Design of a blockchain based secure and efficient ontology generation model for multiple data genres using augmented stratification in healthcare industry

Suniti Purbey (sunitynu@gmail.com)
Amity University Chhattisgarh
Brijesh Khandelwal
Amity University Chhattisgarh
Ashutosh Kumar Choudhary
GMR Institute of Technology

Research Article

Keywords: Ontology, blockchain, correlation, security, OWL, augmented, stratification

Posted Date: January 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2473807/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Signal, Image and Video Processing on May 10th, 2023. See the published version at https://doi.org/10.1007/s11760-023-02576-1.

Abstract

Ontology generation is a process of relationship analysis, and representation for multiple data categories using automatic or semi-automatic approaches. This process requires a domain knowledgebase that describes given input data using entity-to-entity relations. A wide variety of approaches are proposed for this purpose, and each of them processes & converts input data using multiple relationship evaluation stages. These stages include data-preprocessing, correlation analysis, entity mapping, and ontology generation. A very few of these approaches are dataset independent, and most of them do not implement security measures during ontology generation, which limits their security, scalability & deployment capabilities during real-time implementation. Thus, in this text a blockchain based secure & efficient ontology generation model for multiple data genres using augmented stratification (BOGMAS) is described. The BOGMAS model uses a semi-supervised approach for ontology generation from almost any structured or unstructured dataset. It uses a variance-based method (VBM) for reduction of redundant numerical features from the dataset, while textual features are converted to numerical values via standard word2vec model, and then processed using VBM. This model uses a combination of linear support vector machine (LSVM), and extra trees (ET) stratifiers for variance estimation, which makes the model highly efficient, and reduces redundant features from the output ontology. These feature sets & their variances are given to a correlation engine for relationship estimation, and ontology generation. Each ontology record is secured using a mutable proof-of-work (PoW) based blockchain model, which assists in imbibing transparency, traceability, and distributed peer-to-peer processing capabilities. The generated ontology is represented using an incremental OWL (W3C Web Ontology Language) format, which assists in dynamically sizing the ontology depending upon incoming data. Performance of the proposed BOGMAS model is evaluated in terms of precision & recall of representation, memory usage, computational complexity, and accuracy of attack detection. It is observed that the proposed model is highly efficient in terms of precision, recall & accuracy performance, but has incrementally higher computational complexity & delay of ontology formation when compared with existing approaches. Due to this incremental increase in delay, the proposed model is observed to be applicable for a wide variety of real-time scenarios, which include but are not limited to, medical ontology generation, sports ontology generation, and internet of things (IoT) ontology generation with high security levels.

1. Introduction

Ontology generation is a multidomain task which involves database analysis, rule evaluation, attribute checking, class-based analysis, class-level variance-based relationship estimation, ontology acquisition, and post-processing. These tasks require efficient models for design of each operation, and a combination of these efficient designs results into an effective ontology generation model. A typical model that generates ontologies from relational databases can be observed from figure 1 (a), wherein database schema is available for OWL generation.

Due to availability of relational data schema, this model can evaluate table-to-table mapping, attribute relationships, hierarchy rules, and attribute level constraints. Each of these rules is given to an ontology generation model [1], which groups similar database entities together, and generates an ontology tree from this cluster relationships. But real-time datasets are generally non-structured, and do not provide relationship information due to multiple application-specific issues [2]. To remove this drawback, and to establish entity-based relationships, a wide variety of system models are proposed by researchers over the years. A review of these system models is discussed in the next section of this text, which will allow readers to evaluate various advantages, nuances, limitations, and future research scopes in these models. Inspired by these observations, section 3 proposes design of BOGMAS, which is a blockchain based secure & efficient ontology generation model for multiple data genres using augmented stratification. This model uses a combination of language processing, clustering, variant feature detection, relationship classification, OWL generation, and blockchain-based security as depicted in figure 1 (b) for improving efficiency of ontology generation. This section is followed by performance evaluation of BOGMAS, wherein it is applied for ontology generation of different datasets, and its performance is compared with various state-of-the-art models.

It can be observed that 2 parties agree for a transaction, which is triggered by events, and is stored on the blockchain for future reference operations. Such a model is used in the proposed work for security of data samples. Finally, this text concludes with some interesting observations about the proposed model and recommends methods to improve their efficiency.

2. Literature Review

A wide variety of algorithmic models are available for ontology generation, and each of these models are applied to different fields for application-specific deployments. For instance, the work in [3, 4] proposes industrial ontology generation with domain identification support, and biomedical ontology generation (BOG) for Variant Call Formats (VCFs). These models have high efficiency of data representation and require low delay due to their application-specific deployment behaviour. This behaviour limits the number of supported data types, thereby accelerating the process of ontology generation. Applications of these models are observed from [5, 6], wherein researchers have used control flow graph generation, and online repository generation using ontology generation models. These applications assist in reducing processing delays, thereby improving efficiency of data processing. An advanced model for ontology creation is proposed in [7], wherein researchers are able to generate ontologies from unstructured property graphs via deep learning approach. This model has high accuracy of representation, but has limited security, which limits its usability for real-time deployments. Inspired by this, the work in [8] proposes a multiple aspect ontology model (MAOM), which assists in improving decision support for human computer interfaces (HCIs). The proposed model is highly secure, and has been tested on various applications, which assists in evaluating its deployment capabilities.

Extensions to these models are discussed in [9, 10, 11, 12], wherein Compact Brainstorm Algorithm (CBSO), personal feedback generation using student ontologies, blockchain based time-based protocol (BTBP), and machine learning for art-ontology are defined. These models showcase different applications which can be used for generation and deployment of blockchains in real-time networks.

Similarly, the work in [13, 14, 15] discusses use of ontologies in e-commerce applications, effect of different attacks on ontology systems, and fuzzy ontologies that use blockchain for improved security. These application-specific models are observed to be highly effective in terms of precision, recall, & storage cost measures, and thus are directly used for improved system design. Ontologies can be used for multimedia data representation [16], Gastroenterology and other medical fields which include infectious disease ontologies [17, 18], transaction ontologies based on smart contracts [19], and health record management ontologies [20], which assist in reducing system dependency on external sources, thereby improving their query performance. Similarly, the work defined in [21, 22, 23, 24] also proposes ontology models for renewable energy sources, keyword-based search applications, eLearning applications, and internet of medical things (IoMT) applications, where high efficiency of system design with minimum system complexity is needed. Efficient models are also proposed in [25, 26, 27, 28] which discuss use of secure blockchains for real-time scenarios under different applications. Similar models are discussed in [29, 30, 31, 32, 33, 34, 35], which propose use of crowdsensing with blockchains for efficient representation of data samples. These applications suggest that a very few ontology generation models utilize blockchain, and variance-based approaches for reducing redundancies, which can be explored for improving system performance. Based on this observation, the next section describes design of a blockchain based secure & efficient ontology generation model for multiple data genres using augmented stratification. This model is evaluated on multiple datasets and compared with some of the reviewed techniques to estimate its performance improvement with respect to standard ontology generation models.

3. Proposed Blockchain Based Secure & Efficient Ontology Generation Model For Multiple Data Genres Using Augmented Stratification

From the literature review, it is observed that a wide variety of models are available for ontology generation, and each of these models have their own limitations. These include limited security for the generated ontology, data redundancy, limited traceability capabilities, etc. In order to remove these drawbacks, this text proposes a blockchain based secure & efficient ontology generation model for multiple data genres using augmented stratification. The proposed model is depicted in Fig. 2, wherein different datasets, and their final ontological classification status is described.

From Fig. 2, it is observed that input structured/unstructured data is given to a pre-processing layer, wherein data-specific feature extraction process is applied. These features are given to a feature selection layer, which uses combination of linear support vector machine (LSVM), and extra trees (ET) classifier for finding most variant features. Extracted features are given to an ontology generation framework, which stores these features, along with tagged classes in order to generate an RDF based ontology.

Details about BOGMAS

For the purpose of ontology creation, the BOGMAS model employs a semi-supervised technique that can be applied to almost any structured or unstructured dataset. It reduces the number of redundant numerical features in the dataset by utilizing a variance-based technique (VBM), whereas textual characteristics are first transformed to numerical values using a typical word2vec model, and then it processes those numerical values using VBM. This model utilizes a mix of linear support vector machine (LSVM), and extra trees (ET) stratifiers for variance estimation. This not only makes the model incredibly efficient, but it also decreases the number of redundant features that are included in the output ontology. A correlation engine is provided with these feature sets together with their variances so that it may estimate connection strengths and provide an ontology. Every ontology record is protected by a blockchain system that is built on a changeable proof-of-work (PoW) architecture, which helps to infuse the system with transparency, traceability, and the capacity to do distributed peer-to-peer processing. An incremental OWL (W3C Web Ontology Language) format is used to express the created ontology. This format lends a hand in dynamically scaling the ontology in accordance with the data that is being received from different sources.

