International Journal on Emerging Technologies **11**(4): 233-239(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Multi-agent Distributed Data Mining: Challenges and Research Directions

***V. Devasekhar*[1] *and P. Natarajan*[2]**
[1]Research Scholar, SCOPE, VITU, Vellore & Associate Professor,
Department of CSE-GNITC, Hyderabad (Telangana), India.
[2]Professor, SCOPE, VITU, Vellore (Tamil Nadu), India.

(Corresponding author: P. Natarajan)
(Received 20 May 2020, Revised 20 June 2020, Accepted 04 July 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Multi-agent-based computing aim is to deal with complex data systems has revealed chances to improve distributed data mining systems frameworks for distributed problem solving. Multi-agent systems consist of a network of problem solvers which are modelled in agents and provides a powerful abstraction that helps to model systems where multiple entities exhibiting self-directed behavior's must co-exist within an environment.

This paper presents the overview of multi agent-based distributed data mining techniques available in the literature. Various distributed data mining and machine learning methods with their efficiencies on different data sets are presented in detail. Efficient tools, applications and several problems are discussed. Comparative analysis of different techniques and their advantages and disadvantages are discussed. This paper also presented a framework that based on multi-agent approach to data mining of large datasets and research directions.

Keywords: Data Mining, Machine Learning, Classification, Clustering, Multi-Agents, Smart space.

Abbreviations: DAI, Distributed Artificial Intelligence, DPS, Distributed Problem Solving, MAS, Multi Agent Systems, DDM, Distributed Data Mining, KDD, Knowledge Discovery, DML, Distributed Machine Learning, ANN, Artificial Neural Nets, SVM, Support Vector Regression, FSVR, Fuzzy Support Vector Regression. ABM, Agent-Based Models, MADM, Multi-Agent Distributed Data Mining,

I. INTRODUCTION

Distributed Artificial Intelligence (DAI) or Decentralized Artificial Intelligence [1] is a class of technologies and methods that is emerged as part of Artificial Intelligence (AI) for over three decades. DAI systems have been comprising of different independent entities that are connected specific area. In general, DAI divided into Distributed Problem Solving (DPS) focus on the information management, second one is Multi-agent Systems (MAS) [1, 2] focus on behavior management.

MAS [3] permits the subproblems of an imperative fulfilment issue to be subcontracted to various problem-solving agents with their very own advantages and objectives. Moreover, domains with different agents of any sort, including self-sufficient vehicles and even some human agents, are starting to be considered.

Distributed Data Mining [15, 16] for multi-agent systems has attracted much attention and growing interests from researchers recently. Distributed data mining is the mining of distributed data and it intends to obtain global knowledge from the local data distributed sites.

Multi-Agent Systems [12, 14] are a genuinely old class of methods, where each agent connects between one another dependent on pre-decided guidelines/limitations and, as a result, an aggregate behavior that is sufficient from these interactions.

II. MAS: APPLICATIONS

Smart space [25] is a distributed surrounding environment with existing, inside it, dynamic large number of occupants ("living and non-living, for example robots") solving their own and basic tasks. The strategic this condition is to give, for smart space occupants, with universal correspondence, customized administrations, updates and customized proposals in an easy to understand mode where and when required in pervasive and unpretentious style. This definition covers numerous significant current applications [16-19], for example "smart home, smart city, spatial security frameworks, emergency management, environmental monitoring, health care and disability person's assistance, smart grid", and numerous others.

III. DISTRIBUTED DATA MINING AND MACHINE LEARNING

Data mining [40] is commonly used mining techniques to analyse different kind of data sets. Data mining is a process in knowledge discovery in databases (KDD) [13]. Data mining deals with algorithms that generates knowledge from given data. In general, data mining is used for "(i) classification, (ii) clustering, (iii) regression, and (iv) association rule learning".

Classification [21] sorts the data into predefined categories. For example, web pages may be classified into categories of reading, such as technology, science, financial, etc. Classification algorithm learns from labelled examples in order to discover patterns of the data. Clustering [14, 15] groups the data which are similar without knowing predefined categories. In this regard, there is no labelled example for the algorithm, instead clustering algorithm tries to discover clusters that best describe the data. Regression finds a function, which poses the data with the least error, this includes multiple regression and other statistical methods. Finally, association rule learning examines relationships between features of the data.

