EasyChair Preprint

№ 1833

A literature survey: semantic technology approach in machine learning

L Rachana and S Shridevi

November 5, 2019

A literature survey: semantic technology approach in machine learning

Rachana L, Shridevi S
SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
Vellore Institute of technology, Chennai
rachana.l2019@vitstudent.ac.in, shridevi.s@vit.ac.in

Abstract. Semantic technology approach in machine learning is an emerging technique to solve the problems in the machine learning. Semantic technology has been the improvised from decades according to the human needs and industrial demands. This new era is all about teaching a machine to learn on its own and to make it understand the concept and the purpose for what it is used, using algorithms. This paper, condenses the work of semantic technology approach in machine learning and its idea put forward. The introduction details with brief explanation followed by description of the semantic technology and machine learning, important role. The literature survey contains summarized view of the papers and concluded with review analysis.

Keywords: semantic technology, machine learning, ontology.

1. Introduction

The approach of the semantic technology in machine learning is just like icing on a cake, the technology enables to provide them with an interface that allows them to directly model their knowledge in the system, and to review the knowledge in the system. Semantic systems also excel at provability. In a well-structured semantic system, you can get formal logical proofs, expressed in fairly understandable terms, backing up the answers produced by the system. You can have a very high degree of confidence in the answers which the system produces. Of course, sometimes there are flaws or holes in the underlying ontology but when there are, they're generally easier to identify. The inclusion of semantic technology has also given great opportunity to the machine learning to interact with human, has for machine learning is quiet complex for human to understand at one go. The datasets are categorized and processed as well as the relationship of the data within the sets is found. In this paper it deals the method used to solve the problems which evolved during the comparison of data for more accurate values.

1.1 Semantic technology

The term "semantic" refers to the sense of words. Semantic technology uses Artificial intelligence to simulate the understanding and processing of language information by people. Semantic software reads and tries to grasp language and words in its context with respect to the approach. Technically speaking, this method is based on various levels of research: analysis of morphology and grammar. In Cogito's case, the semantics design requirements include various integrated elements. The following are the most important elements

I. Parser performs morphological, grammatical and syntactic study of the expression

II. Lexicon understands terms and all their meanings memory keeps track of analysis outcomes

III. memory records empirical outcomes information represents real-world knowledge

IV. Content representation is language content in the context of a theoretical and mental graph.

The lexicon includes the so-called "semantic network." We name our collection of semantic networks "Sensigrafo" at Expert System. It is not an ordinary dictionary but tools that have been designed for programmatic use, where word types are knots connected to each other by multiple links denoting semantic or lexical relationships. As human beings, it's easy to understand our everyday language and words ' meanings. It's not that easy to pass these same skills to a computer. It takes time to learn, and the same applies to a computer. There are no shortcuts or magic formulas: learning a language is challenging, and it takes time and work for automated processes. [18]

1.2 Machine learning

Machine learning is an artificial intelligence application that provides machines with learning skills and automatically learns from experience without direct programming. Machine learning focuses on building software tool for processing and using data for their own purposes. The learning process starts with observations, evidence including examples, direct experience, and preparation based on our examples, to interpret knowledge patterns and make better future decisions. The main goal is to make intelligent machines without human intervention to understand and change the behavior.

Some other methods in machine learning are often referred to as supervised or non-supervised algorithms for machine learning.

I. Techniques of supervised machine learning can use defined occurrences to predict future events to use what has already been observed from new data. Beginning with the evaluation of the existing learning data collection, the learner algorithm produces a system that is supposed to predict the performance values. The device will set a goal for any new input after proper training. The analysis of algorithm can also adapt its output to the correct output and detect errors so that the design can be improved accordingly.

II. In contrast, when the software used for training is not labelled or numbered, unguarded algorithms are used for machine learning. Unsupervised training investigates how constructs in order to explain a secret framework could derive a feature from unlabeled data. The computer does not find the correct performance, but analyses the information and can extract details from data sets to clarify the secret properties of unlabeled software.

III. Half-supervised machine learning algorithms fall between supervised and non-supervised learning because they use marked and unlabeled training information, usually with a small proportion of labelled and a large proportion of unlabeled information. The programs that use this approach can greatly enhance reading accuracy. Semi-supervised training is usually handled when it is needed to educate and benefit from the information gained when the programs are educated or relevant. Typically, however, additional resources are not required to collect unlabeled information.

