Predicting Energy Consumption of Ontology Reasoning over Mobile Devices

Isa Guclu[1]([✉]), Yuan-Fang Li[2], Jeff Z. Pan[1], and Martin J. Kollingbaum[1]

[1] University of Aberdeen, Aberdeen, UK
r01ig15@abdn.ac.uk
[2] Monash University, Melbourne, Australia

Abstract. The unprecedented growth in mobile devices, combined with advances in Semantic Web (SW) Technologies, has given birth to opportunities for more intelligent systems on-the-go. Limited resources of mobile devices demand approaches that make mobile reasoning more applicable. While Mobile-Cloud integration is a promising method for harnessing the power of semantic technologies in the mobile infrastructure, it is an open question how to decide when to reason over ontologies on mobile devices. In this paper, we introduce an energy consumption prediction mechanism for ontology reasoning on mobile devices that allows an analysis of the feasibility of performing an ontology reasoning on a mobile device with respect to energy consumption. The developed prediction model contributes to mobile–cloud integration and helps to improve further developments in semantic reasoning in general.

Keywords: Energy · Semantic web · Ontology reasoning · Mobile computing · Prediction · Random forests

1 Introduction

Server and desktop machines have been the main environment for ontology reasoning in assisting knowledge management so far. With the rapid improvement of hardware capabilities, as well as software developments in mobile devices (e.g., smartphones, tablets, PDAs, smartwatches), semantic reasoners start to become adopted [24] in mobile environments. Mobile applications (apps) that use semantic technologies, such as for the integration with diverse data sources and knowledge due to inferences made during semantic reasoning, are also being developed. However, this potential seems not fully utilized yet.

According to Yus and Pappachan's research [25] on semantic mobile apps, 23 out of 36 apps implement a client-server architecture, where the mobile app is used as an interface for the results processed by a server and just 6 apps utilize a semantic reasoner directly on the device to infer facts. According to their observation, the use of Semantic Web technologies on mobile devices is on the rise and there is a need for the development of more tools to facilitate this growth [25]. Groth [6] asserts that, while there has been a large amount of effort in Semantic

Web Services, even going as far as developing a standard for describing those services, we have not seen a corresponding take-up in using these languages to enable the execution of actions either on the Web or in the real world.

Our goal is to make semantic technologies more feasible for a new era of mobile and cloud computing by building an energy prediction mechanism that will guide us "to what extent ontology reasoning can be made on mobile devices". *Mobile-Cloud Integration* can significantly enhance the capabilities and benefits of semantic technologies. For a successful mobile-cloud integration, a mechanism is needed that will (1) predict the cost of data processing (including loading, parsing, reasoning, query answering) on a mobile device itself in terms of time and energy consumption, (2) predict the cost of data processing on the cloud, and (3) ultimately determine where data processing should be conducted in an optimal way.

In this paper, we focus on the *Prediction of Energy Consumption* aspect of ontology reasoning *on the mobile front*, using statistical methods and execution data collected during experiments. We present an energy consumption prediction mechanism that predicts how much energy a new ontology will consume and whether this ontology may be processed within a predefined time, using previous reasoning results and specific metrics for ontologies [12].

We focus on the energy consumption aspect of semantic data processing, because, as [15,19] pointed out, *energy consumption* is a principal design concern for mobile platforms, rather than just a desirable attribute. Our investigation (see Sect. 5.2) shows that it *cannot* be assumed that the energy consumption of a reasoning process on a mobile device correlates with its time consumption. The main contributions of this paper can be summarised as follows:

1. We show that metrics of ontologies are very effective for accurate prediction (having R^2 between 0.8985 and 0.9859, and a maximum $RMSE$ of 10.86) of energy consumption of ontology reasoning on the Android platform, as validated by our comprehensive evaluation.
2. A comprehensive dataset ontologies in OWL 2 EL profile (a tractable profile in OWL 2) is made available for assessing and improving the performance of reasoning algorithms in terms of energy consumption.

2   Related Work and Background

Kleemann [14] discusses resource limitations in terms of computing power, memory and energy and presents a study for the development of a reasoner suitable for resource constrained environments such as mobile devices. Cerri et al. [3] propose the "knowledge in the cloud" approach, extending "data in the cloud" with support for handling semantic information, such as organising and finding it efficiently and providing reasoning and quality support. Despite presenting an efficient approach for harnessing the power of the cloud, this study is limited to cloud computing and doesn't take into account the capabilities of mobile devices. Rietveld and Schlobach [20] present a study about how the constraints in computing environments influence SW applications. In their study, they take *battery*

power as one constraint, however, no deeper study is provided about the relationship between the energy need of the application and the structure of semantic data. Corradi et al. [4] propose an architecture and describe a prototype system for a mobile–cloud support of semantically enriched speech recognition in social care. In their approach, they move resource-demanding tasks that consume a high amount of energy on a mobile device to the cloud computing infrastructure. Hogan et al. [11] discuss scalability issues of reasoning and propose an approach for making the processing of a billion triples of open-domain Linked Data feasible. While they contribute to the feasibility of semantic data processing with regard to complexity, energy consumption of these approaches haven't been investigated.