A. Distributed Data Mining (DDM)

The principle factors are prompted the development of Distributed Data Mining [31, 32] includes– "privacy of sensitive data, transmission cost, calculation cost and memory cost". The goal of these technique is to remove valuable data from information situated at heterogeneous points. DDM [30, 33, 34] pursues decentralized mining techniques. Alfredo Cuzzocrea [30] expressed that encircling a system for DDM is testing by circulated condition, yet in addition for its productive asset sharing and limited computational multifaceted nature details. DDM [36] for the most part contains two varieties — information dispersed and calculation disseminated. In the previous strategy, information is circulated between "heterogeneous sites" [35] at nearby level and calculation is facilitated at worldwide level. In the last strategy, calculation is disseminated among heterogeneous destinations at neighborhood level and information is facilitated at worldwide level.

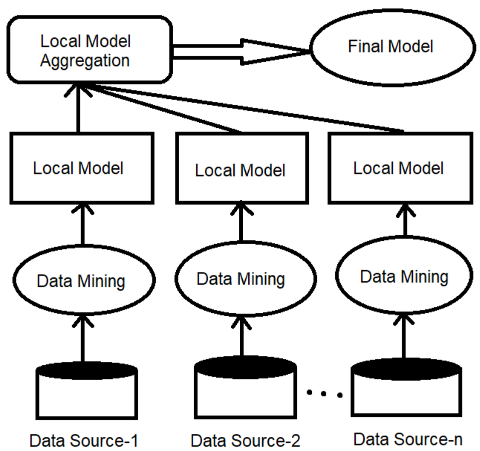

Fig. 1. Distributed Data Mining.

In Fig. 1, presented Distributed Data Mining framework. DDM methods on information at heterogeneous destinations as nearby approach lastly the DM methods provide outcome resultant to global model [29]. Authors in [31, 32] examined that few specialists divided DDM in two different ways: using computational resources efficiently and performing data mining techniques for knowledge discovery at top level. In [35] presented about specific problems in creating DDM methods, in particular defining appropriate DDM calculations for "heterogeneous datasets", limiting complexity issues, upgrading information protection at disseminated destinations and keeping up neighborhood datasets self-governance. Grigorios Tsoumakas *et al.* [34] displayed local and global framework for Distributed Data Mining.

For example, the title of the web page may relate to the type of reading, in term of DDM, rules may be discovered from distributed data source and apparently related. The rules can be combined and pruned to present the most meaningful result to the user.

Data mining presents various challenges includes distributed data source, high dimensionality data [23], etc., and all other related processes are dependent.

Fig. 2 presents overview of distributed data mining techniques that includes multi agent environment and preserving the privacy of the information.

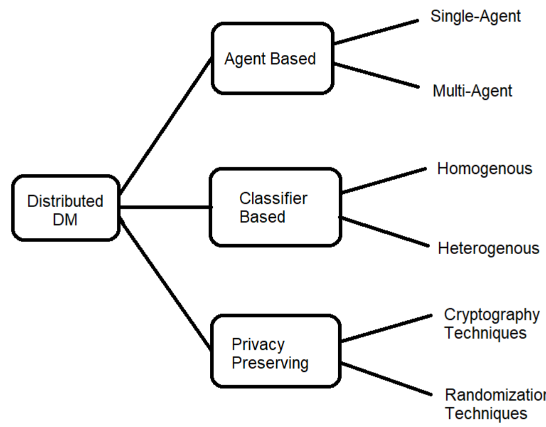

Fig. 2. DDM with other techniques.

B. Distributed Machine Learning

There has been a developing interest for creating adaptable learning methods utilizing appropriated advancement [17], motivated by the rise of big data sets[18], security and protection imperatives [19], and the idea of helpful and community oriented learning agents [20]. Distributed Machine Learning (DML) enables different agents to cooperatively classify data from full scale information. Almost all these methods are "graph based", where every node in the diagram speaks to a part of data or an agent. A few approaches accept there is a "central server (or a group of server clients)" in the system and the sub-agents transmit privately learned data to the server(s), which at that point perform information combination [18]. Different approaches accept that processing power is appropriated between agents [19].