IV. Enhancing machine learning algorithms is a form of learning that interacts with its environment and finds errors and rewards. Evaluation and error detection and delayed compensation are the most important features for optimizing training. This approach allows devices and computer agents to immediately examine the optimal actions to improve their efficiency in a particular context. Clear feedback is needed to let the agent know what behavior is needed, known as the reinforcement signal.

Machine learning involves the storage of massive data. Using the computer. Although it usually delivers faster and more accurate results to detect favorable opportunities and harmful threats, additional time and resources may also be required to train properly. The combination of machine learning with AI and cognitive technology can make it even more productive to produce large amounts of information.[18]

1.3 Role of semantic technology in machine learning

The Semantic Web offers structured templates for both information and ontological context awareness representation. Semantic Web concepts are used to define Meta data, but as a general framework for interaction and information management, they also have great potential. Machine learning will play an extremely important role in a wide range of possible applications: Machine learning alternatives have been evolved to support ontology management, semi-automatic analysis unstructured data, and web mining integration of semantic data. Increasingly, machine learning will be used to evaluate distributed data sources represented in semantic web formats and to help theoretical system reasoning and querying.

2. Literature Survey

The survey is all about to understand the approach of semantic technology in machine learning and realize the state of art of the works in this field for finding the research gap, so that it can motivate for better research problems.

Oge Marques et al [1],"Semi-automatic Semantic Annotation of Images Using Machine Learning Techniques”, in order to semi automatically annotate objects using machine learning methods, we suggested a three-layer structure. The research expands the previous work into a Semantic Web-oriented approach in the related field of content-based picture recovery. The proposed architecture allows for a good degree of independence between efforts to improve its supporting technologies and algorithms, such as: development of better algorithms for visual feature extraction. Implementation of better clustering algorithms. Improved communication methods inquiry. Deployment of intelligent ontology agents. Enlargement, preservation and binding of ontologies with keywords. It is also recognized that the quality of the algorithms – especially those used to obtain visual characteristics, to collect, to detect similarity and to map the bottom-middle layer of our model is an important factor in the success of the methodology proposed.

Matthew E. Taylor et al[2],"Autonomous Classification of Knowledge into an Ontology ",The problem of how to automatically determine where new knowledge can be placed in existing ontologies is seen, though distinct, in this work. Alternatively, it is increasingly important for machine learning technologies to simplify a job rather than relying on human knowledge engineers to carefully identify information. The rates of ontology building through automatic knowledge acquisition techniques are increasing. This paper compares three well-established machine learning approaches and illustrates that they can be used effectively for this function. In the Cyc knowledge base program our processes are fully implemented and evaluated.

Khin Phyu Phyu Shein [3],"Ontology based combined approach for Sentiment Classification ", In order to enhance the sentiment classification we suggested the combination of POS tagging, FCA-based domain ontology, and SVM classification. By using this approach, we can take a closer look at the strength or weakness of the products or artifacts and we hope that it will be useful for further creation and enhancement of the product or object growth. We need to experiment with a large amount of data sets as the future work and need more learning and classification to solve the comparative sentence problem.

Khin Phyu Phyu Shein et al [4],"Sentiment Classification based on Ontology and SVM Classifier ", the suggested hybrid method of POS labeling, FCA-based domain ontology and SVM classifier to boost the identification of feelings.
By using this method, we may take a closer look at the strength or weakness of the goods or artifacts and we expect that it will be helpful for further creation and enhancement of the product or artifact growth. This approach is still being further developed.

Azadeh Haratian Nezhadi et al[5],”ontology alignment using machine learning techniques” ,This paper proposes a method of effective ontological consistency based on the combination of various categories of similarity in one input sample. The proposed model specifies the coordination mechanism without having to have ontology instances

beforehand, making it easier to coordinate. Our proposed exemplary does not require user intervention and it has a consistent performance, both for aligned and non-aligning entities, through a detailed operational enchantment process. AdaBoost (DT) model provides the best overall accuracy, particularly when using feature selection scheme. Experimental results indicate that up to 99 percent of F-measurement parameters were higher than other relevant research trials that used up to 23 similarity tests.
Although this research uses only eight similarities measurements in its optimal model, it has compensated for the potential impacts of the decrease of feature through the use of ontology observations, the increase in sample diversity and the choice of more effective similarity measures. This provides greater precision and lower computing costs, making a model suitable even for a task to align ontology online.

