# An Ontology-based Hybrid Approach for Accurate Context Reasoning

Muhammad Asif Razzaq[1], Muhammad Bilal Amin[2] and Sungyoung Lee[1]
Department of Computer Science and Engineering,
Kyung Hee University, 446-701, Republic of Korea
Email: {[1]asif.razzaq, [1]sylee}@oslab.khu.ac.kr, [2]m.b.amin@ieee.org

***Abstract*—The combination of ontology based context-awareness and machine learning context classification is an interesting research area. The determined contexts are obtained using semantic reasoning based on context ontology developed by expert using domain specific rules. This reasoning suffer challenges of soundness and completeness in real-time deployment. This paper addresses the aforementioned challenges from semantic reasoning by embracing machine learning modeling and classification benefits. Machine learning relies on data, for this we developed training and deployment phase for ontological ABox assertions. Approximately 99.99% precision through machine learning approach was achieved over 91.5% accuracy with semantic reasoning. The statistical evaluation proves the improvement in terms of accuracy for context prediction and overall performance.**

***Keywords*—Ontologies, Reasoning, Classification, Context-awareness.**

## I. Introduction

Context awareness provides good approximation to the user interaction with the surroundings. Context awareness is an essential ingredient of ubiquitous and pervasive computing systems [5] acting as a key technology, which results for applications to be aware of situation of their users and interaction with the environment. Dey and Abowd [1] defined the context as "Context is any information that can be used to characterize the situation of an entity." In [2] introduced the details of context aware systems, and their role in developing applications for end users. Nowadays, the most dominant context information consumers are mobile applications, which engage different capabilities of sensors in the shape of context [14]. A comprehensive survey was conducted by [12], which gives an overview of the challenges, opportunities and approaches for using machine learning in the Semantic Web. In the study by [6], provides sufficient evidence towards use of machine learning techniques in developing context-aware applications. These applications aim at sensing the context of the operating environment for optimizing interaction among applications and users. In the discussion by [7], compared accuracies of ten machine learning algorithms involved in real-time classification for interactive applications. Their work aimed for increasing processing efficiency, which they achieved by reducing training time for internet traffic real-time. A similar work was also conducted by [9] in which they used SPARQL query endpoint for learning relational Bayesian classifiers over RDF data. Nishihara et al. [11] highlighted latency and high throughput as the most important requirements for emerging machine learning applications. They proposed new distributed execution framework with underlying candidate approach having proof-of-concept architecture with significant increased performance. In an IoT environment Kim et al. [8] proposed seamless and effective integration of machine learning and semantic technology to compensate each other. In [10] authors illustrated comprehensive, highly inter-operable, reproducible and exchangeable classification methodology for ontology-based knowledge management and machine learning approaches using spatial data sources. In this paper, we focused on providing a machine learning based solution to increase the reasoning accuracy in addition to ontology based context-awareness in Mining Minds framework. The rest of the paper is structured as follows, Section II presents the motivation and objectives for this study. Section III describes overview of Mining Minds framework. Section IV discusses our proposed approach. Section V compares experimental evaluations. Finally, Section VI draw conclusions and future work.

## II. Motivation

Nowadays, more and more dynamic user's contextual information is becoming available to decision makers in context-aware systems. In a multi-user, environment, while working in real-time, with passage of time, ontological reasoning is challenging, as it suffers delay, and data loss, as contexts are needed in real-time. Also there is no mechanism of retransmission for sensory data and benefits of delayed reasoning. In order to meet reasoning QoS, there is dire need to strengthen real-time ontological reasoning for missing contexts and overcome unnecessary delays while processing them. In the study[13], it was observed that there are contexts for whom membership is not inferred, and were named as *unidentified high-level contexts*. These are contexts i.e. *unidentified high-level contexts*, which do not belong to any of the classes defined due to missing underlying contexts. Therefore, an approach needs to be designed to make *unidentified high-level contexts* to be part of context-aware system for better decision making. The key contributions in this study includes: motivating the use of machine learning algorithms for context awareness; classifying *unidentified HLC*; filling the gap of missing LLC and HLC, and increase in the accuracy and effectiveness for over all context-aware reasoning.