IV. TOOLS AND TECHNIQUES FOR DATA MINING

In this section, discussed techniques that are used in classification and clustering, and tools, libraries for data mining.

A. Classification methods for data mining

(a) *Artificial Neural Nets*

i. ANN [12, 41] requires training of data, identify the complex nonlinear correlations variables, identify the interactions between predictor variables and running multiple training algorithms.
ii. ANN is a data-driven and self-adaptive procedure.
iii. ANN is nonlinear statistical methods and it is easy to use and it is a non-parametric model.
iv. ANN methods are easy to implement in sequential and parallel computing facility like cluster computing. Parallel training reduces the processing.

v. ANN selects topological network structure which is a complex procedure provides significantly accurate and efficient results.

(b) *Support Vector Regression [23]*

i. SVM techniques are considered as applicable in which non-linear separations exist in data sets.

ii. SVM models create feature space, called as finite-dimensional vector space.

iii. In SVM [12], classification calculated by obtaining hyper-plane to make separation in more than one classes.

iv. SVM works well and provide clear margin of separation.

v. In SVM, Due to availability of variety of kernel tricks, the method is more flexible according to nature of input dataset.

(c) *Maximum likelihood*

i. This technique [13] is computationally intensive and so extremely slow procedure. The result is dependent on the model of evolution used.

ii. It is less flexible method especially in terms of probabilities and statistical measures calculation of uncertainty involved in data.

(d) *Minimum distance classifier*

i. The method [42] is used to classify unknown image data and it depends heavily on the spectral distinctness of the classes.

ii. In the event that at least two classes are fundamentally the same as one another regarding their unearthly reflectance then misclassifications will in general be high.

(e) *Decision tree fuzzy classifier*

i. Decision tree classifier [41] has some drawbacks such as the trustworthiness of the information depends on supplying the accurate internal and external knowledge.

ii. Some change in input data that provide big variations in the decision tree.

(f) *Fuzzy Support Vector Regression*

i. FSVM [42] recognizes Several stochastic correlations.

ii. Before computation FSVM need information about knowledge data sets such as stochastic and probabilistic information.

iii. Comparatively it functions slower.

Applications of classification techniques [41, 42] are "email spam analysis, Bank customer loan data prediction, climate data analysis, Sentiment analysis" and automotive car driving etc. Tools analyzing raw data are "Jupyter, NumPy, Matplotlib, Pandas, ScikitLearn, NLTK, TensorFlow, Seaborn, Basemap", etc.

B. Clustering methods for data mining

Clustering [37] is the method of partitioning the population or data focuses into various groups to such an extent that information focuses in similar groups are increasingly like other information focuses in a similar group than those in different groups.

Clustering methods [39] can be classified into the following categories: "(i) Partitioning, (ii) Hierarchical method, (iii) Density-based method and (iv) Grid-based methods". Well known techniques [38] for clustering are K-Means clustering, Agglomerative Algorithms, Divisive Algorithms Probabilistic Clustering, K-Medoids Methods, Subspace Clustering and Constraint-Based Clustering.

Major application of clustering methods [38, 40] are "scientific data exploration, information retrieval and text mining, spatial database applications, Web analysis, CRM, marketing, medical diagnostics, computational biology" and many others.

C. Tools for Clustering & Classification of Datasets

In this section, presented list of tools [7-9] available for analyzing different type of data sets.

List of Tools.