This ontology is stored on a PoW based blockchain model and stored on the database after elliptic curve-based encryption. Internal details of each of these models is described separate sub-sections of this text. Readers can refer these sections implement the proposed model in parts and, depending upon the requirements.

3.1. Pre-processing layer design

Initially, all input data is given to a pre-processing layer for feature extraction. Here numerical features are directly passed to the feature selection layer, while textual features are given to a word2vec model. This model is depicted in Fig. 3, wherein components like context builder, vocabulary builder, continuous bag of words (CBoW) engine, etc. are defined.

A large number of context-sensitive vocabulary models are available for this purpose. In this work, we are using Bidirectional Encoder Representations from Transformers (BERT) model, because of its coverage extensiveness, and reduced dependency on external sources. This block creates a vocabulary from input data and provides it to the context builder block. This block generates word pairs and finds neighbourhood combinations from these pairs. These combinations are counted and given to a numerical layer for evaluating number of occurrences for each pair. These occurrence values are considered as initial word2vec features, and are evaluated using backpropagation, hierarchical SoftMax, and negative sampling layers. These layers further reduce feature redundancy by removing non-action words, which assists in feature reduction, and accuracy improvement of the proposed feature extraction model. The extracted features are processed using a 2-layered neural network, where each layer maps input word pairs with respective features. The result of this model is a single context-sensitive feature vector for the entire sentence. This feature vector is given to a feature selection layer, where variance-based features are extracted.

3.2. Feature selection layer design

After features are extracted using Word2Vec layer, they are given to a variance-based selection layer. This layer uses combination of linear SVM, and Extra Trees (ET) classifier for removal of non-variant features from input dataset. Both these classifiers are used in their standard form and are given per-feature intra-variance value for training and validation. This value is extracted using Eq. 1 as follows,

$$F_{int} = \sqrt{\frac{\sum_{a=1}^{m}(f_a - \frac{\sum_{i=1}^{m}\sqrt{\frac{\sum_{j=1}^{n}(f_j - \frac{\sum_{k=1}^{n} f_k}{n}^{2})}{n-1}}^{2}}{m})}{m-1}} \ldots(1)$$

Where, $F_{int}, m, and n$ represents inter-variance feature value for feature f, total number of features of current type, and total number of other features available in the dataset. This inter-variance feature value is an indicative of variance-level of this feature w.r.t. all other features in the dataset. This value is given to both ET and SVM classifiers for estimation of feature redundancy. Parameters for both these classifiers, along with reason for selection is depicted in Table 1 as follows,

Table 1
Parametric values for each classifier

Classifier	Feature	Reason
SVM	Kernel: Linear Tolerance = 0.1% Decision type = One v/s Rest	A linear regression kernel is used for evaluation of feature-to-feature variance and is accompanied with low error tolerance between current features, and rest of the evaluated features.
Extra Trees	Number of estimators = 10 * Number of features Entropy criterion = Gini Impurity Random State = random (0, Number of features) Class weights = Intra variance of all features	The extra trees classifier is trained w.r.t. number of available features and is allowed to shift to any feature vector for variance checking. Gini impurity is used for one-to-one mapping, while class weights are initialized with intra-variance between different features. Due to which, the model can reduce dependency on default tree weights, and estimate redundancy with better efficiency

Both these classification engines output their own set of features. A union of these features is used as the final feature vector, and can be obtained using Eq. 2 as follows,

$$F_{out} = \bigcup F_{out}\left(SVM\right), F_{out}\left(ET\right)\ldots(2)$$

Both these feature outputs are given to an ontology generation engine, which is described in the next sub-section of this text.

3.3. Ontology generation process

Upon feature selection, only the feature vectors which are non-redundant are extracted from input dataset. These features are given to a correlation engine, which evaluates relationships between them. Correlation value of each feature w.r.t. other features ($Corr_{F_1F_2}$) is extracted using the Eq. 3,

$$Corr_{F_1F_2} = \frac{\sum_{i=1}^{N_f} F_{1i_{int}} - F_{2i_{int}}}{\sqrt{\sum_{i=1}^{N_p} (F_{1i_{int}} - F_{2i_{int}})}^2} \ldots (3)$$

Where, $F_{1i_{int}}, andF_{2i_{int}}$ represents intermediate feature values for the compared features, and is estimated using Eq. 1, while N_f represents number of extracted features for the given comparison. Correlation values for each feature pair is extracted, and their average correlation is evaluated using Eq. 4,

$$AVG_{corr} = \frac{\sum_{i=1}^{N_f} \sum_{j=1}^{N_f} Corr_{i,j}}{N^2} \ldots (4)$$

All features with correlation more than AVG_{corr} are clubbed together, while other features are stored in a separate group. These groups are combined with entity information, and an output RDF ontology is created using the format described in Table 2 as follows,

Table 2
RDF format used for ontology formation

Entity	Class	Grouped Features	Ungrouped Features	Time Stamp

In this format, the entity is an application dependent entry, which can be 'disease type' for medical applications, 'product type' for E-commerce applications, etc. Class represents the category of this feature, while grouped and ungrouped features represent similar & dissimilar feature values, and Timestamp indicates the time at which this entry was generated. The RDF data is given to a blockchain based model for improving security, which is described in the next sub-section.

3.4. Blockchain based security model for storage

The RDF data is stored using a blockchain model, which ensures immutability, traceability, distributed processing, and improved trust-levels. In order to store the data into blockchain, a chain similar to the one described in Fig. 4 is formed, and the following operations are performed,

- Every time a new entry is added to the RDF, a new block is created
- The following information is added to the block,
    - Source of input data
    - Timestamp at which this data arrived in the system
    - A random nonce value, which is used to form uniquely identifiable hashes for each block
    - RDF data generated from section 3.1, 3.2, and 3.3
    - Hash value of previous block (this value is blank for the Genesis block)
- Each block is encrypted using elliptic curve cryptography (ECC) model for improved security.

During addition of a new block, a random nonce number is generated for hash calculation. After generation of this nonce value, the following Eq. 5 is evaluated,

$$Hash = SHA256(Source, Timestamp,$$

$$RDFdata, Nonce) \ldots (5)$$

If this hash value is already present in any of the blocks, then a new random nonce is generated again, otherwise it is used for blockchain creation. The block is also encrypted using ECC, where the following encryption curve is used,

$$y^2 = x^3 + 5x + 4 \ldots (6)$$

Here, curve constants are selected based on multiple evaluations of the model and observing delay for each curve type. This is a standard secp256 curve and is proven to have high encryption efficiency. This curve can be observed from Fig. 5, wherein its nature is visualized. Using this curve, and standard ECC model, each block was encrypted before storage. Due to which, the model is observed to be highly secure, and possess lower delay, and better network & representation efficiency when compared with existing approaches. This evaluation is done in the next section of this text, wherein the proposed BOGMAS model is compared with BOG [3], MAOM [8], and BTBP [11] models, which have similar representation capabilities.

4. Result Analysis And Comparison

The proposed BOGMAS model was evaluated on a wide variety of datasets including, heart disease (http://www.informatics.jax.org/disease/DOID:114), blood reports (https://bioportal.bioontology.org/ontologies), EEG (https://maayanlab.cloud/Harmonizome/gene_set/eeg+abnormality/GWASdb+SNP-Phenotype+Associations), ecommerce (https://www.sciencedirect.com/science/article/pii/S2352340922000968), social media (https://ieee-dataport.org/documents/ontosnaqa-multi-domain-ontology-social-network-analysis), and news datasets (https://www.v7labs.com/open-datasets/visual-sentiment-ontology). These datasets were combined together to form a large-dataset, and then given to the model for RDF generation. A total of 100k values were evaluated on the dataset, evaluation was done in terms of precision of representation, recall of representation, delay needed for representation, and memory size required for representation. Apart from this, the RDF was mutated after formation using Sybil, & Masquerading attacks, but due to incorporation of blockchain, a very few numbers of blocks were mutated. Based on this strategy, accuracy of attack detection was also evaluated for these models. Upon validation, all affected blocks were restored, thereby confirming high-trust ability of the blockchain-based storage model. In order to evaluate the performance w.r.t. standard models, this performance was compared in terms of accuracy (A), precision (P), recall (R), and delay levels, with BOG [3], MAOM [8], & BTBP [11] approaches for different test set sizes (TSS), and results were tabulated in table 6, 7, 8, 9, and 10, while these parameters were estimated via equations 7, 8, 9 and 10 as follows,

$$A = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \ldots (7)$$

$$P = \frac{t_p}{t_p + f_p} \ldots (8)$$

$$R = \frac{t_p}{t_p + t_n + f_p + f_n} \ldots (9)$$

$$d = t_{complete} - t_{start} \ldots (10)$$

Where, $t\&f$ represents true and false rates for different inputs, while t represents the timestamps for different instances.