## III. Mining Minds: Framework

Mining Minds[1], an innovative framework based on concepts of the digital health and wellness paradigm, which

[1]http://www.miningminds.re.kr/

provides personalized healthcare and wellness support. To provide innovated services and utilize different tools Mining Minds framework architecture is divided into five distinct layers, *Data Curation Layer*, *Information Curation Layer* (ICL), *Knowledge Curation Layer*, *Service Curation Layer* and *Supporting Layer*. A detailed description of framework can be found in [4].

### A. High Level Context Awareness in a nutshell

In Mining Minds, the core technologies designed for the inference and modeling of the user's context constitutes the ICL, which is further subdivided to *Low-Level Context Awareness* (LLCA) and *High-Level Context Awareness* (HLCA) [13]. LLCA contains four main context categories `Activities`, `Locations`, `Emotions` and `Food` contexts recognized through respective recognizers. They are realized by applying machine learning approaches over sensory data obtained through smart-watch (accelerometer and gyroscope), smart-phone (accelerometer, gyroscope, GPS, camera and audio features), and kinect (skeletal postures using depth camera) [3]. `Food` items taken during meals are processed through user's specific tagged meal images using smart-phone. The HLCA relies on Mining Minds Context Ontology which is described in Section III-B.

### B. Mining Minds Context Ontology

The Mining Minds Context Ontology[2] ($M_2CO$) model contexts for human behaviour identification, additionally it also triggers the framework to provide personalized health and wellness services to its users. This work uses extended $M_2CO$ by Villalonga et al. [13] with the inclusion of additional nutrition information of the users in Mining Minds platform. This scalable $M_2CO$ comprehensively models Low-Level Contexts (LLCs), High-level Contexts (HLCs) such as: Physical Activity High-Level Context (PA-HLC) and Nutrition High-Level Context (N-HLC).

In the development of OWL2 semantic ontologies, `Classes` provide an abstraction mechanism for organizing resources with similar characteristics. Every OWL class can be linked with a set of individuals named as class extension. These individuals in the class extension are identified as the instances of the class. In Mining Minds, the `subclass` constraint is used for designing and assigning classes and individuals for LLC, HLC, PA-HLC and N-HLC `superclasses`. In Protégé, OWL `subclass` construct is used for defining class hierarchies. The superclass concept in this case is *LowLevelContext* with `subclass` concepts of *Activity*, *Emotion*, *Locations* and *Food*. Each `subclass` concepts have further their `subclass` concepts associated with `subclass` object property. Its meaning in OWL is exactly the same for instance if the class description *ClimbingStairs* is defined as `subclass` of class description *Activity*, then all set of individuals in the class extension of *ClimbingStairs* must be a subset of the set of individuals in the class extension of Activity. This fulfills definition requirement for `subclass`. These classes are associated to the *HighLevelContext* class through the object properties hasActivity, hasLocation, hasEmotion and hasFood. The hasActivity property has domain as the *LowLevelContext* Context class and range as the *Activity* class. Similarly, hasLocation, hasEmotion and hasFood properties have domains as *LowLevelContext* class and ranges *Location*, *Emotion* and *Food* classes respectively. The different recognized activities are modeled as 17 disjoint subclasses of the *Activity* class. Similarly the *Location* class has 9 disjoint subclasses used to model the detected locations. The recognized emotions are modeled through the 9 disjoint subclasses of the *Emotion* class. PA-HLC are modeled based on `activities`, `locations` and `emotions`. PA-HLC includes `OfficeWork`, `Housework`, `Sleeping`, `Commuting`, `Exercising`, `Amusement`, `Gardening`, and `Inactivity` discussed in detail with definitions by Villalonga et al. [13]. N-HLC consists of major dominating nutrient like `Carbohydrates`, `Protein` and `Fats` determined using LLC. These N-HLC are modeled on LLC `Food` based on 57 food-item list, categorized further into 10 broader groups (Fig. 1). This categorization and major nutrient identifications are performed in accordance with guidance and suggestions provided by United States Department of Agriculture[3] (USDA) on daily food consumptions.