Tool	Features/Characteristically and Computational Aspects
"Rapid Miner"	This tool has features – "data cleansing, filtering, clustering, built in templates etc.," It supports Python and R into work flows.
"Weka"	This tool has features – "data pre-processing, classification, regression, clustering, association rules and visualization". It supports "Java based free and open source software licensed Linux, Mac OS X and Windows".
Knime	This tool has features- "conduct data pre-processing, collection, analysis, modelling and reporting". It supports Java and built upon Eclipse.
DataMelt	This tool has features- "statistics, numeric and symbolic computations, scientific visualization, etc". DMelt also provides "data mining classification tools and interactive visualizations using 2D/3D plots and histograms".
Apache Mahout	This tool has features- "clustering, classification and frequent pattern mining". It also supports in "distributed mode that helps easy integration with Hadoop".
ELKI	This tool has features- "cluster analysis and outlier detection with a compilation of numerous algorithms from both these domains".
KEEL	Knowledge Extraction for Evolutionary Learning Java based tool has features- "data pre-processing, statistical libraries and some standard data mining and evolutionary learning algorithms".
Rattle	This tool has features- "R Analytical Tool and has been developed using the R statistical programming language" it support platforms "Linux, Mac OS and Windows, and features statistics, clustering, modelling and visualization with the compuwer of R".
Data Preparator	This tool has features- "Performs cleaning, extraction and transformation of data".
Lisp Miner	This tool has features- Achieves data pre-processing by scrutinizing the stream and the corpus collected.
RStudio	This tool has features- "statistical computing, multimedia data analysis and graphics" It supports "RStudio and open-source integrated development environment for R".
GRETL	It is a cross-platform system software for

	econometric reasoning, written in the C programming script.
RATS	Regression Analysis of Time Series (RATS), is an "Analytical package tool for the time series analysis and geometric".
ORANGE	Performs data analysis and knowledge exploration with an attractive data visualization. It performs statistical assessments, box plots and scatter plots.
TANAGRA	This tool has features- "Visualization, Descriptive statistics, Instance selection, feature selection, feature construction, regression, factor analysis, Data mining techniques".
Sewebar-Cms	This tool helps while opting rules among diversified rules in association rule mining.
WebViz	Analyze the patterns and caters them in the style of graphical patterns.
"Scikit-learn"	This tool has features- "machine learning library and various classification, regression and clustering algorithms and DBSCAN". It supports "Python numerical and scientific libraries NumPyandSciPy".

V. MULTI-AGENT SYSTEMS

A multi-agent setting with "tuple<N,S,A,R,P,O,γ>", in which N is the number of agents, S is state space, "A={A1,...,AN}" is the set of actions for all agents, "P is the transition probability among the states, R is the reward function and O={O1,...,ON}is the set of observations" for all agents. Within any type of the environment, use a to denote the vector of actions for all agents, a-i the vector of all agents except agent i, τi represents observation-action history of agent i, and τ is the observation-action of all agents. Also, T,S, and A are the observation-action space, state space and action space. Then, in a cooperative problem with N agents with full observability of the environment, each agent I at time-step t observes the global states t and uses the local stochastic policy πi to take action ati and then receives reward rti. If the environment is fully cooperative, at each time step all agents observe a joint reward value rt, i.e., rt1 =· · ·=rtN = rt. If the agents are not able to fully observe the state of the system, each agent only accesses its own local observation oti.

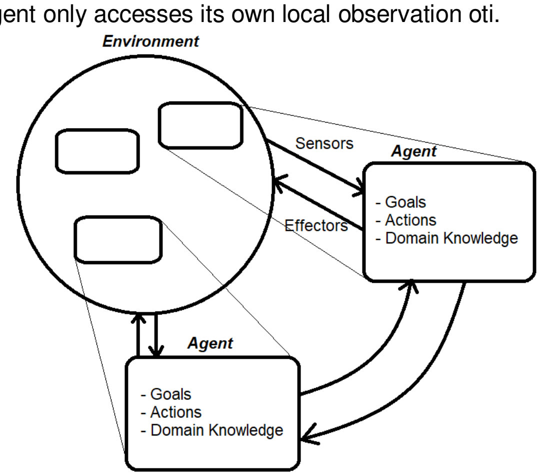

Fig. 3. Multi-Agent systems.

An agent playing chess is in a two-agent condition. There are, be that as it may, some inconspicuous issues. Does a agent A (the cab driver for instance) need to treat an item B (another vehicle) as an operator, or would it be able to be dealt with just as a stochastically carrying on object, closely resembling waves at the sea shore or leaves blowing in the breeze? In Fig. 3, architecture of Multi Agent System is given below.