From the accuracy values it can be observed that the proposed model is 19% more accurate than BOG [3], 35% more accurate than MAOM [8], and 34% more accurate than BTBP [11], which makes it useful for highly secure network applications. This increase in accuracy is due to use of blockchain for storage, which reduces probability of any network attack. Similar observations are made for precision (P) values, and can be observed from the following Fig. 7,

From the precision values it can be observed that the proposed model is 15% more efficient than BOG [3], 18% better than MAOM [8], and 29% better than BTBP [11], which makes it useful for high precision data representation applications. Similar observations are made for recall (R) values, and can be observed from the following Fig. 8,

From the recall values it can be observed that the proposed model is 16% more efficient than BOG [3], 22% better than MAOM [8], and 31% better than BTBP [11], which makes it useful for high recall data representation applications. The value of recall is very high than existing models due to use of variance-based representation, which makes the model highly efficient in real-time scenarios. Similar observations are made for representation delay, and can be observed from the following Fig. 9,

From the delay values it can be observed that the proposed model is 2% slower than existing implementations. But as the delay difference is not very large, the model is applicable for real-time system design. Similar observations are made for storage cost (SC) values, and can be observed from the following Fig. 10,

From the storage cost values, it can be observed that the proposed model is 30% more efficient than BOG [3], requires 28% lower space than MAOM [8], and 16% lower space than BTBP [11], which makes it useful for applications with limited memory requirements. Due to which, the model is capable of lowering storage costs, while improving accuracy & precision of ontological representations. Because of lower storage costs the model is also capable of reducing blockchain-based deployment costs, thus improving scalability. The model was evaluated on Proof-of-Stake (PoS) based blockchain that uses smart contracts but can be used with any other blockchain type with minimum reconfiguration operations. These advantages are due to use of variance-based representation, and because of reduced redundancy in the output RDF data representation, which makes the model highly effective for real-time applications.

5. Conclusion And Future Work

Due to utilization of variance-based approach for feature selection, and blockchain for security improvement, the proposed model can reduce storage cost, and improve efficiency of data representation. It is observed that the proposed model is 15% more efficient than BOG [3], 18% better than MAOM [8], and 29% better than BTBP [11] in terms of precision of data representation, while it is 16% more efficient than BOG [3], 22% better than MAOM [8], and 31% better than BTBP [11], which makes it useful for high recall data representation applications. Similarly, in terms of storage cost the proposed model is 30% more efficient than BOG [3], requires 28% lower space than MAOM [8], and 16% lower space than BTBP [11], which is due to use of enhanced feature selection capabilities of the system. Security of the proposed model is also very high, which is due to use of blockchain for securely storing ontological data. It is observed that the model is 19% more accurate than BOG [3], 35% more accurate than MAOM [8], and 34% more accurate than BTBP [11], against Masquerading & Sybil attacks, which makes it useful for highly secure network applications. The model has been observed to have a higher delay in comparison with compared models, which is due to use of blockchain. This delay can be reduced via extension of blockchains using various sharing models, which needs to be researched. Furthermore, in future researchers can also aim at developing recommendation models based on the proposed approach and estimate it performance on different application scenarios.

Declarations

Ethical Approval

This research did not contain any studies involving animal or human participants, nor did it take place on any private or protected areas.

Competing Interests

No funding was received to assist with the preparation of this manuscript.

Authors' contributions

Conceptualization, Suniti Purbey; methodology, Suniti Purbey; software, Dr. Brijesh Khandelwal; validation, Suniti Purbey, Dr. Brijesh Khandelwal, and Ashutosh Kumar Choudhary; formal analysis, Suniti Purbey; investigation, Suniti Purbey; resources, Suniti Purbey; data curation, Dr. Brijesh Khandelwal;

writing—original draft preparation, Suniti purbey; writing—review and editing, Dr. Brijesh Khandelwal; visualization, Ashutosh Kumar Choudhary; supervision, Dr. Brijesh Khandelwal.

Funding

There is no funding related to this paper.

Availability of data and materials

Data will be made available on request.

References

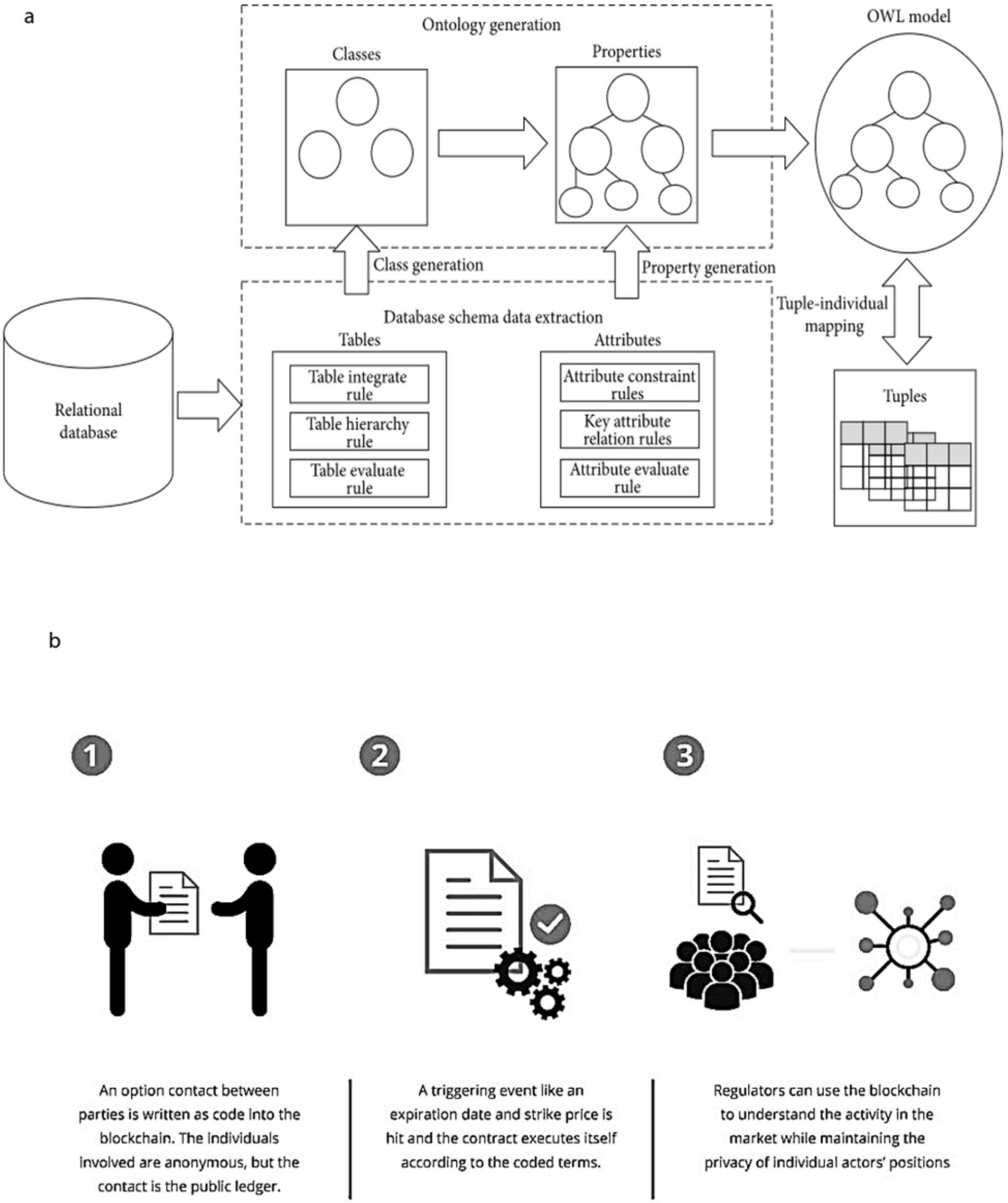
1. S. Hashemikhabir, R. Xia, Y. Xiang and S. C. Janga, "A Framework for Identifying Genotypic Information from Clinical Records: Exploiting Integrated Ontology Structures to Transfer Annotations between ICD Codes and Gene Ontologies," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 4, pp. 1259-1269, 1 July-Aug. 2018, doi: 10.1109/TCBB.2015.2480056.
2. O. Sadio, I. Ngom and C. Lishou, "A Novel Sensing as a Service Model Based on SSN Ontology and Android Automotive," in IEEE Sensors Journal, vol. 19, no. 16, pp. 7015-7026, 15 Aug.15, 2019, doi: 10.1109/JSEN.2019.2911913.
3. R. Kumaravel, S. Selvaraj and C. Mala, "A Multidomain Layered Approach in Development of Industrial Ontology to Support Domain Identification for Unstructured Text," in IEEE Transactions on Industrial Informatics, vol. 14, no. 9, pp. 4033-4044, Sept. 2018, doi: 10.1109/TII.2018.2835567.
4. J. Liu, Z. Qu, M. Yang, J. Sun, S. Su and L. Zhang, "Jointly Integrating VCF-Based Variants and OWL-Based Biomedical Ontologies in MongoDB," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 17, no. 5, pp. 1504-1515, 1 Sept.-Oct. 2020, doi: 10.1109/TCBB.2019.2951137.
5. C. Yang, V. Dubinin and V. Vyatkin, "Automatic Generation of Control Flow From Requirements for Distributed Smart Grid Automation Control," in IEEE Transactions on Industrial Informatics, vol. 16, no. 1, pp. 403-413, Jan. 2020, doi: 10.1109/TII.2019.2930772.
6. T. Chakraborty, S. Jajodia, J. Katz, A. Picariello, G. Sperli and V. S. Subrahmanian, "A Fake Online Repository Generation Engine for Cyber Deception," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 2, pp. 518-533, 1 March-April 2021, doi: 10.1109/TDSC.2019.2898661.
7. N. Fathy, W. Gad, N. Badr and M. Hashem, "ProGOMap: Automatic Generation of Mappings From Property Graphs to Ontologies," in IEEE Access, vol. 9, pp. 113100-113116, 2021, doi: 10.1109/ACCESS.2021.3104293.
8. A. Smirnov, T. Levashova, A. Ponomarev and N. Shilov, "Methodology for Multi-Aspect Ontology Development: Ontology for Decision Support Based on Human-Machine Collective Intelligence," in IEEE Access, vol. 9, pp. 135167-135185, 2021, doi: 10.1109/ACCESS.2021.3116870.