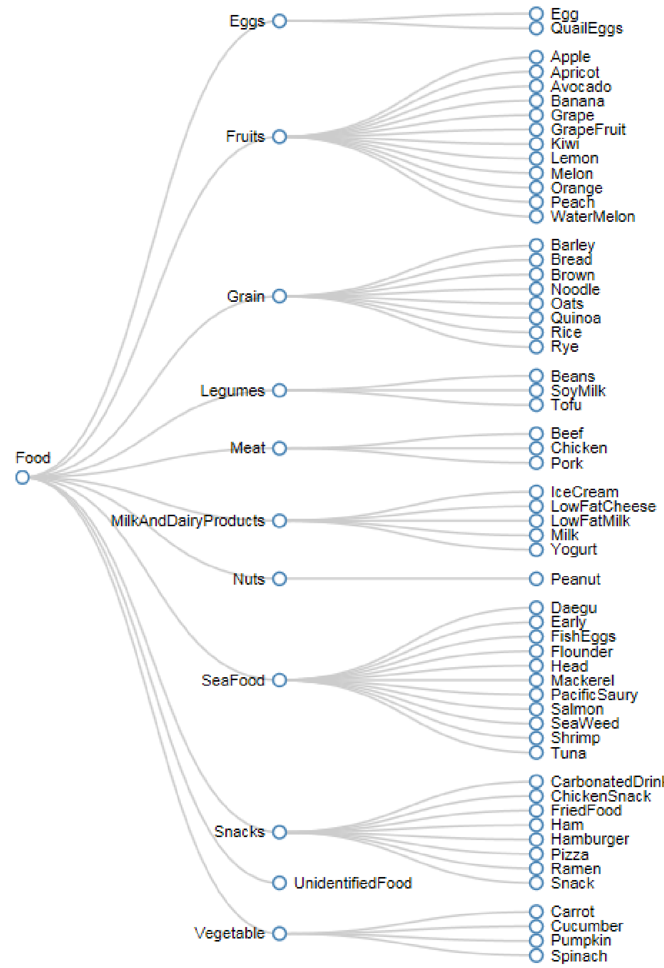


Fig. 1. Partial view of Mining Minds Context Ontology ($M_2CO$)

---

[2] $M_2CO$: https://goo.gl/EBA4nO

[3] http://www.usda.gov/wps/portal/usda/usdahome

## IV. Our Approach

In this work, we presented a novel conjunctive approach for ontology based reasoning and machine learning for real-time context-awareness modeling and classification. Machine learning supports inferencing by classifying high-level contexts based on ABox dataset stored in Jena triple store, in *Mining Mining Context Ontology Storage*. We performed all experiments on a single 64-bit machine, having Microsoft Windows 7 on the top of AMD A10-5800K APU with 12 GB of RAM. Java engine was executed with Eclipse IDE Luna that ran using JRE 1.8.0_45-b15 64-bit version. The implementation[4] of high-Level context architecture is performed in Maven[5] based project management with Java along-with available open source library, Apache Jena[6] (v2.11.2). For experimental evaluation, we collected real-time dataset by involving 20 users on different days and timings. User's were provided with smart-watch, smart-phone, and skeletal postures were detected using indoors kinect facilitated with depth camera. We collected 26,298 HLCs (i.e. PA-HLCs and N-HLCs) through HLCA reasoning. By analyzing these contexts, it was found that there are almost 83,757 LLC, which fulfilled the ontological rule definitions for HLCs.