The key differentiation is whether B's conduct is best portrayed as augmenting a presentation measure whose worth relies upon operator A's conduct. For instance, in chess, the adversary substance B is attempting to expand its exhibition measure, which, by the principles of chess, limits operator A's presentation measure. Thus, chess is a competitive multiagent environment. In the taxi-driving condition, then again, maintaining a strategic distance from crashes boosts the exhibition proportion all things considered, so it is a mostly agreeable multiagent condition. It is additionally mostly aggressive in light of the fact that, for instance, just a single vehicle can consume a parking spot.

Some of the MAS (Multi-Agent System) of DDM are "PADMA (Parallel DM Agents), JAM (Java Agents for Meta-learning), bodhi and papyrus".

A. Multi-agent distributed data mining

The Multi-agent distributed data mining (MADM)technique [4] is a distributed agent-based DM method (see figure 4) which is linked with multi-agents and processing in each level.

Agent [4,5] is the principle class on the graph that uses three areas: "Roles, Attributes and Perception". In general an agent ought to have:

- Interface operator interfaces with the client at local level. The agents having combined techniques to interact other sites and to the global level.
- Resource agent effectively keeps up the meta-database and maps heterogeneous database at global level.
- Result agent screens development of mining agent and get the outcome from it. The layout results put away is displayed to the end client.
- Broker agent diffuses operators for preparing in conveyed condition. It monitors number of agents, screens the capacity of them, and so forth.
- Query agent [8] is created on request by the environment. It stores the local level blueprint and global level outline.
- Mobile agent screens [8] the whole system for appropriate usefulness. The drawback here is expanded system traffic and necessity of introducing stage for operators at each disseminated information site.

Multi-Agent Systems: Related work

Recent methodologies [2, 7, 42] for sure hope to profit by the joint utilization of MAT and Machine learning (and all the more explicitly "support adapting, deep learning and profound convolutional systems)", since ML can utilize ABM as a situation and a prize generator, while Agent based Modeling (ABM) can utilize ML to refine the inner models of the specialists [1, 43, 44]. The neural systems are in this manner utilized as a computational guess of the non-direct, multivariate time arrangement created by the ABM, or by and large as computational emulators of whole ABMs [4].

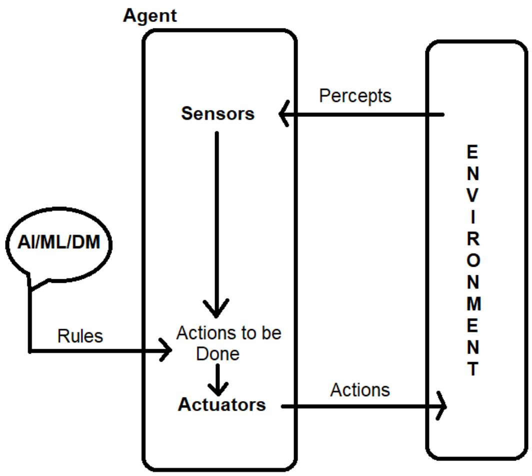

Fig. 4. Multi-Agent DDM.

The Agent-Based Models (ABM) [4] are used in several applications from "urban planning to epidemiology, from economics to transportation". In [25] presented about an agent and antique way to deal with DDM by "Collaborative Filtering (CF)" among agents.

For the most part 3 artefacts were utilized by agents, in particular J48 classifier, Instance-Base and Oracle. Coordinating agents utilize use Oracle for extricating data for learning information and laborer agents utilize the split information for preparing. These information in Instance-Base. An underlying CF technique model is worked by J48 classifier. For the most part "Adult, German, letter, poker and waveform datasets from UCI (University of California, Irvine)" storehouse were utilized.

As discussed about CoLe2, [26, 27] examined Smart Space condition and Decision of virtual neighbors. In [28] examined synergistic model and multi-system DDM. In [29] examined specialist advanced DARM, which is joint collaboration of agents is discussed [30].

