

# A Survey Paper on Ontology-Based Approaches for Semantic Data Mining

Prof. Priti V. Bhagat  
Asst. Prof., Department of CSE  
DMIETR, Sawangi(M), Wardha  
email:bhagat.preetee@gmail.com

Prof. Palash M. Gourshettiwar  
Asst. Prof., Department of CSE  
DMIETR, Sawangi(M), Wardha  
email:palash9477@gmail.com

**Abstract-** Semantic Data Mining alludes to the information mining assignments that deliberately consolidate area learning, particularly formal semantics, into the procedure. Numerous exploration endeavors have validated the advantages of fusing area learning in information mining and in the meantime, the expansion of information building has enhanced the group of space learning, particularly formal semantics and Semantic Web ontology. Ontology is an explicit specification of conceptualization and a formal approach to characterize the semantics of information and data. The formal structure of ontology makes it a nature approach to encode area information for the information mining utilization. Here in Semantic information mining ontology can possibly help semantic information mining and how formal semantics in ontologies can be joined into the data mining procedure.

**Keywords** – information mining; ontology; semantic data mining

\*\*\*\*\*

## I. INTRODUCTION

Data mining, otherwise called knowledge discovery from database (KDD), is the methodology of nontrivial extraction of verifiable, beforehand obscure, and possibly valuable data from information [1]. With the recent advances in data mining strategies lead to numerous momentous upsets in information investigation and big data. Data mining likewise joins methods from insights, computerized reasoning, machine learning, database framework, and numerous different controls to examine substantial information sets. Semantic Data Mining alludes to data mining errands that deliberately fuse space learning, particularly formal semantics, into the methodology. Past semantic data mining exploration has bore witness to the positive impact of space learning on data mining. Amid the seeking and pattern generating procedure, area learning can function as an arrangement of former information of requirements to help decrease hunt space and aide the inquiry way [2], [3].

To make utilization of area information in the data mining process, the first step must record for speaking to and building the learning by models that the PC can further get to and process. The multiplication of knowledge engineering (KE) has astoundingly improved the group of space learning with strategies that fabricate and utilization area information in a formal manner [4]. Ontology is one of effective knowledge engineering advances, which is the unequivocal determination of a conceptualization [5], [6]. Ontology is created to determine a specific area. Such a ontology, regularly known as space ontology, formally determines the ideas and connections in that area. The encoded formal semantics in ontologies is essentially utilized for viably imparting and reusing of

learning and information. Research in the region of the Semantic Web [7] has prompted truly develop benchmarks for demonstrating and arranging area information. At present, Semantic Web ontologies turn into a key innovation for wise learning handling, giving a system to offering calculated models about an area. The Web Ontology Language (OWL) [8], which has developed as the true standard for characterizing Semantic Web ontologies, is broadly utilized for this reason. The Semantic Web advancements that speak to space learning including organized gathering of former data, derivation rules, information advanced datasets and so on, could hence create systems for precise joining of area information in a savvy data mining environment.

Essentially the study is on three points of view of ontology based methodologies in the exploration of semantic data mining [9]:

- Role of ontologies: Why area information with formal semantics, for example, ontologies, are important in all phases of the data mining methodology.
- Mining with ontologies: How ontologies are spoken to and prepared to help the data mining methodology.
- Performance assessment: How ontologies can enhance the execution of data mining frameworks in applications.

## II. ROLE OF ONTOLOGIES IN SEMANTIC DATA MINING

The viewpoint and component of using ontologies in semantic data mining fluctuates crosswise over diverse frameworks and applications. The accompanying are three reasons that ontologies have been acquainted with semantic data mining:

- To scaffold the semantic hole between the information, applications, data mining calculations, and data mining results.
- To furnish data mining calculations with from the earlier learning this controls the mining process or lessens/ compels the inquiry space.
- To give a formal approach to speaking to the data mining stream, from information preprocessing to mining results.