9. X. Xue and J. Lu, "A Compact Brain Storm Algorithm for Matching Ontologies," in IEEE Access, vol. 8, pp. 43898-43907, 2020, doi: 10.1109/ACCESS.2020.2977763.
10. M. N. Demaidi, M. M. Gaber and N. Filer, "OntoPeFeGe: Ontology-Based Personalized Feedback Generator," in IEEE Access, vol. 6, pp. 31644-31664, 2018, doi: 10.1109/ACCESS.2018.2846398.
11. M. Hardini, Q. Aini, U. Rahardja, R. D. Izzaty and A. Faturahman, "Ontology of Education Using Blockchain: Time Based Protocol," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020, pp. 1-5, doi: 10.1109/ICORIS50180.2020.9320807.
12. W. M. W. Isa et al., "An Ontological Approach for Creating a Brassware Craft Knowledge Base," in IEEE Access, vol. 8, pp. 163434-163446, 2020, doi: 10.1109/ACCESS.2020.3022795.
13. Y. Huang, Y. Chai, Y. Liu and J. Shen, "Architecture of next-generation e-commerce platform," in Tsinghua Science and Technology, vol. 24, no. 1, pp. 18-29, Feb. 2019, doi: 10.26599/TST.2018.9010067.
14. M. Iqbal and R. Matulevičius, "Exploring Sybil and Double-Spending Risks in Blockchain Systems," in IEEE Access, vol. 9, pp. 76153-76177, 2021, doi: 10.1109/ACCESS.2021.3081998.
15. I. Huitzil, Á. Fuentemilla and F. Bobillo, "I Can Get Some Satisfaction: Fuzzy Ontologies for Partial Agreements in Blockchain Smart Contracts," 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2020, pp. 1-8, doi: 10.1109/FUZZ48607.2020.9177732.
16. P. Kudumakis, T. Wilmering, M. Sandler, V. Rodriguez-Doncel, L. Boch and J. Delgado, "The Challenge: From MPEG Intellectual Property Rights Ontologies to Smart Contracts and Blockchains [Standards in a Nutshell]," in IEEE Signal Processing Magazine, vol. 37, no. 2, pp. 89-95, March 2020, doi: 10.1109/MSP.2019.2955207.
17. Y. Shen et al., "Gastroenterology Ontology Construction Using Synonym Identification and Relation Extraction," in IEEE Access, vol. 6, pp. 52095-52104, 2018, doi: 10.1109/ACCESS.2018.2862885.
18. Y. Shen et al., "IDDAT: An Ontology-Driven Decision Support System for Infectious Disease Diagnosis and Therapy," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 1417-1422, doi: 10.1109/ICDMW.2018.00201.
19. O. Choudhury, N. Rudolph, I. Sylla, N. Fairoza and A. Das, "Auto-Generation of Smart Contracts from Domain-Specific Ontologies and Semantic Rules," 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2018, pp. 963-970, doi: 10.1109/Cybermatics_2018.2018.00183.
20. G. -W. Kim and D. -H. Lee, "Intelligent Health Diagnosis Technique Exploiting Automatic Ontology Generation and Web-Based Personal Health Record Services," in IEEE Access, vol. 7, pp. 9419-9444, 2019, doi: 10.1109/ACCESS.2019.2891710.
21. M. Y. Khan, M. Ali, S. Qaisar, M. Naeem, C. Chrysostomou and M. Iqbal, "Placement Optimization for Renewable Energy Sources: Ontology, Tools, and Wake Models," in IEEE Access, vol. 8, pp. 72781-72800, 2020, doi: 10.1109/ACCESS.2020.2984901.

22. Abad-Navarro, Francisco & Martínez-Costa, Catalina & Fernandez-Breis, Jesualdo. (2021). Semankey: A Semantics-Driven Approach for Querying RDF Repositories Using Keywords. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3091413.
23. F. Škopljanac-Mačina, I. Zakarija and B. Blašković, "Towards Automated Assessment Generation in e-Learning Systems Using Combinatorial Testing and Formal Concept Analysis," in IEEE Access, vol. 9, pp. 52957-52976, 2021, doi: 10.1109/ACCESS.2021.3070510.
24. F. Alsubaei, A. Abuhussein and S. Shiva, "Ontology-Based Security Recommendation for the Internet of Medical Things," in IEEE Access, vol. 7, pp. 48948-48960, 2019, doi: 10.1109/ACCESS.2019.2910087.
25. Ali A, Almaiah MA, Hajjej F, Pasha MF, Fang OH, Khan R, Teo J, Zakarya M. An Industrial IoT-Based Blockchain-Enabled Secure Searchable Encryption Approach for Healthcare Systems Using Neural Network. *Sensors*. 2022; 22(2):572. https://doi.org/10.3390/s22020572
26. Ali A, Rahim HA, Pasha MF, Dowsley R, Masud M, Ali J, Baz M. Security, Privacy, and Reliability in Digital Healthcare Systems Using Blockchain. *Electronics*. 2021; 10(16):2034. https://doi.org/10.3390/electronics10162034
27. Almaiah MA, Ali A, Hajjej F, Pasha MF, Alohali MA. A Lightweight Hybrid Deep Learning Privacy Preserving Model for FC-Based Industrial Internet of Medical Things. *Sensors*. 2022; 22(6):2112. https://doi.org/10.3390/s22062112
28. Almaiah MA, Hajjej F, Ali A, Pasha MF, Almomani O. A Novel Hybrid Trustworthy Decentralized Authentication and Data Preservation Model for Digital Healthcare IoT Based CPS. *Sensors*. 2022; 22(4):1448. https://doi.org/10.3390/s22041448
29. E. Wang et al., "Trustworthy and Efficient Crowdsensed Data Trading on Sharding Blockchain," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3547-3561, Dec. 2022, doi: 10.1109/JSAC.2022.3213331.
30. S. Zou, J. Xi, G. Xu, M. Zhang and Y. Lu, "CrowdHB: A Decentralized Location Privacy-Preserving Crowdsensing System Based on a Hybrid Blockchain Network," in IEEE Internet of Things Journal, vol. 9, no. 16, pp. 14803-14817, 15 Aug.15, 2022, doi: 10.1109/JIOT.2021.3084937.
31. C. Zhang, L. Zhu, C. Xu and K. Sharif, "PRVB: Achieving Privacy-Preserving and Reliable Vehicular Crowdsensing via Blockchain Oracle," in IEEE Transactions on Vehicular Technology, vol. 70, no. 1, pp. 831-843, Jan. 2021, doi: 10.1109/TVT.2020.3046027.
32. C. Cai, Y. Zheng, Y. Du, Z. Qin and C. Wang, "Towards Private, Robust, and Verifiable Crowdsensing Systems via Public Blockchains," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 4, pp. 1893-1907, 1 July-Aug. 2021, doi: 10.1109/TDSC.2019.2941481.
33. R. Zhao, L. T. Yang, D. Liu, X. Deng and Y. Mo, "A Tensor-Based Truthful Incentive Mechanism for Blockchain-Enabled Space-Air-Ground Integrated Vehicular Crowdsensing," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 2853-2862, March 2022, doi: 10.1109/TITS.2022.3144301.