In [31] talked about a novel structure by connection and DM for measurements, to be specific "learning, knowledge, interface, interaction, social, application and performance". Chayapol Moemeng *et al.*, [32] examined another model, I-analyst by the executives of assets stage. Authors in [33, 34] talked about conveying "mobile agents" for looking through information, which were dispersed [18, 24]. It can defeat organize "congestion, security and unreliability".

B. Multi agents for smart space

Multi-agent based DDM [16, 17, 45] frameworks are made out of heterogeneous specially "ad-hoc wireless sensor network, effectors, as well as of diverse software" planned for "monitoring and on-line processing data produced by the sensors, performing intelligent data analysis, human's behavior pattern recognition to predict his/her needs, classification, context detection, user profile learning", etc. In this manner, MADDM [14] is developed utilizing numerous systems, of organized calculation, correspondence, detecting [22, 23] and activity that decide the first testing issue. The subsequent test is brought about data perceived by different sensors just as information, so as to give, for "the smart space, with intelligent, adaptive, self-organizing, learnable, actionable and efficient performance".

Research Directions: The scope of the research is to analysis and investigate methodologies and techniques published in existing literature and design an efficient framework that provide an effective result for smart space data sets.

Challenages: Homogeneous Non-Communicating MAS- "Several different agents with identical structure (sensors, effectors, domain knowledge and decision functions)", but they have different sensor input and effector output. In other words, they are arranged diversely in the earth and they settle on their own choices with respect to which actions to make.

Issues- "Reactive vs. deliberative agents, Local or global perspective, Modeling of other agents, states", how to affect others.

Heterogeneous Non-Communicating Multi-agent Systems-The agents may have different goals, actions, and/or domain knowledge.

Issues- "Benevolence vs. competitiveness, Fixed vs. learning agents (arms race, credit-assignment), Modeling of others' goals, actions, and knowledge Resource management (interdependent actions), Social conventions".

Homogeneous Communicating MAS- As in the "homogeneous, non-communicating case", the agents are indistinguishable aside from that they are arranged distinctively in nature. Nonetheless, in this situation, the agents can impart legitimately as showed by the arrows associating the agents.

Issues- Distributed sensing, Communication content

Heterogeneous Communicating Multi-agent Systems- In this situation, enable the agents to be "heterogeneous to any degree from homogeneity to full heterogeneity". Many multi-agent frameworks incorporate people as at least one of the agents.

Issues- "Understanding each other, Planning communicative acts, Benevolence vs. competitiveness Negotiation, learning opportunities, Resource management, Commitment/decommitment Evolving language, Collaborative localization Effects of speech acts on global dynamics, Changing shape and size".

Research Directions

1. In the scenario of multi-agent data, obtained from distributed (non-centralized) sources, the instances and variable's nature may be heterogeneous along with the presence of outliers. The amount of diversity in statistical distribution, while developing the prediction and classification model is a key challenge.

In this research work, develop new soft computing and applied statistics-based approach to address this issue and test their performance using standard text datasets/corpuses from repositories.

2. Data mining is an efficient tool to retrieve knowledge from big data sets. There is a high probability for the presence of duplicate instances in the data that is streamed and collected from distributed sources. Also, the number of predictor variables may be massive.

To overcome this challenge, develop the methodology for detecting the duplicate instances and selection of optimally minimal set of variable vectors in data sample collected from multi-agent distributed sources. This process will reflect the benefits like- reduction in the computational time complexity, which will play a vital role in further modeling with reduced dimensioned data.

3. Process of data mining need the important data into a central location from distributed sites. In this scenario, privacy is also a major concern, which may prevent the distributed agents from sharing their data with others due to confidentiality concerns.

To deal with this challenge, develop the protocol which provides privacy preservation in the de-centralized mining and knowledge exploration process. Here, adopt some specific cryptographic primitives and building blocks that enable "data sharing, while protecting data privacy".

VI. CONCLUSION

This paper presents literature survey of multi-agent systems and data mining techniques. In this paper, presented evolution of multi-agent systems and their applications to different environments. Further, listed the tools for data analysis along with their advantages and disadvantages.