Agnese Pinto et al [6],"A semantic-based approach for Machine Learning data analysis", the wide range of technology and products are primarily centered on the intelligent presentation of data obtained in the atmosphere by heterogeneous sensors. Traditional Machine Learning (ML) approaches often do not go beyond a simple description without a clear description of the events observed. This paper delivers an early draft of a semantic-enhanced machine learning study on sensor flow information, doing much better on ubiquitous smart objects that are resource-constrained. The architecture blends ontology-driven features of statistical data delivery with non-standard matching resources. Encourages the identification of fine-grained events by addressing the standard asset discovery classification issue of ML.

M. Syamala Devi et al [7],"machine learning techniques with ontology for subjective answer evaluation ", in this study, the methods examined and applied show a high association with human Performance. This is because the duration of the response is largely influenced by human interpretation. Presence of keywords and keyword meaning. Use of Ontology, searches for keywords as well as keywords that occur in the correct Context. This dimension is absent in ontology-free techniques. Use Ontology tests for term existence, synonyms, correct word meaning, and scope of all definitions. It is assumed that, owing to systematic analysis, utilizing ML strategies with Ontology provides satisfactory results.

Aurawan Imsombut et al [8],"An Alternative Technique for Populating Thai Tourism Ontology from Texts Based on Machine Learning ",This study proposed the use of Conditional Random Fields (CRFs) to identify examples of concepts with dictionary knowledge and a cue word list. In contrast, the post processing of instance abstraction is used as the method to boost the effects of NE boundary recognition, and the lexico-syntactic template is used to classify the relationships between instances. The experiments were carried out in the tourism domain on Thai language web documents.
The method can derive instances with appropriate results according to these preliminary results.

Jingyuan Xu et al [9],"An OWL ontology representation for machine-learned functions using linked data ", Machine-learned tasks in the world of big information are incredibly useful. It is difficult to create new learned functions, but linguistic awareness can be used to allow anyone to replicate current functions effectively. We demonstrated the use of OWL ontology to canonize variable names and function properties, such as characteristics and performance metrics. Using hierarchical relationships between variables, in the semantic web, we may find missing function values based on other data provided by the client. We also have a database interface that allows users to use these generic objects to submit information.

Marie Kim, et al [10],"Augmented Ontology by Handshaking with Machine Learning ", we are exploring two methods and finding a good solution that maximizes and mitigates the advantages of both strategies. This study indicates a new idea of convergence to make up for each innovation with the other: that is semantic filtering. This paper involves a semantic simulation of toys and an adaptation of a machine learning functionality to understand the suggested term, semantic sorting. Recognition of the situation is AI's secret and the essence of AI is a data analysis. This paper suggests smooth and secure implementation in the IoT world for machine learning and semantic software. Overall framework needs to be refined as a further work, and real integrated implementation needs to be followed.

Fabiano A. Dorc et al [11],"An Approach for Automatic and Dynamic Analysis of Learning Objects Repositories through Ontologies and Data Mining Techniques for Supporting Personalized Recommendation of Content in Adaptive and Intelligent Educational Systems", Adaptive and Intelligent Educational Systems are valuable resources to support teaching-learning activities. These systems utilize innovative strategies to tailor educational content to students ' real needs. With the growing abundance of educational content, there is a good reason to believe that smart data analytics and machine learning technologies will become essential ingredients of educational advancement. Hence, this research suggests a method to systematic and complex study of databases of learning artifacts in which ontology models the relationship between properties of learning objects and styles of learning. Better outcomes are collected, and in this study they were discussed.

Ms.K.Saranya et al [12],"Onto-based sentiment classification was using Machine Learning Techniques ", In this article, for better outcomes, the use of semantics and ontology to define text is coupled with machine learning techniques. The experiment shows that SVM, NB and kNN classifiers in the existing literature have been shown to be more useful for emotional research. The process began by defining and removing the affective class hierarchy in WordNet. The next move was to attribute emotions to the linguistic positions of the derived emo-words to establish a hierarchy of emotions. With more accurate results, the machine learning algorithm or a mixture of the algorithms (hybrid algorithm) can be implemented. Semi structured text grouping and clustering have some dispute and new opportunities.