Metrics of ontologies have been used for assessing the quality [2], complexity [26], cohesion [23], population task [16] and time consumption [13] of ontology reasoning. Hasan and Gandon [8] implemented a machine learning approach for predicting the performance of SPARQL queries using previous execution data. These investigations targeted server machines and efficient results were obtained. In our investigation, we are going to make use of metrics to deal with the energy bottleneck of mobile devices.

2.1   Electric Power and Energy Consumed

(Electric) Power (P) is the rate of doing work, measured in *watts*. The electric power in watts produced by an electric current I passing through an electric potential (voltage) difference of V is,

$$P = V * I, \quad (watts = volts * amperes). \tag{1}$$

Energy (E) is equal to the power *(P)* times the time period *(t)* is,

$$E_J = P_W * t_s, \quad (joules = watts * seconds). \tag{2}$$

We measure *Energy Consumed*, in *watt-seconds (Ws.)*, which is equal to joules.

2.2   Measuring Energy Consumption Programmatically

Various techniques [7,22] have been used to measure and predict energy consumption on mobile devices. For measuring energy consumed in reasoning, Patton and McGuinness propose a power benchmark [18] using a physical device setup that consists of a power monitor and a notebook computer to collect data. Because of the difficulty of implementing hardware-dependent (requiring any extra equipment not natively available on the mobile device) techniques to a solution that is desired to be applicable to all mobile devices, we searched for a software-based technique that can be programmatically implemented.

We adopted the Power Consumption Benchmark Framework [21] proposed by Valincius et al., which is hardware-independent and easily programmable[1].

---

[1] https://github.com/evalincius/PowerBenchMark,   https://github.com/evalincius/HermitOWLAPI.

Energy is calculated using the properties in Android's `BatteryManager` class, `BATTERY PROPERTY CURRENT NOW` and `EXTRA VOLTAGE`, *Current* and *Voltage* are retrieved (see Eq. (1)). Valincius et al. measured total energy consumption with interval of 1 s (see Eq. (2)).

Observing and measuring overall energy consumption and battery drainage of the mobile device in 1 sec intervals poses a problem – a measurement with a resolution of 1 s shows the energy consumed by the mobile device during this second, independent of whether during such a measuring interval the processing of an ontology lasts 1 ms or 1000 ms. In order to get a more accurate measurement of how much energy the processing of a single ontology consumes, we, therefore, increased the precision of measurements by shortening the interval to 100 ms. In order to do this, we recorded the value of $\frac{V*I}{10}$ with intervals of 100 ms from the start of the data processing to the end. The cumulation of these values constitutes the ***total energy consumed***. With that, we reach a precision of 100 ms. With this method, an ontology processed in 1 ms is measured to consume the energy calculated for 1 interval of 100 ms. And, an ontology processed in 1000 ms is measured to consume the energy calculated for 10 intervals of 100 ms cumulatively.

2.3   Ontology Metrics

To be able to capture the complexity of ontologies thoroughly, we have adopted the set of 91 metrics proposed by Kang et al. [12,13]. These metrics include the number of general class inclusions, number of individuals, and the count of additional types of logical axioms (including reflexive properties, irreflexive properties and domain/range axioms). There are 24 ontology-level metrics to measure the overall size and complexity of an ontology, 15 class-level metrics to measure characteristics of OWL classes in an ontology, 22 anonymous class expression metrics to capture different types of class axioms, 30 property definition and axiom metrics to capture different types of property declarations and axioms. The complexity of all the metrics calculation algorithms is polynomial [13] in the size of the graph representation of the ontology.

2.4   Statistical Methods for Energy Prediction

We use a series of statistical methods for our energy prediction. *Regression Analysis* [9] is a statistical tool for the investigation of relationships between variables using some predictor variables and an output variable. We have built a regression model in which metrics are the predictor variables and the overall energy consumption of processing an ontology is the output variable. The output variable is denoted by Y, and the set of predictors by a vector X ($X_1, X_2, \ldots, X_n$, where n is the number of predictor variables). A regression model is formalized as $Y \approx f(X, \beta)$, where β is the *unknown parameters*, X is the *independent variables* and Y is the *dependent variable*. *Classification* identifies to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. In our

classification model, metrics are the predictor variables and the output variable states either "able to process the ontology in 100 s " or "not able to process the ontology in 100 s". *Random Forest* [1] is an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. In this paper, we train Random Forest-based classification models to predict whether an ontology can be processed within a predefined time and Random Forests-based regression models to predict energy consumption for an ontology using the power benchmark introduced above and syntactic metrics as features. *A Moving Average* is a calculation to analyse data points by creating series of averages of different subsets of the full data set in statistics. A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. We will use moving average to see whether there is a trend in the energy consumption while the battery level is decreasing from 100 % to 1 %. *The Coefficient of Variation* (CV), also known as "relative variability", is a standardized measure of dispersion of a probability distribution or frequency distribution. It is often expressed as a percentage, and is defined as the ratio of the standard deviation to the mean (or its absolute value). In our work, we will use CV for examining the variability of the energy measurement results from 100 % to 1 % battery level.