VII. FUTURE SCOPE

The study of multi-agent systems is "concerned with the development and analysis of sophisticated AI problem-solving and control architectures for both single-agent and multiple-agent systems. Major research areas include agent-oriented software engineering, and coordination distributed constraint optimization, distributed problem-solving multi-agent learning, language evolution, and economics.

In future work, propose an efficient technique for distributed data mining based on multi-agent, classifiers and privacy preserving techniques for analyzing smart space data sets.

REFERENCES

[1]. Stone, P., & Veloso, M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. Autonomous Robots, *8*(3): 345–383.

[2]. Rand, W. (2007). Machine learning meets agent-based modeling: when not to go to a bar. Unpublished Paper.

[3]. Ponomarev, S., & Voronkov, A. E. (2017). Multi-agent systems and decentralized artificial superintelligence. arXiv:1702.08529.

[4]. Van Der Hoog, S. (2017). Deep Learning in (and of) Agent-Based Models: A Prospectus. Preprint arXiv: 1706.06302.

[5]. Yeoh, W., Yokoo, M. (2012). Distributed Problem Solving. *AI Magazine, 33*(3): 53–65

[6]. Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., Wang, J. (2018). Mean Field Multi-Agent Reinforcement Learning". arXiv:1802.05438.

[7]. Mguni, D., Jennings, & J., Munoz de Cote, E. (2018). Decentralised Learning in Systems with Many, Many Strategic Agents. arXiv:1803.05028v1.

[8]. Prasad, B.R., Agarwal, S., 2016. Comparative study of big data computing and storage tools: a review. *Int. J. Database Theory App., 9*, 45–66.

[9]. Acharjya, D., & Ahmed, K. P., (2016A). A survey on big data analytics: challenges, open research issues and tools. *Int. J. Adv. Comput. Sci. App. 7*, 511–518.

[10]. Acharjya, D., & Ahmed, K.P., (2016b). A survey on big data analytics: challenges, open research issues and tools. *Int. J. Adv. Comput. Sci. App. 7*, 511–518.

[11]. Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process*, 1–16.

[12]. R. O. Duda, P. E. Hart, and D. G. Stork (2001). Pattern classification. Wiley-interscience.

[13]. G.V. Trunk (1979). A problem of dimensionality: A simple example. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence, 3,* 306–307.

[14]. Moemeng, Chayapol, Gorodetsky, Vladimir, Zuo, Ziye, Yang, Yong, Zhang, Chengqi (2009). Agent-Based Distributed Data Mining: A Survey. 10.1007/978-1-4419-0522-2_3.

[15]. M. Wooldridge (2009). An Introduction to multiAgent systems", 2nd ed., John Wiley & Sons, ISBN-10: 0470519460, 2009.

[16]. M. Li, D. G. Andersen, A. J. Smola, and K. Yu. (2014). Communication efficientdistributed machine learning with the parameter server. In Advances in NeuralInformation Processing Systems, 19–27.

[17]. R. Caruana (1997). Multitask Learning. *Machine Learning,* 41–75.

[18]. V. Cevher, S. Becker, and M. Schmidt (2014). Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics.*IEEE Signal Processing Magazine*, 32–43.

[19]. J. Chen, C. Richard, and A. H. Sayed (2014). Multitask diffusion adaptation overnetworks. *IEEE Transactions on Signal Processing, 62*(6), 4129–4144.

[20]. J. Chen and A. H. Sayed (2012). Diffusion adaptation strategies for distributedoptimization and learning over networks. *IEEE Transactions on Signal Processing, 60*(8), 4289–4305.

[21]. Xing, E. P., Ho, Q., Dai, W., Kim, J. K., Wei, J., Lee, S., ... & Yu, Y. (2015). Petuum: A new platform for distributed machine learning on big data. *IEEE Transactions on Big Data*, *1*(2), 49-67.

[22]. Wang, J., Kolar, M., & Srerbo, N. (2016, May). Distributed multi-task learning. In *Artificial Intelligence and Statistics* (pp. 751-760).

[23]. Forman, G., Zhang, B. (2000). Distributed Data Clustering Can Be Efficient and Exact. *SIGKDD Explorations, 2*(2), 34–38.