#### A. *Bridging the semantic crevice*

Specialists guarantee that there exists a learning hole between the information, data mining calculation, and mining results in all phases of data mining including preprocessing, calculation execution, and result era. There exists semantic crevice between the data mining calculation and information also. Data mining calculations are typically intended for information gathered from distinctive areas and situations. Along these lines, information from a particular area more often than not convey space particular semantics. The bland data mining calculations do not have the capacity to distinguish and make utilization of semantics crosswise over distinctive areas and applications [10],[11],[12]. Ontologies are helpful to determine space semantics and can decrease the semantic hole by commenting the information with rich semantics. Semantic annotation goes for allotting the essential component of data connections to formal semantic portrayals. Such components ought to constitute the semantics of their source. Semantic annotation is significant in acknowledging semantic data mining by conveying formal semantics to information [13], [14]. The explained information are exceptionally helpful for the later ventures of semantic data mining in light of the fact that the information are elevated to the formal and organized organization that unites ontological terms and relations. The information/content mining results are sets of organized data and learning with in regards to the area. To speak to the organized and machine-coherent data, it is nature to speak to the data with ontology. Ontology Based Information Extraction (OBIE) [9] has broadly utilized this representation. With OBIE, the data removed is all around organized as well as spoken to by predicates in the ontology which are simple for offering and reuse.

#### B. *Providing former learning and requirements*

The definition and reuse of former learning is a standout amongst the most vital issues for semantic data mining. Ontology is a nature approach to encode the formal semantics of former learning. The encoded earlier learning can possibly guide and impact all phases of the data mining methodology, from preprocessing to result sifting and representation. Ontologies are fused into the chart representation of the information as the priori learning to predisposition the diagram structure furthermore speaking to the separations in the middle

of terms and ideas in the diagram. The methodology changes the hypergraph and weighted hyperedges into a bipartite chart to speak to both the information and ontology in a uniformed structure. Arbitrary strolls with restart over the bipartite diagram are performed to create semantic affiliations. At whatever point the arbitrary walk experiences the ontology based edges, the area learning encoded in ontologies connects the inert semantic relations underneath the information with rich semantics [15].

#### C. *Formally speaking to data mining results*

The very much composed data mining frameworks ought to present results and found examples in a formal and organized arrangement, with the goal that data mining results are fit to be deciphered as area learning and to further advance and enhance current information bases. Ontology is one of the best approach to speak to the data mining results in a formal and organized way. ontology can encode rich semantics for distinctive areas. The data mining results from distinctive areas and undertakings acclimate commonly with the representation of ontology, for instance, data extraction and affiliation standard mining. In ontology based information extraction (OBIE) [9] the separated data is a situated of clarified terms from the record with the relations characterized in the ontology. It is hence straight forward to speak to the separated data with ontology.

### III. MINING WITH ONTOLOGIES

With formally encoded semantics, ontology can possibly support in different data mining errands. semantic data mining calculations planned in a few imperative errands, including affiliation principle mining, order, grouping, proposal, data extraction, and connection expectation.

#### A. *Ontology-based Association Rule Mining*

Affiliation tenet mining is a central data mining undertaking and very much utilized as a part of diverse applications. Ontology in this work gives the limitations to inquiries in the affiliation mining methodology. The hunt space of affiliation mining is compelled by the question came back from the ontology that a few things from the yield affiliation tenets are avoided or to be utilized to portray fascinating things as indicated by a deliberation level. The client limitations incorporate both pruning imperatives, which are utilized for sifting an arrangement of non-intriguing things, and reflection requirements, which allow a speculation of a thing to an idea of the ontology [3].

#### B. *Ontology-based Classification*

Arrangement is a standout amongst the most well-known data mining undertakings that discovering a model (or capacity) to portray and recognize information classes or ideas [16]. In semantic data mining, one normal utilization of ontology is to expound the arrangement names with the arrangement of relations characterized in the ontology. Research by Balcan et al. [2] shows that with the ontology clarified order names, the semantics encoded in the characterization errand has the potential not just to impact the named information in the arrangement assignment additionally to handle extensive number of unlabeled information. They fused ontology as consistency requirements into various related order errands. These errands group numerous classes in parallel. Ontology indicates the limitations between the different order undertakings. An unlabeled lapse rate is characterized as the likelihood the classifier relegates a name for the unlabeled information that disregards the ontology. This arrangement assignment delivers the order theory with the classifiers that create the slightest unlabeled slip rate and accordingly most grouping consistency.