Fuchao Liu et al [13],"The Extension of Domain Ontology Based on Text Clustering ",This work discusses the relevant principles and approaches of constructing ontology and expanding ontology, introducing an automated ontology extension approach focused on supervised training and text clustering. This method uses the K-means clustering algorithm to isolate the information of the domain and to direct the construction of the Naive Bayes classification training collection. Terms are applied to the goal ontology in the nominee array, while noise terms are introduced to the stop-word dictionary at the same time. This method's input system was designed to promote the structure of the consistency of ontology, and eventually it will semi-automatically expand ontology. Ontology editors and other development tools are also necessary throughout the infrastructure construction of automatic cycle of ontology. Future work will also include attempts to improve the reliability of proposed methods and implement the Fuzzy Inference Rules training process.

Kyoung Soon Hwang et al [14],"Autonomous Machine Learning Modeling using a Task Ontology ", we extracted essential keywords from articles and textbooks on machine learning in this paper to establish ontology. In addition, we have designed a MEX-based ontology task. We have also researched the automated machine learning system workflow. For automated workflows at a specified autonomous stage, the following protocol applies. Therefore, non-experts are able to perform complex tasks by the system and can quickly apply the machine learning framework in a particular application.

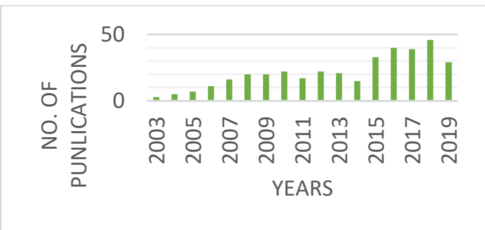
Michele Rutaa et al[15],"Machine Learning in the Internet of Things: a Semantic-enhanced Approach", In this paper we presented a new approach for semantic enhanced machine training in the Internet of Things on heterogeneous data streams. Carrying raw data on ontology-based labels of concept provides a low-level linguistic understanding of the numerical distribution of information, whereas the conjunctive aggregation of concept components enables automated representation of events during model training phases to be rich and substantive. Finally, the use of non-standard match-making inferences allows the discovery of finest occurrences by treating the topic of ML identification as an asset seek.

Bo Fu et al [16],"Using Behavior Data to Predict User Success in Ontology Class Mapping – An Application of Machine Learning in Interaction Analysis", Representation in ontology usually consists of one-size-all approaches that offer the same representation for a given task / system to each client. Not every visualization technique, however, can result in a successful task outcome for each user, and thus it may be beneficial to develop systems that can provide adaptive assistance during visualization use. We must first determine if a client is likely to succeed in a job to create these scalable systems. For this reason this article provides a possible solution for the avoidance of user failure by exploring how effectively, in terms of accuracy, completeness and overall performance, can be accomplished through a user-interaction information (including eye tracking and click-data). We execute three classification tests to forecast success based on a data set obtained in a client sample utilizing two separate ontology modeling methods and two different types of activities. Results showed that for these predictions communication information could potentially be used and all three classifiers statistically outperform simple classifiers substantially.

Dominik Sacha et al [17],"VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning”, we suggest an ontology (VIS4ML) in this paper for a VA sub-area, called "VA-assisted ML." VIS4ML is intended to identify and explain current VA workflows used in ML as well as to recognize deficiencies in ML processes and the ability to apply new VA techniques to such processes. Ontologies are commonly used in biology, medicine, and many other disciplines to map the reach of a subject. We follow the empirical methodologies for VIS4ML development, including ontology definition, conceptualization, formalization, deployment, and validation. In general, to include model-development workflows, they reinterpret the conventional VA pipeline.
In order to formulate VIS4ML, we add the requisite concepts, laws, syntaxes and graphical notes and use semantic web technology to incorporate it in the Web Ontology Language (OWL). VIS4ML incorporates the information from past workflows at a high level where VA is used to aid in ML. It is aligned with the existing VA principles and will continue to evolve in tandem with potential VA and ML innovations.
While this ontology is an effort to build the theoretical basis for VA, clinicians can use it in real-world applications to refine model-development workflows by routinely analyzing the potential benefits which computer and person ability may bring about.

2.1. Review analysis

After the analysis of the above the we can say that every experiment has its own drawbacks the such that, algorithms requires the storage of massive data sets, which should be diverse / impartial and of good quality. There may also be occasions when they have to wait for the generation of new information. The next key challenge is the ability to analyze algorithm-generated findings correctly. You also need to pick the algorithms carefully for your intent. Mistakes may set up a chain of errors which can go undetected for prolonged periods of time. And it takes quite a while to identify and even lengthy to correct the source of the problem when it becomes noticed. The below fig 1 allows us to track the number of publications during the last decade.