34. J. An, Z. Wang, X. He, X. Gui, J. Cheng and R. Gui, "PPQC: A Blockchain-Based Privacy-Preserving Quality Control Mechanism in Crowdsensing Applications," in IEEE/ACM Transactions on Networking, vol. 30, no. 3, pp. 1352-1367, June 2022, doi: 10.1109/TNET.2022.3141582.
35. Y. Hui et al., "BCC: Blockchain-Based Collaborative Crowdsensing in Autonomous Vehicular Networks," in IEEE Internet of Things Journal, vol. 9, no. 6, pp. 4518-4532, 15 March15, 2022, doi: 10.1109/JIOT.2021.3105547.

Figures



An option contact between parties is written as code into the blockchain. The individuals involved are anonymous, but the contact is the public ledger.

A triggering event like an expiration date and strike price is hit and the contract executes itself according to the coded terms.

Regulators can use the blockchain to understand the activity in the market while maintaining the privacy of individual actors' positions

Figure 1

(a). Ontology generation model for relational database

(b). Working of smart contracts

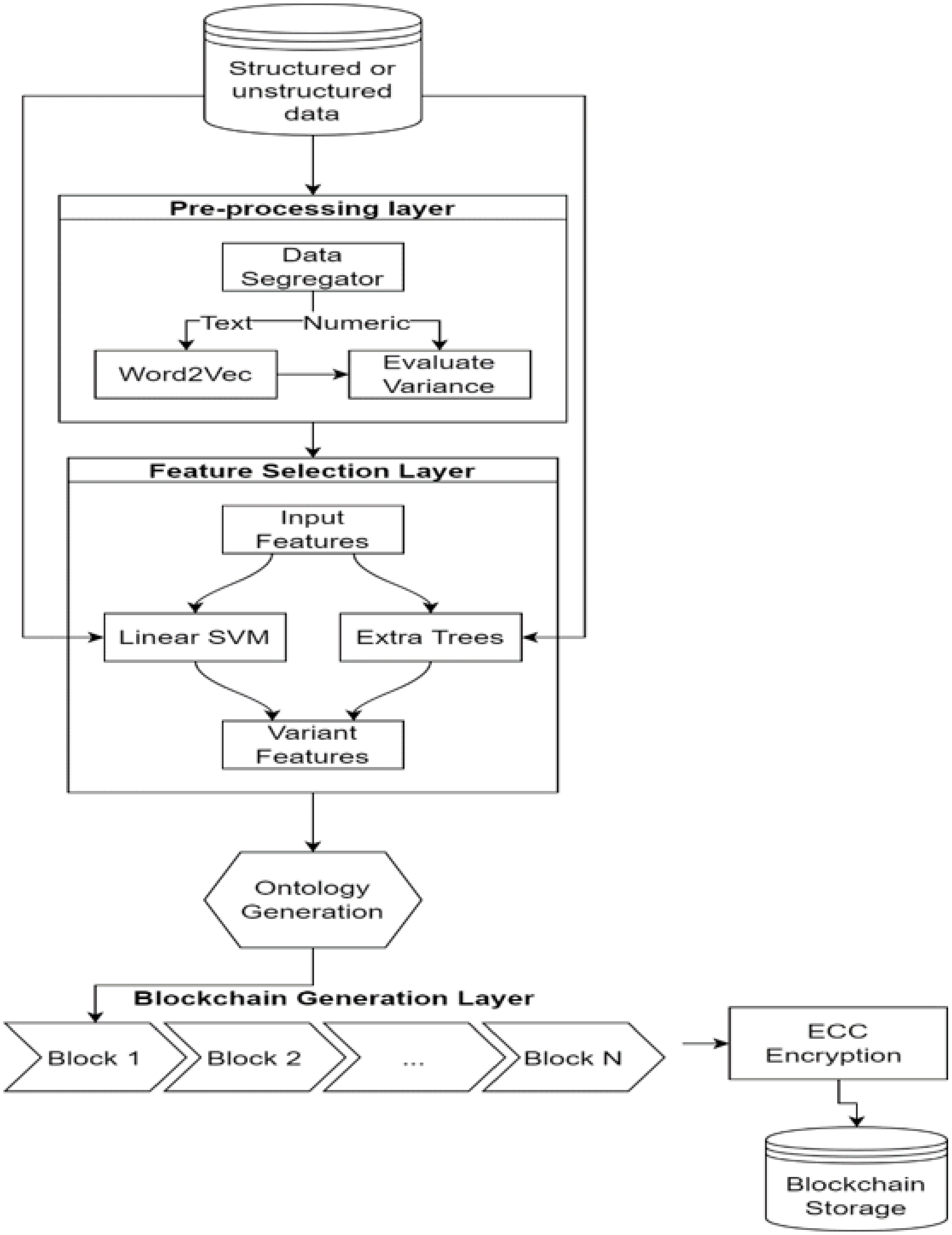


Figure 2

Model for the proposed blockchain-based ontology generation framework

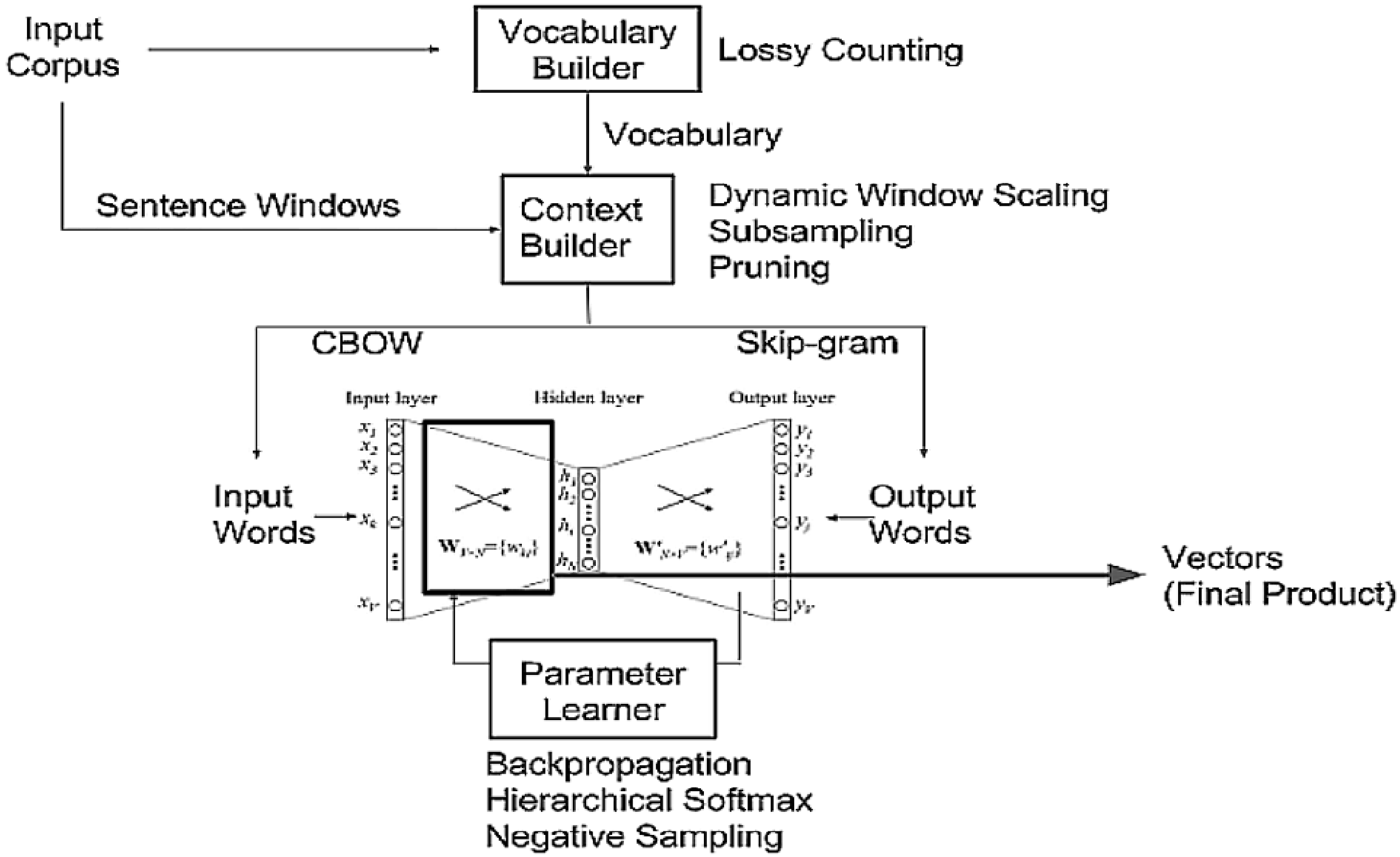


Figure 3

Word2Vec internal working

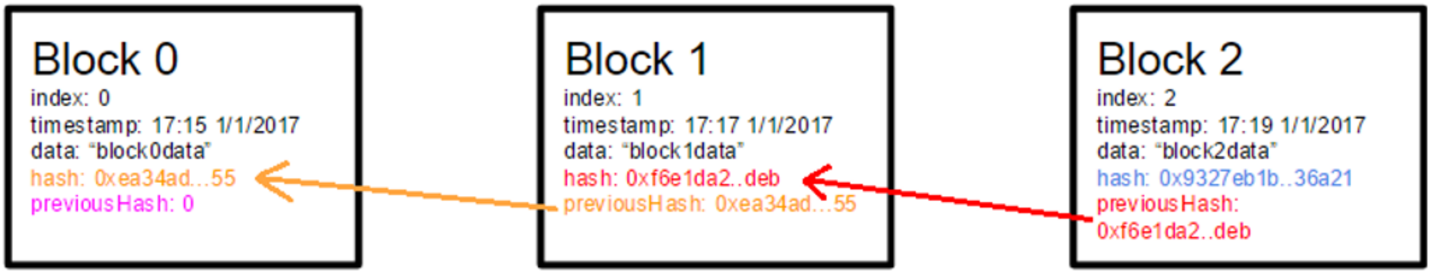


Figure 4

The blockchain structure

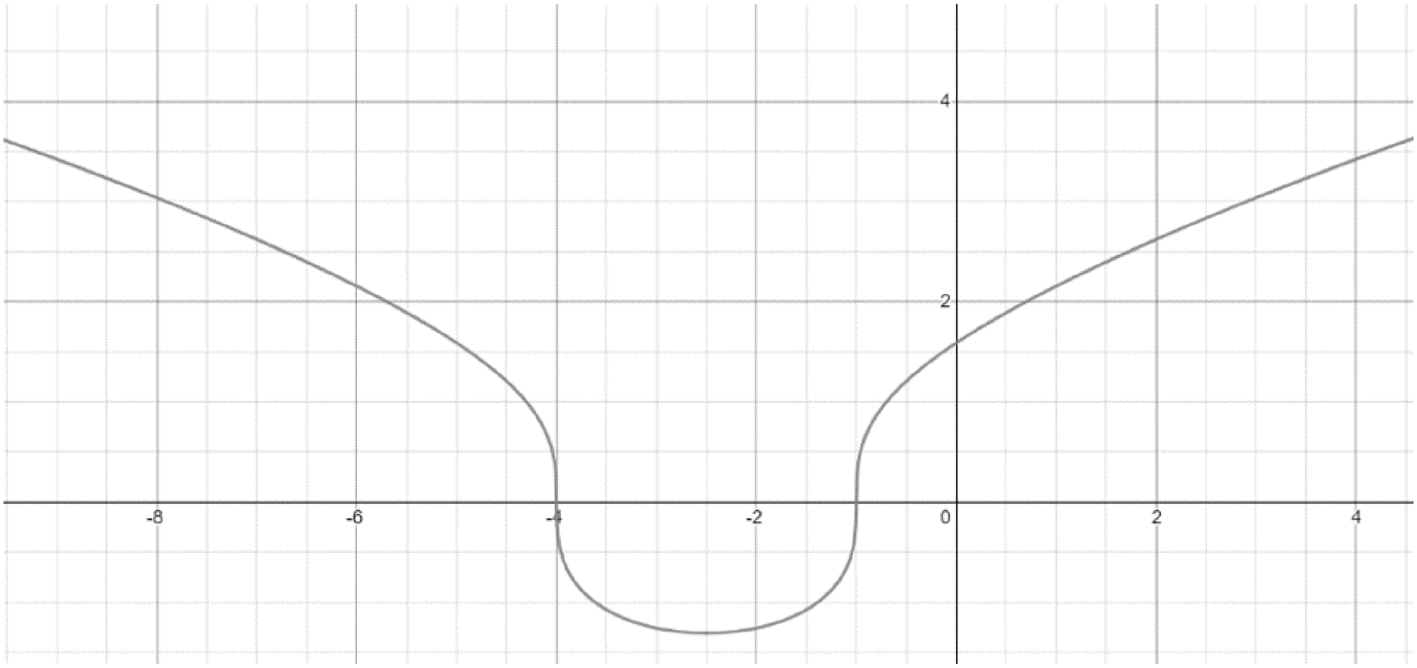


Figure 5

The ECC curve used for block encryption

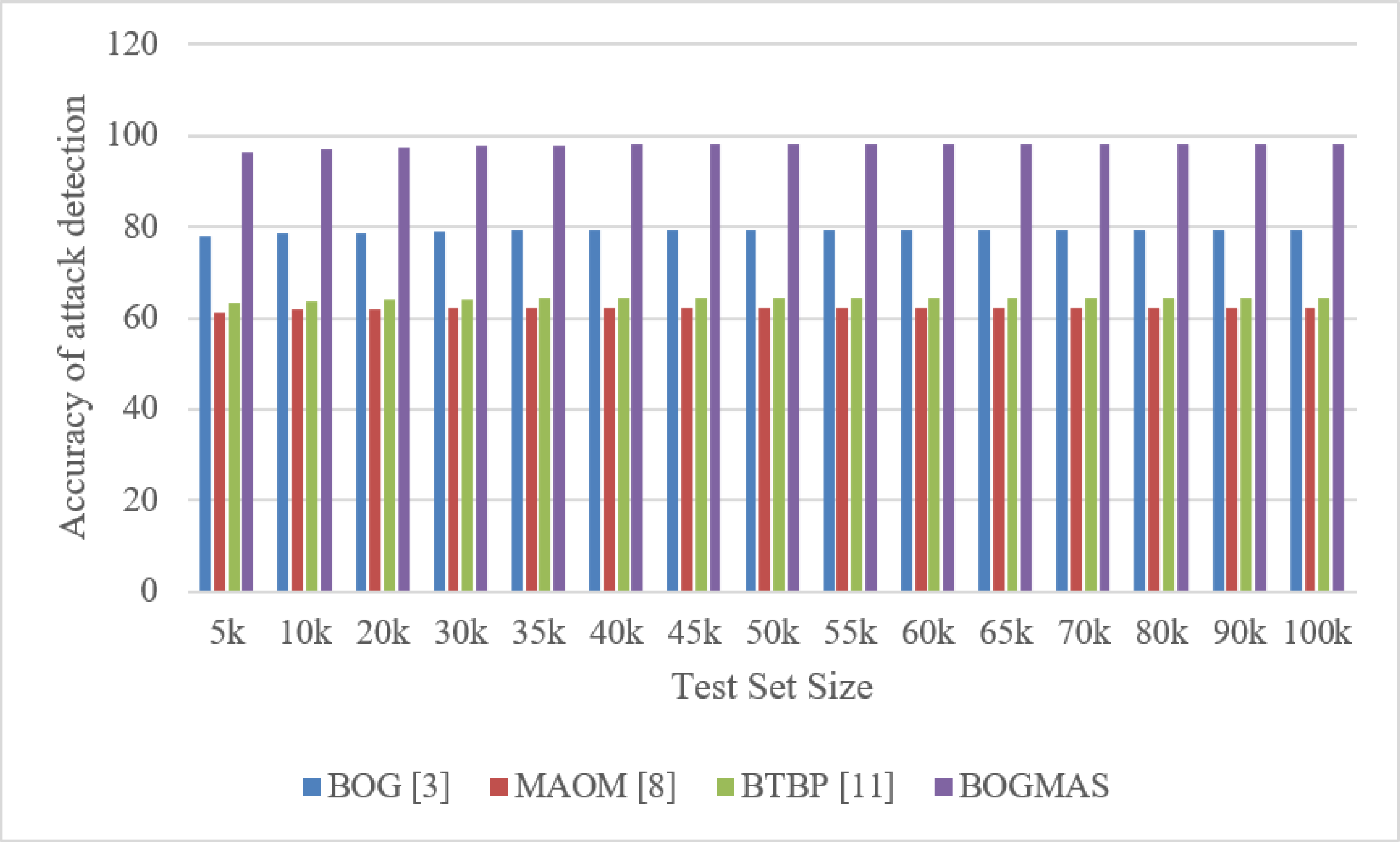


Figure 6

