Computers and Electrical Engineering 74 (2019) 511–519

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Improved scalability in mining using ontology record linkage algorithm☆

T. Prabhu[a,*], C. Suresh Gnana Dhas[b]

[a] Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Thirunelveli, Tamilnadu 627012, India. E-Mail: prabhu.cse@drmgrdu.ac.in

[b] Department of Computer Science and Engineering, Vivekanadha College of Engineering for Women, Tiruchengode, Tamilnadu 637205, India.E-Mail: sureshc.me@gmail.com

ARTICLE INFO

Article history:
Received 17 April 2017
Revised 18 January 2018
Accepted 18 January 2018
Available online 21 February 2018

Keywords:
Record linkage
Data mining
Angle based neighborhood
Ontology
Conventional method

ABSTRACT

Record linkage offers wide role in record identification and relevant datasets matching. The conventional researchers use probabilistic approach to identify reliable and unique datasets. Record linkage with probabilistic approach exploits data, which are common to an individual record pair. Classical methods have equality based record linkage in common fields. Therefore, errors associated with record linkage reduce the scalability. In this paper, a similarity between individual values of record pairs is improved using ontology-based semantic similarity model. Semantic similarity between the records is tested successfully using angle based neighborhood graph. To validate the proposed approach, a conventional record linkage algorithm is compared with angle based neighborhood ontology record linkage technique, which achieves improved accuracy and scalability. Finally, the accuracy of identifying similar semantic matches is more scalable in proposed technique than conventional methods.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Record linkage is a technique for identifying the records of similar entities in different databases. Initially, it is used in data integration, data cleaning and duplicate entry detection. As a result, several centralization procedures is used to link the similar data from the different database, as a common entity [1]. The process of record linkage becomes more powerful while identifying the combined variables, partially and hence unique patterns are generated. Usually, two strategies are applied to find record linkage between documents, which includes: deterministic and probabilistic record linkage. In deterministic linkage, predefined subset of linking variables should agree with link pairs, whereas, probabilistic linkage is processed using agreement or disagreement weights between variable. This helps to identify the probability difference among non-matched and matched samples. Finally, a link is created when entire weights between the records are higher than threshold level [2].

Past data linkage techniques includes: Abril et al. (2011) used distance-based record linkage method to find distance function between the records. It is used as a weight aggregation of distance function between single attribute values of each record. A supervised model is used to determine the parameters for measuring optimal distance between the records.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. V. Suma.

* Corresponding author.
E-mail address: prabhu.msu@rediffmail.com (T. Prabhu).

https://doi.org/10.1016/j.compeleceng.2018.01.026
0045-7906/© 2018 Elsevier Ltd. All rights reserved.

It helps to find an utmost linkage between original and protected data. Key variables are then identified using variable weight of individual record linkages [1]. Tromp et al. (2011) utilized probabilistic record linkage procedure to identify the agreement and disagreement weights between matched and non-matched samples from identified datasets. The technique is carried out using measurements between registration error and discriminating power [2]. Vatsalan et al. (2013) linked two different databases using indexing process. An overview of various techniques is used, which links databases of organizations. It prevents unwanted intrusions to affect data privacy. This research addresses the concerns on privacy preserving record linkage [4]. Ong et al. (2014) proposed a triple layered technique that combines weight, distance and linkage concepts to improve the performance of record linkage with missing data. This method used fine grained record linkage software for extending the work of Fellegi–Sunter scoring implementations. With this, three such models are developed that provides a better solution to missed data grievances at the time of record linkage. Three models include weight redistribution, distance imputation, and linkage expansion. Weight Redistribution removes missing data fields from quasi-identifier set and redistribute the weight from missing attribute based on relative proportions across remaining available linkage fields. Distance imputation imputes distance between missing data fields rather than imputing missing data value. Linkage Expansion adds previously considered non-linkage fields to link field set to compensate missing information in linkage field. [5]. Present work considers this method to improve the performance of record linkage by combining both weight and probabilistic based technique. Randall et al. (2014) concentrated on privacy preservation for linking record in real world datasets. This method uses record linkage strategy, which reduces the risk of record privacy on large real world business datasets. Here, bloom filters are used to encrypt personal identification information based on probabilistic record linkage. However, no improvements in encryption are seen after the application of probabilistic record linkage. Hence, the present study considers probabilistic record linkage to measure data linkage performance [6]. However, probabilistic technique is well concentrated to improve the system performance. On contrary, Herranz et al. [8] used global distance estimation to record linkage in large database. Likewise, global distance estimation is estimated by probabilistic record linkage [11]. Further, in [9], probabilistic and distance linking strategies is studied to improve record linkage strategy. Here, ninety-six scenarios are created, which represent real-life situations using non-unique identifiers. Discriminative power, missing and errorrate and file size is systematically introduced to increase linkage patterns and difficulties. In [3], probabilistic and deterministic record linkage is used over anonymized datasets. In [12], k-NNs is used in same record dataset with missing value and their distances to record similarity imputation. In [13], an unsupervised record linkage with ensemble learning and enhanced automatic self-learning is used to improve record linkage. Also, supervised learning with symmetric bilinear form is used to solve optimization problem [15]. In [14], pseudonyms are used for better record linkage that conveys the information related to similarity. Randall et al. (2014) [7] utilized graph based estimation to link records for reducing errors. Effectiveness of graph theory in linking records is studied that identifies the errors in record linkage process. Graph theory technique is applied over two datasets or true sets. A comparison is made to measure the effectiveness of graph theory record linkage. Further, a threshold setting is used to reduce the error in graph theory and it improves the record linkage equality. Finally, a semantic method is used to provide record linkage with Xtensible Access Control Markup Language [16].