### A. Ontology based reasoning in HLCA

The main purpose of ontology based context reasoning is to check the consistency of contexts as well as deducing high-level implicit context information from low-level explicit contexts. Ontological reasoning uses constraint based *subsumption* and *equivalence* rules developed by domain expert while modeling $M_2CO$. We considered Pellet[7] reasoner with essential characteristics of its deductive expressivity, tableau-based methodology and OWL-API support for performing reasoning and provision of ground truth using $M_2CO$ in *HLCA*.

### B. ABox Classification using Machine Learning

In contrast to ontological reasoning, machine learning works on the basis of data-driven approaches. Using input selected features, target class is predicted based on trained model. The dominating feature for machine learning approach is that this does not require domain expert intervention and gives high accuracy. During ontological reasoning the inferred PA-HLCs and N-HLCs are stored in triple form to Jena-TDB[8]. This information can be retrieved using SPARQL[9] from *Context Ontology Storage* i.e. Jena-TDB in RDF/XML graph form, which comprises of recognized HLCs, LLCs and user's meta-data such as userid, starttime, endtime etc. After carefully analyzing the collected dataset, it was learned that almost 91.50% were correctly inferred, we need to investigate other contexts which could not be inferred due to missing low-level contexts. For this a generic machine learning approach is shown in (Fig. 2) comprising of standardized methods for transformation of RDF/XML to arff, essential machine learning tasks for feature vector extraction from RDF graphs, algorithm selection and model training using extracted feature vectors (Activities, Locations, Emotions, Food LLCs and Target HLCs i.e. PA-HLCs & N-HLCs) in off-line manner. The trained models were later tested with test dataset to classify and predict HLCs for *unidentified HLCs* before exhibiting to real-time deployment for context-awareness.

[4]Source code at github: https://goo.gl/0ZFNif
[5]https://maven.apache.org/
[6]https://jena.apache.org/
[7]https://www.w3.org/2001/sw/wiki/Pellet
[8]https://jena.apache.org/documentation/tdb/
[9]https://jena.apache.org/documentation/query/index.html

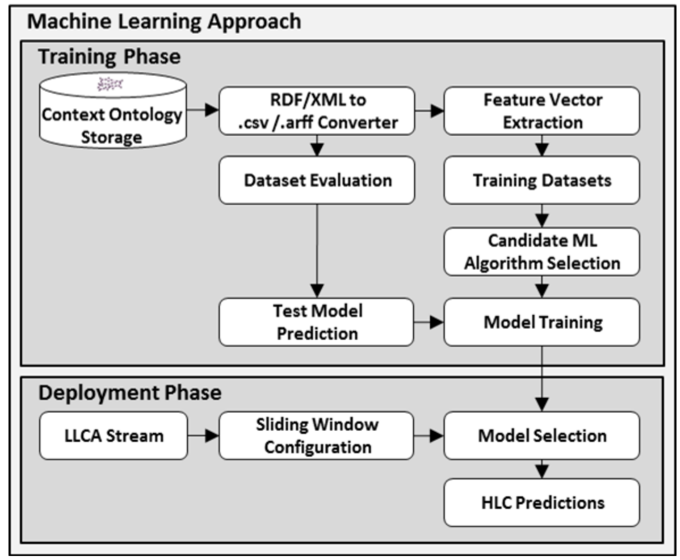


Fig. 2. ML based Training and Deployment phases.

## V. Experimental Evaluations

In this section, we discussed comprehensive evaluations for ontology-based reasoning and ML based prediction of *unidentified HLCs* for HLCA in Mining Minds health platform.

