device-auto configuration lifecycle

IV. CONTRIBUTION AND FURTHER WORK

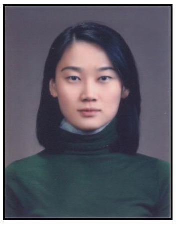
AI is one of the hottest research and development topics nowadays. Recognition of the situation is the key of AI and there is a data analysis at the core of AI. This paper proposes seamless and effective integration of machine learning and semantic technology in IoT environment. As a further work, overall framework needs to be refined, and real integrated implementation needs to follow.

ACKNOWLEDGMENT

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-15-05-ETRI).

REFERENCES

[1] A.P. Sheth, C. Ramakrishnan. Semantic (Web) Technology In Action: Ontology Driven Information Systems For Search, Integration and Analysis. IEEE Data Engineering Bulletin, 2003.
[2] http://cs231n.stanford.edu/syllabus.html.
[3] Peter Harrington, Machine Learning in Action,,April 2012 ,ISBN 9781617290183.
[4] Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, www.deeplearningbook.org.
[5] IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps, IERC position paper, March, 2015.
[6] AKSW, http://aksw.org/About.html
[7] www.semantic-web-journal.net/
[8] EIS blog, http://eis.iai.uni-bonn.de/blog/2016/04/on-semantic-web-machine-learning-and-feature-representation-of-entities-in-knowledge-graphs/



Marie Kim She worked for Samsung Electronics 1998~2000 and is a research engineer at ETRI since 2000.
Interested fields: Internet of Things, machine learning, sw platform.

Hyunjoong Kang He is currently a senior researcher at ETRI since 2009. He graduated from the Suncheon National University with a M.S Degree in Information and Communication Engineering in 2009.
Interested fields: Internet of Things, IoT Platform, knowledgebase and machine learning
.

YongJoon Lee He is principal researcher at ETRI and is currently working at KSB Convergence Research Department, ETRI.
Interested fields: Internet of Things, machine learning for IoT.

Soonhyun Kwon He is a senior researcher at ETRI and is currently working at KSB Convergence Research Department, ETRI. From 2006 to 2009, he was a member of engineering staff of Ontology Reasoner Development Team, IoT Platform Research Section.
Interested fields: Semantic Web, Linked data, and Ontology Inference.

Kwihoon Kim He studied in KAIST, M.S. degree in 2000. He worked in LG Dacom 2000~2005 and is a research engineer in ETRI since 2005. He is a team leader of self-machine learning research team, ETRI now.
Interested fields: Internet of Things, machine learning, reinforcement learning, knowledge-converged intelligent service.

Cheol Sik Pyo He studied in KAIST, M.S. degree in 1999. He is a leader of KSB Convergence Research Department, ETRI now.
Interested fields: Internet of Things, machine learning, knowledge-converged intelligent service.