#### C. *Ontology-based Clustering*

Grouping is an data mining errand that gathering an arrangement of items in the same group which are like one another [17]. Early work of ontology based bunching incorporates utilizing ontology as a part of the content grouping errand for the information preprocessing, enhancing term vectors with ontological ideas, and advancing separation measure with ontology semantics [18], [19], [20].

#### D. *Ontology based Information Extraction*

Information extraction (IE) alludes to the undertaking of recovering certain sorts of data from characteristic dialect message by preparing them naturally. IE is nearly identified with content mining. Ontology based information extraction (OBIE) is a subfield of data extraction, which utilizes formal ontologies to guide the extraction process [21], [9]. Due to this direction in the extraction process, OBIE frameworks have generally actualized after an administered methodology. Albeit not very many semi-regulated IE frameworks are considered as ontology construct they depend with respect to occasions of known connections. Accordingly those semi-directed frameworks can likewise be considered as OBIE frameworks. Early work of OBIE incorporates learning extraction from web archives [22] and information rich unstructured records. Ontology can give consistency checking to the separated data in the IE framework. Kara [23] displayed a ontology based data extraction and recovery framework which utilizes ontology for consistency checking. The yield of a customary

IE framework is changed to ontological examples through ontology populace. The derivation and consistency checking are performed on these ontological occurrences.

#### E. *Ontology based Recommendation System*

Recommender frameworks or proposal frameworks [24], [25] are the frameworks that devote to anticipate the inclination or evaluations that a client would provide for a thing. Suggestion frameworks have ended up greatly famous lately and been connected in an assortment of utilizations including films, music, news, books, examination articles, seek questions, and social labels [26], [27]. In a decent proposal framework, heterogeneous data from different sources is typically needed. Ontology can incorporate the utilization of heterogeneous data and aide the proposal inclination.

#### F. *Ontology based Link Prediction*

Join expectation for informal organizations turns into an exceptionally dynamic examination zone in data mining because of the achievement of online informal organizations, for example, Twitter, Facebook, and Google+. Aljandal et al. [28] introduced a connection forecast structure with ontology improved numerical diagram highlights. The creators asserted that in past interpersonal organization inquire about level representation of interest scientific classifications constrained the change of connection forecast. Ontology accumulated separation measure is proposed to encode the interest scientific classifications in ontology into the separation measure to all the more precisely depict the imparted client intrigues.

The information are initially expounded by controlled vocabulary terms from ontologies. The annotation connections between the information and predicates in ontology fonn an annotation chart. Chart outline and thick subgraph strategy were proposed to channel the diagram and discover promising subgraphs. A scoring capacity in view of various heuristics was proposed to rank the expectations in light of these separated subgraphs [29]. Amakrishnan [30] proposed a strategy to find the educational association subgraphs that relate two substances in the chart. They proposed heuristics for edge weighting that depend by implication on the semantics of substance and property sorts in the ontology and on qualities of the occurrence information. The showcase p-diagram era calculation was proposed to concentrate a little association subgraph from the data chart.

### IV. EXECUTION EVALUATION AND APPLICATIONS

As a formal determination of area ideas and connections, ontology can aid in the data mining process in different viewpoints. It is sensible to expect an execution pick

up in ontology based methodologies contrasted and the data mining methodologies without utilizing ontologies or other type of area learning. Numerous semantic data mining exploration endeavors have validated such enhancements.

#### A. Execution pick up in accuracy, review, and consistency of information mining results

Research in proposal framework recommends that ontology based suggestion frameworks has preferable expectation accuracy over conventional proposal systems [31], [32]. With the enhanced semantics and lessened hunt space, execution velocity increase has been accounted for in the quality grouping assignment from microarray explores different avenues regarding ontology based bunching. In the web use mining and next page forecast assignment, semantics-mindful consecutive example mining calculations was demonstrated to perform 4 times speedier than customary and nonsemantics- mindful calculations[33].

Ontology based methodologies enhance the consistency of data mining results too. Marinica et al. [34], [35] displayed post-handling of the affiliation guideline mining results utilizing a ontology for the consistency checking. Semantically conflicting affiliation standards are pruned and separated out with the assistance of ontology and rationale thinking.