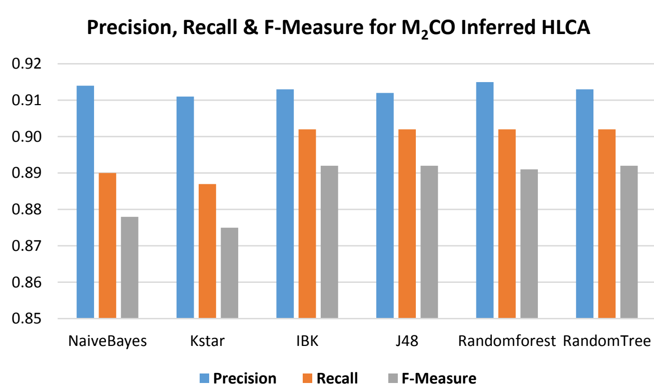


Fig. 3. Precision, Recall and F-Measure for Inferred HLCs using $M_2CO$ on Real-time dataset

*1) Evaluation Metrics:* For a collected dataset, we evaluated how HLCA correctly inferred high-level contexts based on low-level contexts for 15-second sliding window. We used the precision and recall measures to evaluate HLCA in Mining Minds as shown in (Fig. 3) to address our performance challenge. In order to classify *Unidentified HLC* in HLCA, we used six algorithms within Weka Explorer, which are *NăiveBayes*, *KStar*, *IBKm J48*, *RandomForest*, and *RandomTree*. These algorithms were applied under ML approach to both training and deployment phases. Fig. 3 illustrates analysis for Precision, Recall and F-Measure in order to evaluate inference performance of $M_2CO$ using 10-fold cross validation. It's evident that *RandomForest* algorithm produced optimum results of 91.5% precision, compared with other algorithms with high values in Precision, Recall and F-Measure. In addition to that, *Kstar* provided the lowest 87.5% when compared with other algorithms. The high inferencing accuracy proved that HLC are rarely misclassified but the problem arise when missing LLCs lead to wrongly recognized HLCs or *Unidentified HLCs*.

*2) Performance Evaluation:* In this section, we present an analysis of $M_2CO$ and ML based classification. We achieved 91.5% precision through ontology based reasoning, remaining HLCs were not recognized correctly because of missing LLCs, for this we adapted training phase in ML approach (Fig. 2). We trained and tested our dataset for different algorithms as discussed in previous section. We provided test dataset comprising *Unidentified HLC* to the trained model in order to predict their HLC class. The probabilities of predicted classes are plotted in (Fig. 4). As it is clearly seen that instances are classified with high probabilities for class prediction, we ignored values less than 0.50 and analyzed that probabilities greater than 0.50 were predicted correctly, which resulted in accuracy increase up to 99.99%. This proved to be very effective technique used to classify missing HLCs in HLCA.

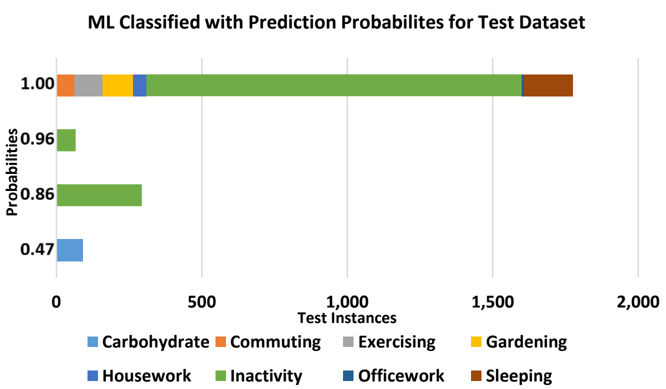


Fig. 4. *Unidentified HLC* Classification using test dataset over learned models

## VI. Conclusions and Future Work

We presented an ontology-driven HLCA under Mining Minds Platform enabling richer semantic basis for interpreting the HLCA, as compared to approaches, which employ the database schema directly as the basis. In this paper, we developed methodology to address inferencing performance using machine learning techniques. The proposed approach meets the stringent accuracy and completeness requirements of real-time reasoning in Mining Minds, a context-aware platform. ML based approach fortifies the identified HLCs, in addition to ontological based approach, by enhancing the precision performance from 91.5% to 99.99%. We proposed an hybrid approach to handle complex context-aware definitions real-time for ontology based context recognition and ML based supervised context classification. In future, we will investigate correlation between ontology characteristics and ML learning algorithms.