3   Our Approach

Making an energy prediction mechanism is a challenging task. Firstly, as detailed in Sect. 5.2, there *may not be always* a linear relation between time and energy consumption of ontology reasoning on every device. Hence, prediction models for reasoning time, and those done in a desktop/server environment (such as [12,13]), cannot be re-used as-is. Secondly, trying to model all the variables of real-world environments for energy consumption prediction, especially for a mobile device, is *very difficult*. Adapting to improvements in mobile environments is another complication for developing predication models, as changes to operating systems or in the utilisation of the CPU may render existing models obsolete.

In addressing these challenges, we developed prediction models by using a programmable (and hardware-independent) energy measurement tool ("Power Consumption Benchmark Framework" [21] proposed by Valincius et al.) and we use metrics that provide us with a numerical representation of particular properties of an ontology and use this information as our data source, in order to deal with the complexity of the semantic web and the uncertainty of internal and external influences on measuring energy consumption during ontology reasoning processes on mobile devices. We have chosen ontologies[2] in EL profile which were used at the ORE 2014 (The OWL Reasoner Evaluation Workshop 2014). We used two substantially different devices for our experiments. The measurement and prediction results will also provide an opportunity to

---

[2] https://zenodo.org/record/10791.

identify unforeseeable effects due to changes in the environment. For example, we observed that energy consumption may not always be correlated with reasoning time during our experiments with Machine2. Our case scenario (work-flow of the mechanism) has the following steps:

1. The mobile device asks the server whether it shall try to perform a reasoning task locally by sending the IRI of the ontology.
2. The server takes measures of the ontology according to the ontology metrics discussed in Sect. 2, and applies those measurement results to the classification model, which is trained to predict whether this ontology can be processed on this mobile device with a chosen reasoner using the metrics of the ontology, and will return either:
   (a) *Positive*: "this ontology can be processed within the predefined time limit (100 s in our experiment), and (using the regression model for predicting the energy consumption) it will consume this amount of energy"; or
   (b) *Negative*: "it cannot be processed in the predefined time".
3. If the mobile device gets a positive result from the server, it will then analyse the remaining energy available on the device and decide whether to proceed locally or in the cloud. If the mobile device receives a negative result, it will wait for the cloud to perform the reasoning task and return the result. For experimental purposes, we will process all the ontologies on the device. If the process exceeds the predefined time limit (100 s in our experiments), the process will be terminated.
4. The server will be informed whether the reasoning finishes with success within the set time limit.
5. The data collected about energy and time consumption will be used to improve our model to produce better prediction results.

Our approach regards mobile device as a "black-box" and accepts all its internal/external influences over data processing as the nature of it. We gather the execution data produced by the device and make inferences using this data with prediction data. We measure overall energy consumed (including loading/parsing of ontology, classification of the ontology TBox and executing the SPARQL query to retrieve the classification result) during the processing of ontologies in EL profile on each mobile device-reasoner pair. We describe experiments with a particular query answering task (explained in Sect. 5.1) by sending the same SPARQL query to the two reasoners we investigate, in order to get results for subsumption reasoning. Experiment results of these ontologies are used to make a prediction model and predict energy consumption of a new ontology on the same device-reasoner pair. This prediction mechanism is validated by the statistical results obtained from experiments.

In our experiments, we have implemented a separate model for each device-reasoner pair to see its validity in that scope. We are planning to work on one model for classification and one model for regression of all device-reasoner pairs as future work.

4   Experimental Setup

For calculating the error rate of our classification model, we divide wrong predictions by total predictions. For deciding whether our regression model is acceptable to describe the relation between the variables and the result obtained from the model, we have referred to R^2 and $RMSE$. The coefficient of determination (R^2) is a key output of regression analysis, which indicates the extent to which the dependent variable is predictable. An R^2 of 0.95 means that 95 percent of the variance in dependent variable can be predictable from independent variables. The Root Mean Squared Error ($RMSE$) is simply the square root of the mean/average of the square of all of the error. $RMSE$ represents the sample standard deviation of the differences between predicted and observed values.

4.1   Data Collection

Reasoners: We have used HermiT [5], a DL reasoner, and TrOWL, an EL reasoner, (version 1.5, ported on Android) as testing reasoners. We implemented an android-ported version[3] of HermiT provided by Yus et al. [24], as the desktop version could not be directly supported by Android Runtime (ART).

Ontologies: The ORE2014 Reasoner Competition Dataset is chosen as the dataset for our experiments. The OWL 2 EL Profile [10] is chosen, because the computational complexities of ontology consistency, class expression subsumption, and instance checking are all *PTIME-Complete* [17] and both reasoners support it natively. From 16,555 ontologies, 8,805 ontologies, which are in EL profile, were filtered. The RDF/XML format was used in the experiments; however, the validity of our prediction mechanism does not depend on a particular input format. As we have built an extendible prediction mechanism, in the future, other formats can be introduced easily. Being aware of the RAM limitation (and reasoning limitation as a consequence), we ordered ontologies according to their file sizes and started with the ones with a smaller file size. Each device-reasoner pair is analysed with the ontologies it could process within mobile-specific and time limitations. There are 17 cases of exceptions throughout the experiment, which contain 14 "InconsistentOntology" exceptions, 2 "ConcurrentModification" exceptions and 1 error for receiving a voltage value of zero (0) from the operating system. Details of the exceptions are available online[4]. The "0 Voltage" problem occurred just once during the processing of ca. 8000 ontologies. Our point of view, therefore, is that such a low frequency of occurrence of these errors do not invalidate our results. These exceptional 17 cases are excluded in the model generation.