3. Conclusion

The above survey gives an insight to recommend ontological data streams for better semantic machine learning. Tracking down raw data to ontological framework labels, gives a low-level semantic understanding of numerical classified information, while the combination of concept components enables a robust and solid description of events to be constructed effectively during the testing period. Ultimately, the use of non-standard corresponding inferences allows fine-grained incidents to be detected by approaching the ML classification issue as an object discovery

4. References

1. Oge Marques and Nitish Barman [1],"Semi-automatic Semantic Annotation of Images Using Machine Learning Techniques", ISWC 2003

2. Matthew E. Taylor, Cynthia Matuszek, Bryan Klimt, and Michael Witbrock [2],"Autonomous Classification of Knowledge into an Ontology ", American Association for Artificial Intelligence 2007

3. Khin Phyu Phyu Shein [3],"Ontology based combined approach for Sentiment Classification ", international conference on communications and information technology 2009

4. Khin Phyu Phyu Shein, Thi Thi Soe Nyunt [4],"Sentiment Classification based on Ontology and SVM Classifier ", International Conference on Communication Software and Networks 2010

5. Azadeh Haratian Nezhadi, Bita Shadgar and Alireza Osareh[5]",ontology alignment using machine learning techniques", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, April 2011

6.Agnese Pinto, Floriano Scioscia, Giuseppe Loseto, Michele Ruta, Eliana Bove, Eugenio Di Sciascio Politecnico di Bari[6]," A semantic-based approach for Machine Learning data analysis", International Conference on Semantic Computing (IEEE ICSC 2015)

7. M. Syamala Devi and Himani Mittal [7],"machine learning techniques with ontology for subjective answer evaluation ", International Journal on Natural Language Computing (IJNLC) Vol. 5, No.2, April 2016

8. Aurawan Imsombut, Chaloemphon Sirikayon [8],"An Alternative Technique for Populating Thai Tourism Ontology from Texts Based on Machine Learning ", ICIS 2016

9. Jingyuan Xu , Hao Wang, Henry Trimbach[9],"An OWL ontology representation for machine-learned functions using linked data ",International Congress on Big Data 2016

10. Marie Kim, Hyunjoong Kang, Soonhyun Kwon, YongJoon Lee, Kwihoon Kim, Cheol Sik Pyo[10],"Augmented Ontology by Handshaking with Machine Learning " ICACT 2017

11. Fabiano A. Dorc, Vitor C. Carvalho, Miller M. Mendes, Rafael D. Ara´ujo, Hiran N. Ferreira, Renan G. Catellan[11],"An Approach for Automatic and Dynamic Analysis of Learning Objects Repositories Through Ontologies and Data Mining Techniques for Supporting Personalized Recommendation of Content in Adaptive and Intelligent Educational Systems", International Conference on Advanced Learning Technologies 2017

12. Ms.K.Saranya, Dr.S.Jayanthy[12],"Onto-based sentiment classification using Machine Learning Techniques ", International Conference on Innovations in information Embedded and Communication Systems (ICIIECS) 2017

13. Fuchao Liu,Guanyu Li[13],"The Extension of Domain Ontology Based on Text Clustering ",International Conference on Intelligent Human-Machine Systems and Cybernetics 2018

14. Kyoung Soon Hwang,Ki Sun Park ,Sang Hyun Lee,Kwang il Kim,Keon Myung Lee [14],"Autonomous Machine Learning Modeling using a Task Ontology ",Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems 2018

15. Michele Rutaa, Floriano Sciosciaa, Giuseppe Losetoa [15], Agnese Pintoa, and Eugenio Di Sciascioa,"Machine Learning in the Internet of Things: a Semantic-enhanced Approach", IOS Press and the authors 2018

16. Bo Fu, Ben Steichen [16],"Using Behavior Data to Predict User Success in Ontology Class Mapping – An Application of Machine Learning in Interaction Analysis", International Conference on Semantic Computing (ICSC) 2019

17. Dominik Sacha, Matthias Kraus, Daniel A. Keim Member, IEEE, and Min Chen Member, IEEE[17],"VIS4ML: An Ontology for Visual Analytics Assisted Machine

Learning",ieee transactions on visualization and computer graphics, vol. 25, no. 1, January 2019

18. Volker Tresp, Markus Bundschus, Achim Rettinger, Yi Huang[18]," Towards Machine Learning on the Semantic Web", Springer-Verlag Berlin Heidelberg 2008