*Acknowledgments.:* This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea), this research was supported by Korea Research Fellowship program funded by the Ministry of Science, ICT and Future Planning through the National Research Foundation of Korea (NRF-2016H1D3A1938039) and this research was also supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-01629) supervised by the IITP (Institute for Information & communications Technology Promotion).

## References

[1] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *Handheld and ubiquitous computing*, pages 304–307. Springer, 1999.

[2] Matthias Baldauf, Schahram Dustdar, and Florian Rosenberg. A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4):263–277, 2007.

[3] O. Banos, C. Villalonga, J. H. Bang, T. H. Hur, D. Kang, S.-B. Park., T. Hyunh-The, L. B. Vui, M.-B. Amin, M.-A. Razzaq, W. Ali Khan, C.-S. Hong, and S. Lee. Human behavior analysis by means of multimodal context mining. *Sensors*, 16(8):1–19, 2016.

[4] Oresti Banos, Muhammad Bilal Amin, Wajahat Ali Khan, Muhammad Afzel, Mahmood Ahmad, Maqbool Ali, Taqdir Ali, Rahman Ali, Muhammad Bilal, Manhyung Han, et al. An innovative platform for person-centric health and wellness support. In *Bioinformatics and Biomedical Engineering*, pages 131–140. Springer, 2015.

[5] Vijay Borges and Wilson Jeberson. Survey of context information fusion for sensor networks based ubiquitous systems. *arXiv preprint arXiv:1309.0598*, 2013.

[6] Antonina Danylenko, Christoph Kessler, and Welf Löwe. Comparing machine learning approaches for context-aware composition. In *International Conference on Software Composition*, pages 18–33. Springer, 2011.

[7] Hamza Awad Hamza Ibrahim, Sulaiman Mohd Nor, Aliyu Mohammed, Abuagla Babiker Mohammed, et al. Taxonomy of machine learning algorithms to classify real time interactive applications. *International Journal of Computer Networks and Wireless Communications*, 2(1):2012, 2012.

[8] Marie Kim, Hyunjoong Kang, Soonhyun Kwon, YongJoon Lee, Kwihoon Kim, and Cheol Sik Pyo. Augmented ontology by handshaking with machine learning. In *Advanced Communication Technology (ICACT), 2017 19th International Conference on*, pages 740–743. IEEE, 2017.

[9] Harris Lin, Neeraj Koul, and Vasant Honavar. Learning relational bayesian classifiers from rdf data. *The Semantic Web–ISWC 2011*, pages 389–404, 2011.

[10] Niklas Moran, Simon Nieland, Birgit Kleinschmit, et al. Combining machine learning and ontological data handling for multi-source classification of nature conservation areas. *International Journal of Applied Earth Observation and Geoinformation*, 54:124–133, 2017.

[11] Robert Nishihara, Philipp Moritz, Stephanie Wang, Alexey Tumanov, William Paul, Johann Schleier-Smith, Richard Liaw, Michael I Jordan, and Ion Stoica. Real-time machine learning: The missing pieces. *arXiv preprint arXiv:1703.03924*, 2017.

[12] Volker Tresp, Markus Bundschus, Achim Rettinger, and Yi Huang. Towards machine learning on the semantic web. In *Uncertainty reasoning for the Semantic Web I*, pages 282–314. Springer, 2008.

[13] Claudia Villalonga, Muhammad Asif Razzaq, Wajahat Ali Khan, Hector Pomares, Ignacio Rojas, Sungyoung Lee, and Oresti Banos. Ontology-based high-level context inference for human behavior identification. *Sensors*, 16(10):1617, 2016.

[14] Meisong Wang, Charith Perera, Prem Prakash Jayaraman, Miranda Zhang, Peter Strazdins, and Rajiv Ranjan. City data fusion: Sensor data fusion in the internet of things. *arXiv preprint arXiv:1506.09118*, 2015.