Mobile Devices: We used two mobile devices (Machine1 and Machine2) that have substantially different hardware specifications[5]. Machine1 had the Android 5.1.1

---

[3] https://github.com/evalincius/Hermit_1.3.8_android.

[4] https://github.com/IsaGuclu/PredictionOfEnergy/blob/master/Exceptions.xlsx.

[5] Unofficial   comparison: http://www.phonearena.com/phones/compare/
Samsung-Galaxy-S6,Sony-Xperia-Z3-Compact/phones/8997,8744.

as the OS and Machine2 had Android 6.0.1 as the OS. To avoid the side-effects of other services and processes, we uninstalled apps that could be uninstalled, closed all services and GSM connection and opened the Wi-Fi connection in all experiments to enable TBox retrieval from the internet if needed. For avoiding side-effects of the sensors, we closed location services and kept the device in a fixed place to avoid triggering sensors, e.g., accelerometer, gyro, proximity, compass, barometer, etc. We closed all sort of energy saving utilities in the settings of the machine to have near standard conditions in experiments. Ontologies are run on the same machine sequentially.

Data Preprocessing: Before training every model, to avoid misleading consequences, predictor metrics with zero standard deviation are discarded.

In the experiments with the TrOWL EL Reasoner, 61 of the metrics have been chosen for training the classification model and 60 of the metrics for the regression model. In the experiments with the HermiT Reasoner, 58 of the metrics have been chosen for training the classification model and 57 of the metrics for the regression model.

Prediction Model Construction: For the 1st prediction ("Will this ontology be reasoned in 100 s on this device-reasoner?"), a random forests based classification model is implemented. For the 2nd prediction ("How much energy will this ontology consume on this device-reasoner?"), a random forests based regression model is implemented. Standard 10-fold cross-validation is performed to ensure the generalizability of models.

5   Results and Evaluation

Before starting experiments, we had questions about how accurately we could collect energy consumption data from a mobile device by taking the measures explained above to eliminate the side-effects of mobile devices. The voltage provided by the battery continuously decreases during each reasoning activity. Heating may have adverse effect on computations as the CPU may slow itself when it reaches some threshold. As the Wi-Fi connection is open throughout the reasoning process, there would be some effects of OS-based or manufacturer specific apps on measurements. If measurement results change through the battery level from 100 % to 1 %, this will make a generalizable approach impossible. This made us investigate the *standard error of the mean caused by these (and those that we may not foresee) side-effects in our experiments.*

Experiments were started with fully charged Machine1 and Machine2. We repeatedly reasoned over the same ontology[6] using the TrOWL EL reasoner until the battery was completely drained. In this experiment, we made the following observations. In *Machine1* (Fig. 1), the average energy consumption for an ontology reasoning task is 151.16 Ws., with a standard deviation of 5.91 Ws. The average duration of the reasoning is 74.06 s and standard deviation is 1.73 s.

---

[6] https://github.com/IsaGuclu/PredictionOfEnergy/blob/master/From100To1PercentBatteryLevel/approximated_00518.owl_RDFXML.owl.
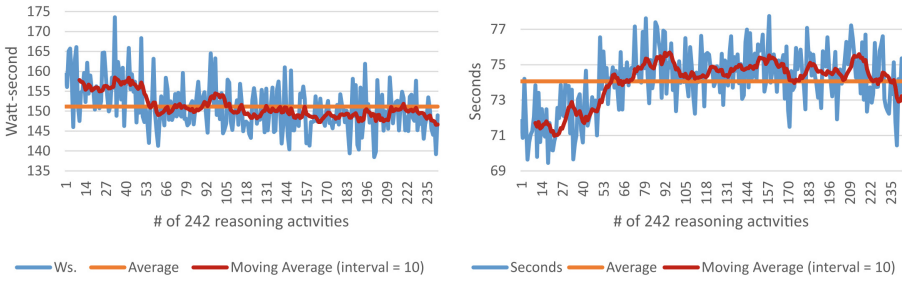
Fig. 1. Machine1 (Android 5.1.1) - Energy/Time consumption with battery level from 100 % to 1 %. *This figure illustrates (1) the energy/time consumed in experiments, (2) the average energy/time consumption of all experiments, (3) moving average of energy/time consumption with interval of 10.*
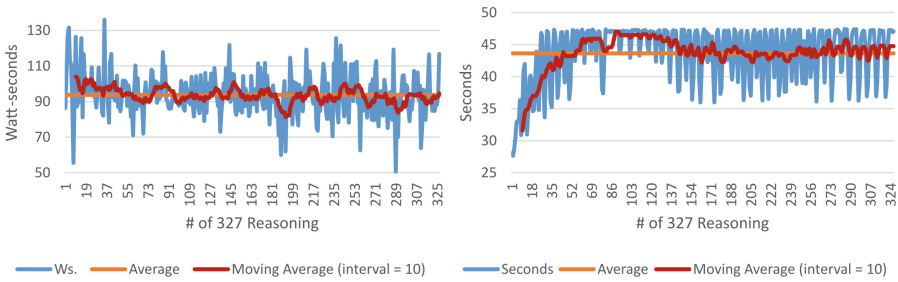