#### B. Semantics rich data mining results

Ontology can likewise aid in advancing data mining results with formal semantics. Semantics rich data mining results are normal from ontology based methodologies contrasted and approaches without utilizing ontologies [23]. Without knowing semantics of the traits or itemsets, affiliation principles mining generally produce an excess of standards or even conflicting guidelines. Ontology based affiliation principle mining extensions the semantic hole of the area information and the affiliation standard mining calculation. It brings about better accumulation and representation of affiliation guidelines by pruning the outcomes or decreasing the pursuit space. The top positioned standards additionally bring about high bolster measure for the focusing on space [3].

#### C. Performing data mining errand that is unachievable with customary data mining systems

Certain data mining errands that are not achievable by customary data mining systems can be fulfilled by ontology based methodologies. Case in point, conventional characterization undertaking more often than not requires in any event sensible measure of named information as former learning. Utilizing ontology as the determination of earlier information, order assignment without enough marked

information is demonstrated to have a similar execution contrasted and customary arrangement systems [2]. Utilizing the ontology as a theoretical consistency imperative, the model with unlabeled information can be tuned into the particular case that have the best consistency with the earlier learning (i.e., ontology). Arrangement assignment without named or commented information is likewise reported in the ontology based content characterization errand [36].

## VI. CONCLUSION

The advances in learning building and data mining advance semantic data mining, which conveys rich semantics to all phases of data mining methodology. Numerous examination endeavors have validated the playing point of joining space learning into data mining. Formal semantics encoded in the ontology is all around organized which is simple for the machine to peruse and process accordingly make it a nature approach to utilize ontologies in semantic data mining. Utilizing ontologies, semantic data mining has points of interest to extension semantic holes between the information, applications, data mining calculations, and data mining results, give the data mining calculation with former learning which either controls the mining process or decreases the inquiry space, and to give a formal approach to speaking to the data mining stream, from information preprocessing to mining results.

To handle and control the enormous information have brought serious examination up in the data mining group. With the improvement of learning designing, particularly Semantic Web methods, mining expansive sum, semantics rich, and heterogeneous information rises as an imperative examination theme in the group. Numerous specialists have brought up, work along semantic data mining is still in its initial stage.

Ontology based semantic data mining is by all accounts one of most encouraging methodologies. The real test is to grow more programmed semantic data mining calculations and frameworks by using the full quality of formal ontology that has very much characterized representation dialect, formal semantics, and thinking instruments for rationale surmising and consistency checking.

## REFERENCES

- [1] J. H an and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers Inc., S an Francisco, CA, US A, 2005.
- [2] N. Balcan, A. Blum, and Y. Mansour. Exploiting ontology structures and unlabeled data for learning. In Proceedings of the 30th International Conference on Machine Learning, pages 1112-1120, 2013.
- [3] A. Bellandi, B. F urletti, Y. Grossi, and A. Romei. Ontology-driven association rule extraction: A case study. Contexts and Ontologies Representation and Reasoning, page 10, 2007.