[24]. Cao, L., Gorodetsky, V. and Mitkas, P. Agent (2009). Mining: The Synergy of Agents and Data Mining. *IEEE Intelligent Systems*.

[25]. Vladimir Gorodetsky (2012). Agents and Distributed Data Mining in Smart Space: Challenges and Perspectives. ADMI: 153-165.

[26]. Cao, L. (2009). Data Mining and Multi-agent Integration (edited). Springer.

[27]. Cao, L., Gorodetsky, V., & Mitkas, P.A. (2009). Agent Mining: The Synergy of Agents and Data Mining. *IEEE Intelligent Systems, 24*(3), 64–72.

[28]. Cao, L., Weiss, G., Yu, P.S. (2012) A Brief Introduction to Agent Mining. *Journal of Autonomous Agents and Multi-Agent Systems, 25*, 419–424.

[29]. Vinaya Sawant and Ketan Shah (2013). A review of Distributed Data Mining using agents, *International Journal of Advanced Technology & Engineering Research (IJATER),* 3(5), 27-33.
[30]. Alfredo Cuzzocrea (2013) Models and algorithms for high-performance distributed data mining. *Elsevier Journal of Parallel and Distributed computing, 73*(93), 281-283.
[31]. Kargupt, Kamath, and Chan (1999). Distributed and Parallel Data Mining: Emergence, Growth and Future Directions. *Advances in Distributed Data Mining*, (eds.), Hillol Kargupta and Philip Chan, AAAI Press, 407-416.
[32]. Zaki, M. J. and Pan, Y. (2002). Introduction: Recent Developments in Parallel and Distributed Data Mining. *Springer Journal of Distributed and Parallel Databases, 11*(2), 123-127.
[33]. Park B. H., and Kargupta H. (2002). Distributed Data Mining: Algorithms, Systems, and Applications", In. Data mining handbook.
[34]. Tsoumakas, G, and Vlahavas I. (2008). Distributed Data Mining. Encyclopedia of Data Warehousing and Mining, 2nd Edition John Wang (Ed.), Idea *Group Reference,* 709-715.
[35]. Fu. Y. (2001). Distributed Data Mining: An Overview In: Newsletter of the *IEEE Technical Committee on Distributed Processing,* 5–9.
[36]. Hillol Kargupta. An Introduction to Distributed Data Mining", http://www.eecs.wsu.edu/~hillol
[37]. Babu, G.P. and Marty, M.N. (1994). Clustering with evolution strategies. *Pattern Recognition, 27*(2), 321-329.
[38]. Jain, A.K, Murthy, M.N., and Flynn P.J. (1999). Data clustering: a review. *ACM Computing Surveys, 31*(3), 264-323.
[39]. Kolatch, E. (2001). Clustering Algorithms for Spatial Databases: A Survey. PDF is available on the Web.
[40]. Zhang, T., Ramakrishnan, R., and LIVNY, M. (1997). BIRCH: A new data clustering algorithm and its applications. *Journal of Data Mining and Knowledge Discovery,* 141-182.
[41]. Han, J., & Kamber, M. (2001). Data Mining: Concepts and Techniques, Academic Press, London.
[42]. V. Devasekhar, & P. Natarajan (2018). Multi-agent based data mining aggregation approaches using machine learning techniques. *International Journal of Engineering & Technology*, *7*(3), 1136-1139.
[43]. Zhinan Penga,Yiyi Zhaob, Jiangping Hua Bijoy & Kumar Ghoshac (2019). Data-driven optimal tracking control of discrete-time multi-agent systems with two-stage policy iteration algorithm, Information Sciences, *481*, 189-202.
[44]. Hao, P. Y. (2019). Dual possibilistic regression analysis using support vector networks. *Fuzzy Sets and Systems*, *387*, 1-34.
[45]. Xianfei Yang, Xiang Yu, & Hui Lu, (2020). Dual possibilistic regression models of support vector machines and application in power load forecasting*. International Journal of Distributed Sensor Networks, 16*.

How to cite this article: Devasekhar, V. and Natarajan, P. (2020). Multi-agent Distributed Data Mining: Challenges and Research Directions. *International Journal on Emerging Technologies*, *11*(4): 233–239.