Fig. 2. Machine2 (Android 6.0.1) - Energy/Time consumption with battery level from 100 % to 1 %. *This figure illustrates (1) the energy/time consumed in experiments, (2) the average energy/time consumption of all experiments, (3) moving average of energy/time consumption with interval of 10.*

We found 3.91 % as the CV (standard deviation of energy consumed divided by the average of energy consumed) of energy measurement for this *machine-reasoner pair*. To see whether this result is generalizable, we made the same experiment with a substantially different machine (i.e., Machine2). In *Machine2* (Fig. 2), the average energy consumption of the ontology is 93.76 Ws., with a standard deviation of 11.92 Ws. The average duration of the reasoning is 43.64 s and the standard deviation is 4.34 s. We found 12.71 % as the CV of energy measurement for this *machine-reasoner pair*. This result made us search for the reason of such a difference. One of the biggest differences between the two machines is that Machine2 has a CPU which has one 2.1 GHz. quad-core processor and one 1.5 GHz. quad-core processor, but Machine1 has one 2.5 GHz. quad-core processor. To see what kind of a behaviour does the CPU have during our experiments, we used Usemon(CPU Usage Monitor)[7] and three sample execution of Machine2 is illustrated with Fig. 3. Figure 3 shows observations made on Machine2, where different cores, which have different clock-pulses, are used

---

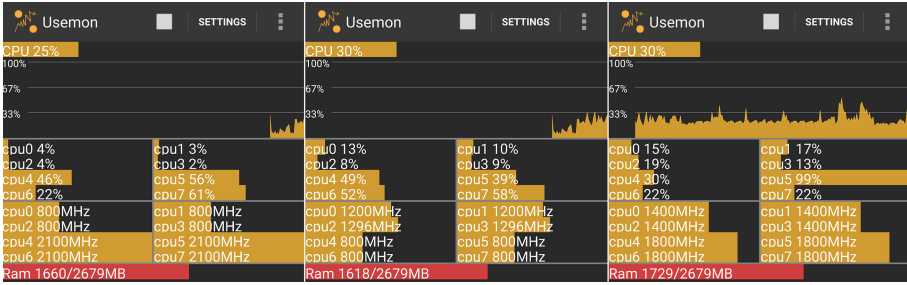[7] https://play.google.com/store/apps/details?id=com.iattilagy.usemon.

Fig. 3. 3 sample CPU utilization graph of Machine2 during reasoning activities. *This figure illustrates 3 sample attitude of Machine2 during processing ontologies.*

during execution. This feature of Machine2 adds one new dimension for predicting. In the first execution, the faster core (2.1 Ghz) makes the processing and results in a shorter time than the average. In the second execution, the slower core (1.5 Ghz) processes the ontology and results in a longer time than the average. In the third execution, faster core executes the reasoning at a slower speed while the other core is used by other processes. This execution finishes in a longer time than the first execution, but in a shorter time than the second execution. This changeability of cores makes the processing longer or shorter. The OS may decide to use the faster or slower core according to its own decision parameters. And this decision will affect time/energy consumption. We accept this internal effect as a nature of this "device" and continue.

To see whether there is a trend in energy consumption of the battery in relation with the remaining battery level, we implemented moving average over the energy consumption with interval of 10 executions, which is illustrated in Figs. 1 and 2. In Machine1, we see that there is a trend of consuming less energy especially when the battery level is less than 50 %. We accept that there is slight trend (probability) when the battery level is low, and it may result in lower energy consumption with this device-reasoner pair. We searched for whether there is the same trend in Machine2 parallel to Machine1. Making the same experiment using the Machine2, we could not find a concrete trend parallel to Machine1. In our work, we assume that our power benchmark will measure the energy, regardless of the battery level, *within the error rate defined*. Seeing this difference in the results of two different machines, we conclude that *it is very difficult to make a generalizable model that can be applied to **all devices***. Thus, regarding "each device" as a black-box in analysing would be more practical.

Machine2 has given us an opportunity to see whether there is a "linear relation" between energy and time, as its energy and time consumption results are varied in time and energy dimension. We ordered execution results of Machine2 according to time consumption and divided into 5 groups as illustrated in Table 1.