Accuracy of attack detection on different test set sizes

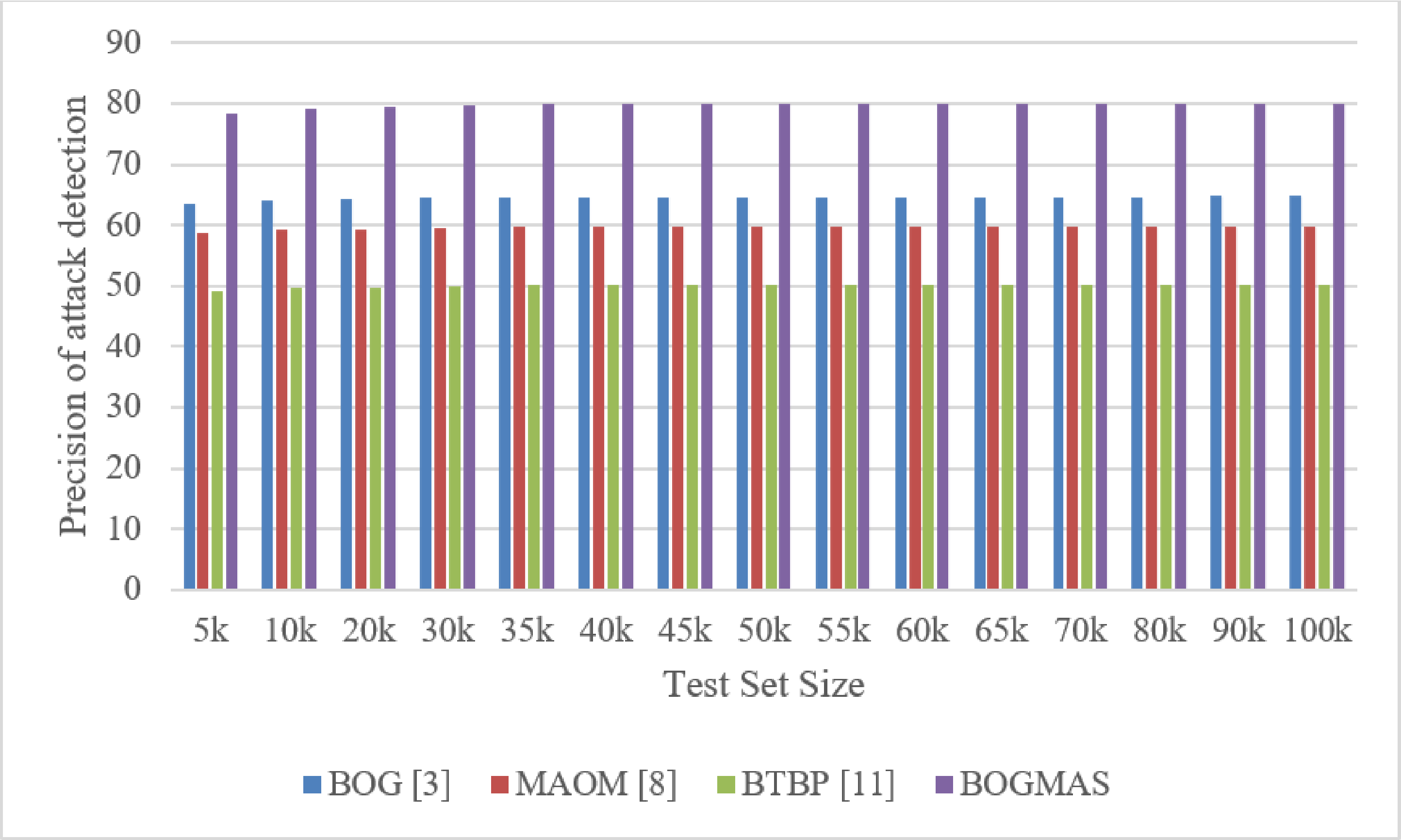


Figure 7

Average precision of representation for different algorithms

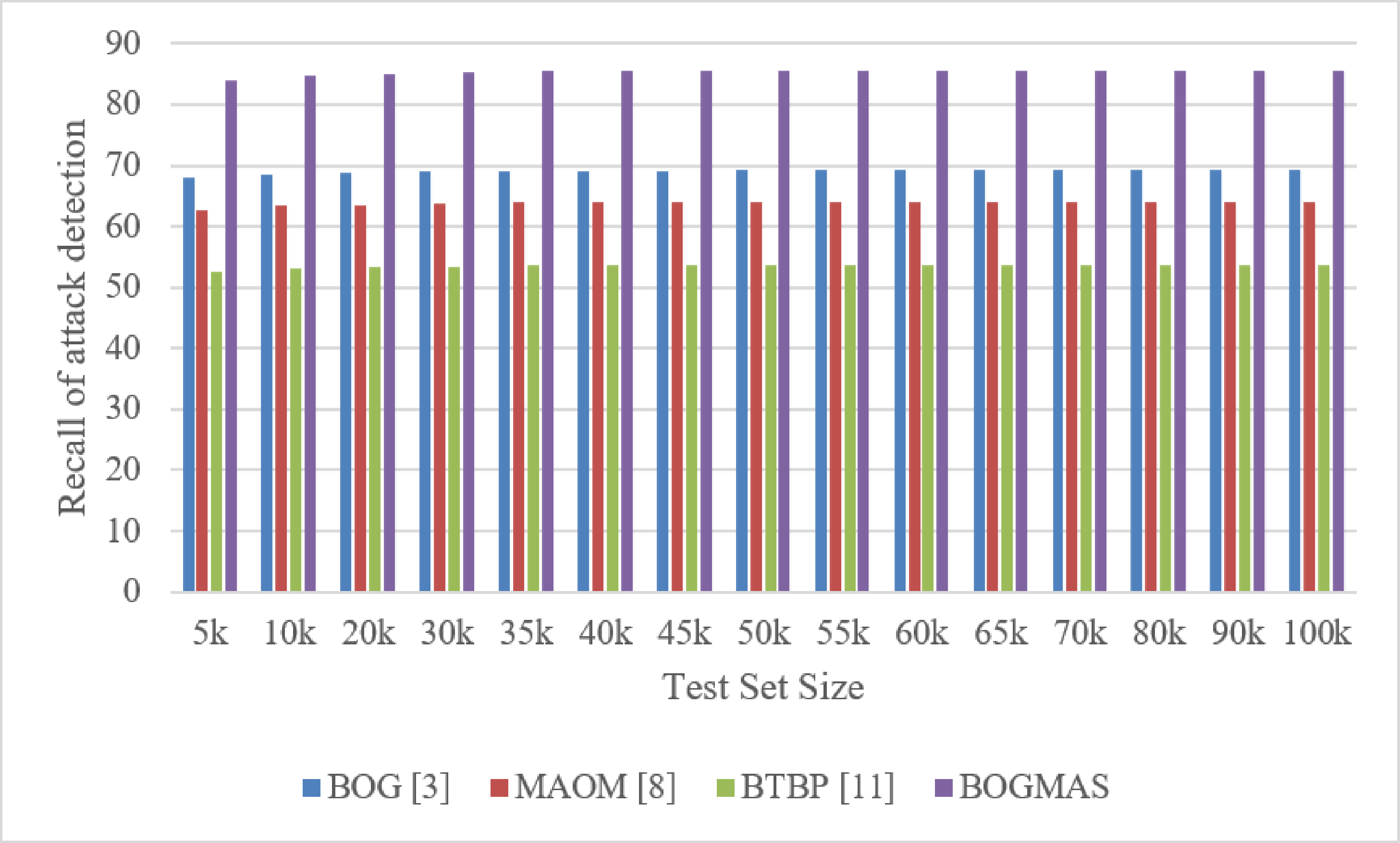


Figure 8

Average recall of representation for different algorithms

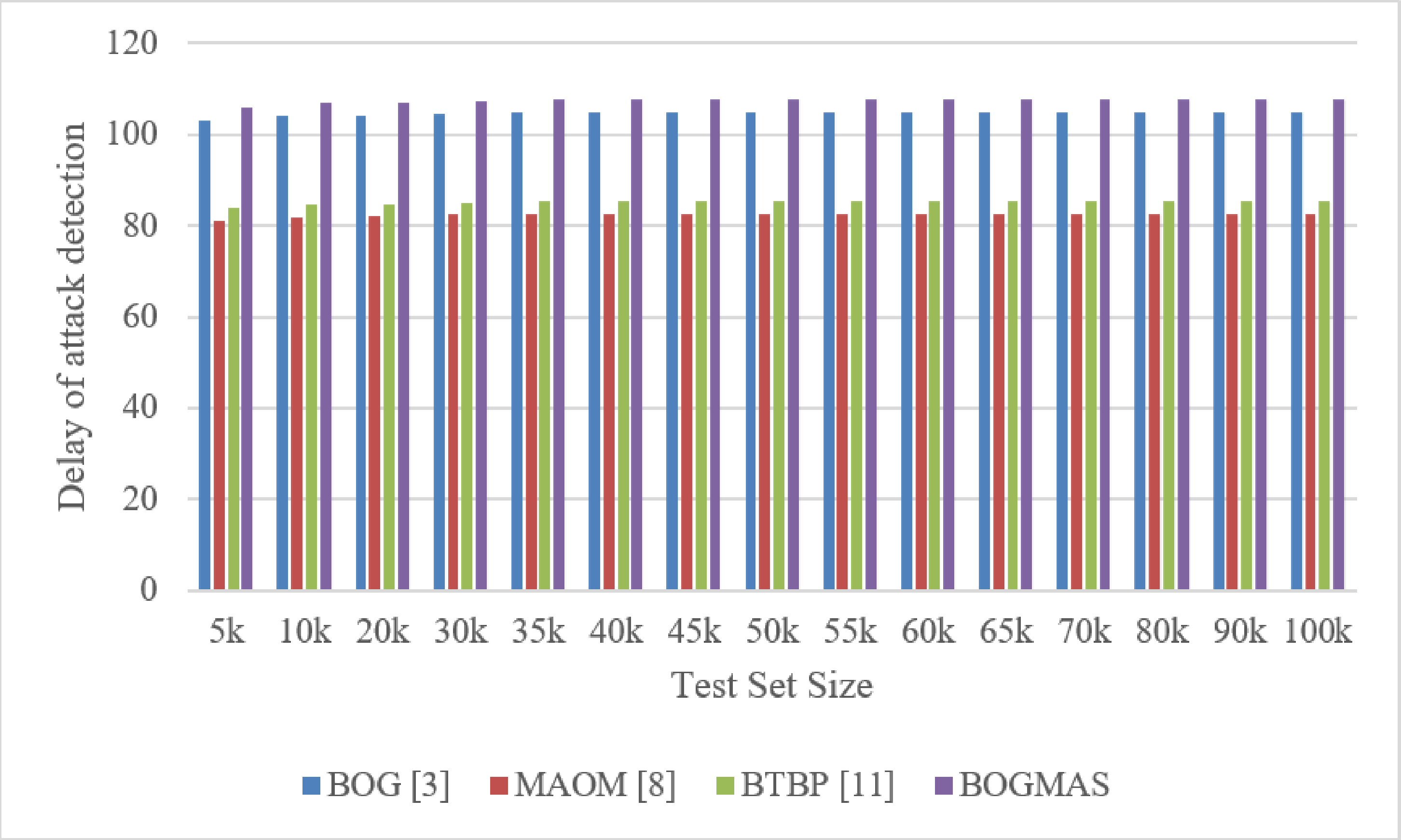


Figure 9

Delay of representation values for different algorithms

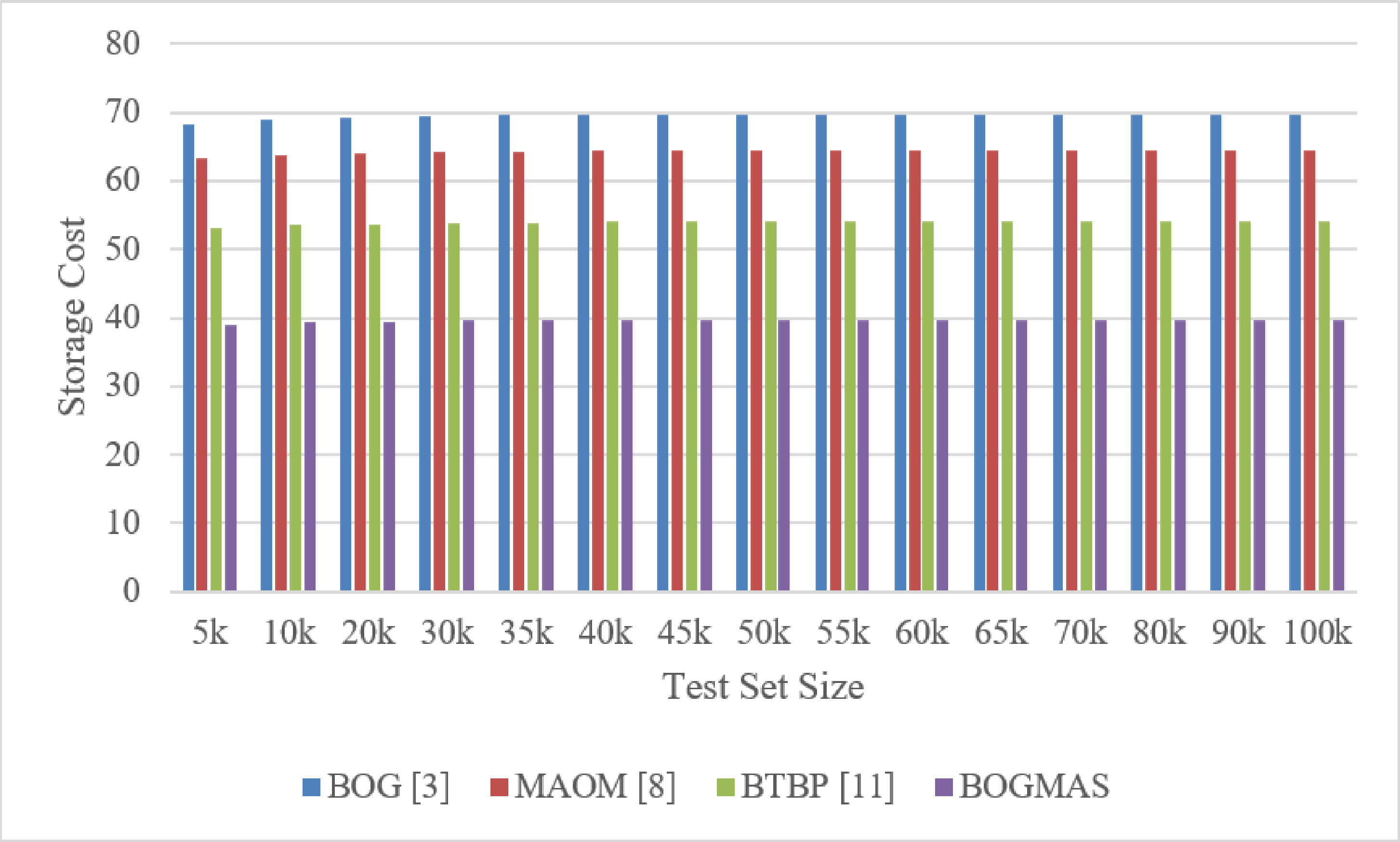


Figure 10

Storage cost for different algorithms