- [4] S. I. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2 edition, 2003.
- [5] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5):907-928, 1995.
- [6] R. S. Tudor, V. R. Benjamins, and D. F. Ensle. *Knowledge engineering: principles and methods*. *Data & knowledge engineering*, 25(1):161-197, 1998.
- [7] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):28-37, 2001.
- [8] OWL Web Ontology Language. <http://www.w3.org/TR/owl-ref/>.
- [9] D. C. Wimalasuriya and D. Dou. Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science*, 36(3):306-323, 2010.
- [10] N. Khasawneh and C.-C. Chan. Active user-based and ontology-based web log data preprocessing for web usage mining. In *Proceedings of the 2006 IEEE IJWICACM International Conference on Web Intelligence*, pages 325-328, 2006.
- [11] D. Perez-Rey, A. Anguita, and I. Crespo. Ontodataclean: Ontology based integration and preprocessing of distributed data. In *Biological and Medical Data Analysis*, pages 262-272. Springer, 2006.
- [12] D. Tanasa and B. Trousse. Advanced data preprocessing for intersites web usage mining. *Intelligent Systems, IEEE*, 19(2):59-65, 2004.
- [13] M. Erdmann, A. Maedche, H. P. Schnurr, and S. Staab. From manual to semi-automatic semantic annotation: About ontology-based text annotation tools. In *Proceedings of the COLING-2000 Workshop on Semantic Annotation and Intelligent Content*, pages 79-85, 2000.
- [14] A. Kiryakov, S. Popov, I. Terziev, D. Manov, and D. Ognyanoff. Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2(1):49-79, 2004.
- [15] H. Liu, D. Dou, R. Jin, P. LePendu, and N. Saha. Mining biomedical ontologies and data using rdf hypergraphs. In *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, volume I, pages 141-146. IEEE, 2013.
- [16] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.
- [17] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651-666, 2010.
- [18] A. Hotho, A. Maedche, and S. Staab. Ontology-based text document clustering. *KI*, 16(4):48-54, 2002.
- [19] A. Hotho, S. Staab, and G. Stumme. Ontologies improve text document clustering. In *Proceedings of the third IEEE International Conference on Data Mining*, pages 541-544, 2003.
- [20] L. Jing, L. Zhou, M. K. Ng, and J. Z. Huang. Ontology-based distance measure for text clustering. In *Proc. of SIAM SDM workshop on text mining*, Bethesda, Maryland, USA, 2006.
- [21] Y. Karkaletsis, P. Fragkou, G. Petasis, and E. Iosif. Ontology based information extraction from text. In G. Paliouras, C. Spyropoulos, and G. Tsatsaronis, editors, *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, volume 6050 of *Lecture Notes in Computer Science*, pages 89-109. Springer Berlin Heidelberg, 2011.
- [22] H. Alani, S. Kim, D. E. Millard, M. J. Weal, W. Hall, P. H. Lewis, and N. R. S. Hadbolt. Automatic ontology-based knowledge extraction from web documents. *Intelligent Systems, IEEE*, 18(1):14-21, 2003.
- [23] S. Kara, O. Alan, O. Sabuncu, S. Akpınar, N. K. Cicekli, and F. N. Alpaslan. An ontology-based retrieval system using semantic indexing. *Information Systems*, 37(4):294-305, 2012.
- [24] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734-749, 2005.
- [25] R. Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331-370, 2002.
- [26] Y. Koren. The bigchaos solution to the netflix grand prize. *Netflix prize documentation*, 81, 2009.
- [27] A. Toscher, M. Jahrer, and R. M. Bell. The bigchaos solution to the netflix grand prize. *Netflix prize documentation*, 2009.
- [28] W. Aljandal, Y. Bahirwani, D. Caragea, and W. H. Hsu. Ontology-aware classification and association rule mining for interest and link prediction in social networks. In *AMI Spring Symposium: Social Semantic Web: Where Web 2.0 Meets Web 3.0*, pages 3-8, 2009.
- [29] A. Thor, P. Anderson, L. Raschid, S. Navlakha, B. Saha, S. Khuller, and X.-N. Zhang. Link prediction for annotation graphs using graph summarization. In *The Semantic Web-ISWC 2011*, pages 714-729. Springer, 2011.
- [30] C. Ramakrishnan, W. H. Milnor, M. Perry, and A. P. Sheth. Discovering informative connection subgraphs in multi-relational graphs. *ACM SIGKDD Explorations Newsletter*, 7(2):56-63, 2005.
- [31] I. Cantador, A. Bellogin, and P. Castells. A multilayer ontology-based hybrid recommendation model. *AI Communications*, 21(2):203-210, 2008.
- [32] L. Zhuhadar, O. Nasraoui, R. Wyatt, and E. Romero. Multi-model ontology-based hybrid recommender system in e-learning domain. In *Proceedings of the 2009 IEEE IJWICACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 03*, pages 91-95, 2009.
- [33] N. R. Mabroukeh and C. I. Ezeife. Using domain ontology for semantic web usage mining and next page prediction. In *Proceedings of the 8th ACM conference on Information and knowledge management*, pages 1677-1680. ACM, 2009.
- [34] C. Marinica and F. Guillet. Knowledge-based interactive postmining of association rules using ontologies. *Knowledge and Data Engineering, IEEE Transactions on*, 22(6):784-797, 2010.
- [35] C. Marinica, F. Guillet, and H. Briand. Post-processing of discovered association rules using ontologies. In *Data Mining Workshops, 2008. ICDMW '08. IEEE International Conference on*, pages 126-133, 2008.
- [36] M. Allahyari, K. J. Kochut, and M. Janik. Ontology-based text classification into dynamically defined topics. In *Semantic Computing (ICSC), 2014 IEEE International Conference on*, pages 273-278, 2014.