According to Table 1, in the 1st group (66 executions with *lowest* time), average time consumption of group is 16.60 % less than average time consumption of all executions. But, energy consumption is not less than general, 1.24 % more than

Table 1. Energy-Time Consumption Relation of Machine2. *Distribution of energy and time consumption of reasoning same ontology from 100 % to 1 % battery level is illustrated. 1st column shows average (Avg.) time consumption of group. 3rd column shows Avg. energy consumption.*

Group	Avg. sec. of the group	Avg. sec. of all	Avg. Ws. of group	Avg. Ws. of all
1	36.397	43.644	94.951	93.785
2	42.124	43.644	90.737	93.785
3	45.446	43.644	94.043	93.785
4	47.035	43.644	93.754	93.785
5	47.273	43.644	95.258	93.785

general. In the 5th group (66 executions with *highest* time), average time consumption of group is 8.32 % more than average time. Whereas, average energy consumption of 5th group is 1.57 % more than the general. Observing the 1st and 5th group has a difference of 24.92 % average time consumption, but 0.33 % of difference in average energy consumption, we conclude that *we could not find a linear relation between time and energy consumption in this device-reasoner pair.*

Hardware doesn't influence the validity of the mechanism but shows varied results which makes us observe the effects of this hardware. For example, we reached the observation that energy consumption may not always be correlated with time consumption (as in Machine2) with help of this model while questioning why there was a higher variance in predictions of Machine2.

5.1 Experiments

Experiment results and source codes are accessible[8]. A re-run requires the preparation of an application development environment, the recompilation of the code and, finally, the generation of predictions in R. The reasoners TrOWL and HermiT are not part of our contribution, we therefore provide the scripts for running the experiments only. While working with TrOWL on Machine1, we observed that ontologies with the file size (in OWL Functional syntax) between 3000 KB and 3999 KB, 29 of 223 (13 %) could be processed within 100 s. Between 4000 KB and 4999 KB, it was about 2.99 % (5 of 167). Seeing this result, we limited our work for TrOWL within the dataset with the file size between 10 KB and 4999 KB (8281 ontologies). While working with HermiT on Machine1, we observed that ontologies with the file size between 500 KB and 599 KB, about 12.01 % (15 of 124) could be processed within 100 s. Seeing this result, we limited our work for HermiT within the dataset with the file size between 10 KB and 599 KB (6487 ontologies). We regard mobile devices as a black-box and do not search for the reasons of the peaks in energy consumption as in Figs. 1 and 2, whether it is because of OS services or manufacturer specific apps or anything we may not

---

[8] https://github.com/IsaGuclu/PredictionOfEnergy.

foresee, as this is the nature of mobile devices to run with this kind of internal (or external) influences. We preferred this overall approach as we are focusing on the energy consumption of the reasoning activity from a holistic perspective. We will not compare lower levels of reasoners but energy consumption in total. Reasoning experiment of one ontology is in this order:

1. The Counter starts calculating time and the energy. The Counter gets the average voltage and current from the OS, measuring the energy consumed in intervals of 100 ms.
2. The reasoner is called to load the ontology and the following query is sent:

```
prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#>
select * where {?X rdfs:subClassOf ?Y}
```

We describe experiments with a particular query answering task by sending the same SPARQL query to the two reasoners we investigate, in order to get results for subsumption reasoning.
3. When the request from the reasoner is provided with success and the query result is parsed, the Counter is stopped.

5.2   Results

After training the classification model with the data provided by the previous executions, we predicted whether a new ontology can be processed within the predefined time (100 s) or not, applying 10-fold cross validation. The results are illustrated in Table 2. This table illustrates successful and wrong predictions of the mechanism. "Positive" denotes reasoning CAN be made on mobile device. "Negative" denotes reasoning CANNOT be made on mobile device. For example, in a "Successful-Positive Prediction", it is predicted that reasoning can be accomplished on mobile device and it is observed so. In a "Wrong-Negative Prediction", it is predicted that reasoning cannot be accomplished on mobile device, but just opposite is observed, it could be processed on mobile device. As shown in Table 2, with TrOWL, the error rate of the 1st prediction in Machine1 is 0.52 % and in Machine2 is 1.51 %. With HermiT, the error rate of the 1st prediction in Machine1 is 0.76 % and in Machine2 is 1.86 %. Working on the ontologies which resulted in wrong predictions, deeper analysis can be made about the energy prediction, but we leave this analysis as a future work. After 1st prediction, we focussed on the prediction of the energy consumption. We trained our regression

Table 2. Classification model assessment.

	Machine1				Machine2			
	Successful prediction		Wrong prediction		Successful prediction		Wrong prediction	
	Positive	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*
HermiT	90.38 %	8.87 %	0.48 %	0.28 %	92.55 %	6.59 %	0.52 %	0.34 %
TrOWL	91.76 %	7.72 %	0.36 %	0.16 %	95.40 %	3.08 %	0.72 %	0.79 %

Table 3. Regression model assessment

	Machine1		Machine2	
	$R^2(\%)$	$RMSE$	$R^2(\%)$	$RMSE$
HermiT	94.05	5.43	89.85	10.35
TrOWL	98.59	4.62	95.64	10.86

Table 4. Percentage of error rates according to actual energy consumption

	Machine1		Machine2	
Group	$Hermit$	$TrOWL$	$Hermit$	$TrOWL$
1 ($Up\ to\ 1\ Ws.$)	20.21 %	18.88 %	43.50 %	50.20 %
2 (1–5 $Ws.$)	24.90 %	13.88 %	56.33 %	47.38 %
3 (5–10 $Ws.$)	25.51 %	16.35 %	58.88 %	33.93 %
4 (10–50 $Ws.$)	21.48 %	15.66 %	40.59 %	28.40 %
5 (50– $Ws.$)	14.80 %	5.43 %	21.11 %	14.06 %
$General$	21.46 %	14.11 %	45.12 %	40.00 %

model with the data provided by the previous executions and predicted how much energy will a new ontology consume, applying 10-fold cross validation. The results are illustrated in Table 3.

In Table 3, R^2 and $RMSE$ values as obtained from the prediction models are shown. Making more observations with different device-reasoner pairs will enhance the precision of the model, which we plan to do in future work.

To see the percentage of this error in prediction according to the amount of actual energy consumptions, we have grouped ontologies according to actual energy consumptions and obtained average percentage of error in prediction according to amount of actual energy consumption, as illustrated in Table 4.

From Table 4, we make the following observations. Machine1, which produces less varied energy consumption results, has less error rate in all groups of actual energy consumption. Whereas, Machine2, which produces more varied energy consumption results, has more error rate in all groups of actual energy consumption.

When the variation of energy consumption of the device-reasoner is lower, percentage of error rate is lower too. This encourages us to obtain more accurate execution data for training our model. Because, the more we can standardize our results for training, the more precise our prediction will be.

The random forests based regression model makes predictions in a very balanced way. As our energy measurement interval is 100 ms, we were expecting that there would be high percentages of error in predictions of 1st group of which the reasoning finishes less than a second. We find the difference of error rate between the 1st group and general acceptable.

These energy prediction results are obtained after predicting whether an ontology can be processed within 100 s with an accuracy of over 98 %. When energy consumption is very small, the energy prediction model can predict with an accuracy of nearly 80 % in Machine1 and nearly 50 % in Machine2. With the increase in reasoning time, the accuracy in Machine1 reaches about 90 % and the accuracy in Machine2 reaches about 80 %.

From all the experiments we have done, we are concluding that:

1. Treating the device (including OS, manufacturer specific apps and hardware specifications, etc.) as a black-box with the reasoner, we obtained affirmative results, indicating that the classification and regression models generated with this approach show a good measure for validity to describe the relation between energy consumption and structure of the ontology.
2. The classification models (which predict whether the ontology will be processed in the predefined time (100 s)) achieve very low error rates. It validates the feasibility and practicality of our approach, as it can be applied to minimize the risk of Out of memory (OOM) exceptions and general uncertainty about whether an ontology can be processed on a mobile device.
3. Using structure of the ontology (metrics) and previous ontology reasoning energy consumption data, actual energy consumption of a new ontology can be predicted with high accuracy. When the execution time of the ontology increases and standardized training data can be supplied, this accuracy reach 94.57 % as in Machine1 with TrOWL reasoner.
4. Patton and McGuinness had hypothesized that the amount of energy used for reasoning would be linearly related to the amount of time required to perform the reasoning, in their power consumption benchmark [18] for reasoners over mobile devices. Seeing experiment results with Machine2 about energy–time relation, we observe that energy consumption is not always parallel to the time consumption and this hypothesis is limited to old CPUs with standard speed. As the device contains many internal (OS policies-services-apps, manufacturer specific apps, hardware specifications, etc.) and external (movement of the user, bandwidth change, temperature, etc.) influences, instead of trying to sort out every variable and their weight in the energy consumption, using a holistic approach and collecting more and more data will be a more effective way for obtaining a more precise energy prediction mechanism. We conclude that the relation between time and energy is *changeable* according to hardware and software specifications of the device and this necessitates making separate prediction mechanisms for *time* and *energy* consumption.

6   Conclusion

Mobile devices, such as smartphones and tablets, have markedly different performance characteristics and requirements, most prominently limited energy, which poses a significant challenge for deploying computation-intensive tasks, such as ontology reasoning on mobile devices. In this paper, we developed statistical methods that predict energy consumption of ontology reasoning on various

mobile devices, using different reasoners and ontologies in the OWL 2 EL profile. Our main contributions include the following. Firstly, high prediction accuracy is achieved for our random forest-based regression models with R^2 of 90 % or higher. It is also observed that the prediction error rate is the lowest for ontologies with the highest actual energy consumption, showing that our prediction models are accurate *when it matters*. Our approach is hardware independent, i.e. hardware specification is not used as a parameter of our prediction model, thus our approach can be applied to devices other than the two that we tested. Secondly, we observe that a linear relation between time and energy consumption on a mobile device is not a valid assumption, especially with new hardware (CPU's containing cores with different speed) and software (multi-threading) improvements. Thirdly, the comprehensive dataset used in our evaluation has been made available to allow for reproducibility and encourage further investigation.

Our plan for future work is to improve our approach and make it applicable to real-world scenarios. First, we will extend our experiments with more devices and combine all models of different device-reasoner pairs into one comprehensive, general model. Second, we are planning to implement this approach in the Android version of TrOWL reasoner and empowering this prediction mechanism by collecting data from devices using this implementation. Third, we will build an optimisation mechanism that will manage the integration of mobile-cloud using this approach with user preferences taken into account.

References

1. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
2. Burton-Jones, A., Storey, V.C., Sugumaran, V., Ahluwalia, P.: A semiotic metrics suite for assessing the quality of ontologies. Data Knowl. Eng. **55**, 84–102 (2005)
3. Cerri, D., et al.: Towards knowledge in the cloud. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2008. LNCS, vol. 5333, pp. 986–995. Springer, Heidelberg (2008)
4. Corradi, A., Destro, M., Foschini, L., Kotoulas, S., Lopez, V., Montanari, R.: Mobile cloud support for semantic-enriched speech recognition in social care. IEEE Trans. Cloud Comput. **PP**(99), 1 (2016). doi:10.1109/TCC.2016.2570757
5. Glimm, B., Horrocks, I., Motik, B., Stoilos, G., Wang, Z.: HermiT: an OWL 2 reasoner. J. Autom. Reasoning **53**, 245–269 (2014)
6. Groth, P.: The rise of the verb. In: SW2022 (2012)
7. Gurun, S., Krintz, C., Wolski, R.: NWSLite : a light-weight prediction utility for mobile devices. In: MOBISYS (2004)
8. Hasan, R., Gandon, F.L.: A machine learning approach to SPARQL query performance prediction. In: WEBI (2014)
9. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2009)

10. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S.: OWL 2 web ontology language primer. W3C Recommendation **27**(1), 123 (2009)
11. Hogan, A., Pan, J.Z., Polleres, A., Ren, Y.: Scalable OWL 2 reasoning for linked data. In: Polleres, A., d'Amato, C., Arenas, M., Handschuh, S., Kroner, P., Ossowski, S., Patel-Schneider, P. (eds.) Reasoning Web 2011. LNCS, vol. 6848, pp. 250–325. Springer, Heidelberg (2011)
12. Kang, Y.-B., Li, Y.-F., Krishnaswamy, S.: Predicting reasoning performance using ontology metrics. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012, Part I. LNCS, vol. 7649, pp. 198–214. Springer, Heidelberg (2012)
13. Kang, Y.-B., Pan, J.Z., Krishnaswamy, S., Sawangphol, W., Li, Y.-F.: How long will it take? Accurate prediction of ontology reasoning performance. In: AAAI (2014)
14. Kleemann, T.: Towards mobile reasoning. In: DLOG (2006)
15. Li, Y.-F., Pan, J.Z., Krishnaswamy, S., Hauswirth, M., Nguyen, H.H.: The ubiquitous semantic web: promises, progress and challenges. Int. J. Semant. Web Inf. Syst. **10**, 1–16 (2014)
16. Maynard, D., Peters, W., Li, Y.: Metrics for evaluation of ontology-based information extraction. In: WWW Workshop on Evaluation of Ontologies for the Web (2006)
17. Motik, B., Cuenca Grau, B., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C., Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P., et al.: OWL 2 web ontology language: profiles. In: W3C Recommendation, 27 October 2009 (2012)
18. Patton, E.W., McGuinness, D.L.: A power consumption benchmark for reasoners on mobile devices. In: Mika, P., et al. (eds.) ISWC 2014, Part I. LNCS, vol. 8796, pp. 409–424. Springer, Heidelberg (2014)
19. Piccolo, L.S.G., Baranauskas, M.C.C., Fernández, M., Alani, H., Liddo, A.D.: Energy consumption awareness in the workplace,: technical artefacts and practices. In: IHC (2014)
20. Rietveld, L., Schlobach, S.: Semantic web in a constrained environment. In: Proceedings of the Downscaling the Semantic Web Workshop (ESWC) (2012)
21. Valincius, E., Nguyen, H.H., Pan, J.Z.: A power consumption benchmark framework for ontology reasoning on android devices. In: ORE (2015)
22. Wen, Y., Wolski, R., Krintz, C.: Online prediction of battery lifetime for embedded and mobile devices. In: Falsafi, B., VijayKumar, T.N. (eds.) PACS 2003. LNCS, vol. 3164, pp. 57–72. Springer, Heidelberg (2005)
23. Yao, H., Orme, A.M., Etzkorn, L.: Cohesion metrics for ontology design and application. J. Comput. Sci. **1**, 107–113 (2005)
24. Yus, R., Bobed, C., Esteban, G., Bobillo, F., Mena, E.: Android goes semantic: DL reasoners on smartphones. In: ORE (2013)
25. Yus, R., Pappachan, P.: Are apps going semantic? A systematic review of semantic mobile applications. In: SEMWEB (2015)
26. Zhang, H., Li, Y.-F., Tan, H.B.K.: Measuring design complexity of semantic web ontologies. J. Syst. Softw. **83**, 803–814 (2010)