These methods stressed the importance of record linkage strategies; however, they failed to use ontology to address the issues. This paper focuses on both distance based estimation and probabilistic record linkage based estimation to improve the effectiveness of proposed work. The link between documents or record is estimated using both distance function and probabilistic measurement [3].

The main contribution of work involves the following: In this study, enhancement in record linkage is introduced using both distance and probabilistic function. The distance function is used as angle based distance similarity measurement and probabilistic function used in Finite Mixture Model (FMM). This provides greater accuracy than normal clustering process. Additionally, ontology concept is introduced to accurately link the best pair in cluster. This is done with parent-child (is-a) relationship. The semantic vector contains parent concept and weights are calculated using distance measurement between pairs. Possible paths are identified by embedding the distance function with ontology for best agreement. Non-agreement pairs are eliminated from database for efficient record linkage. This paper combines both distance and probabilistic records on contrary with conventional works. This attempts to standardize the procedure for improving record linkage on large databases.

The outline of the paper is organized as follows Section 2 contains the proposed technique that combines the clustering and Finite Mixture Model. Section 3 evaluates the present study with conventional techniques and Section 4 concludes the entire work.

2. Proposed model

The present work concentrates on maintaining exact similarity between datasets, which possess binary relations between them. To make the proposed model more realizable, following steps are adopted:

Step 1: Ontology framework is designed to access datasets. Here, datasets of different domains are identified based on reference datasets.

Step 2: Non-binary attributes in dataset is converted into binary values, i.e. it represents the availability of objects using 1's and 0's.

Step 3: Similar group elements are clustered using Fuzzy C- means clustering [10].

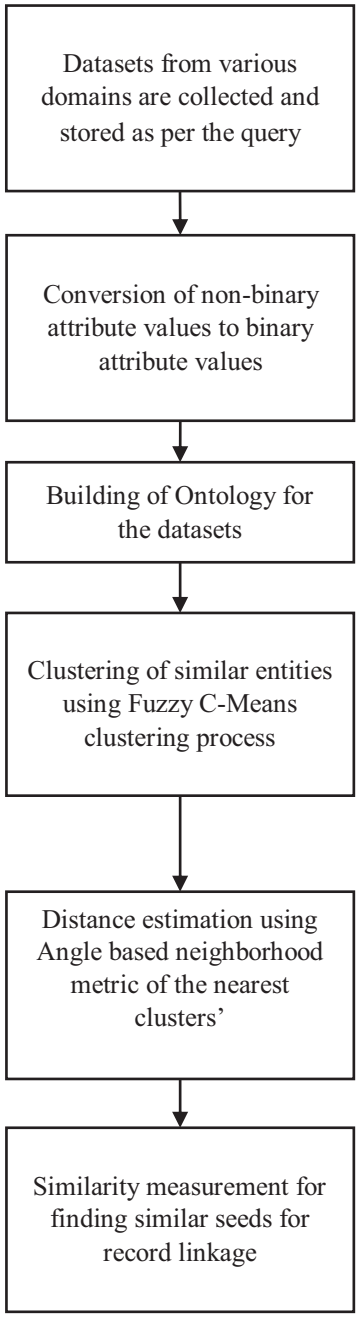


Fig. 1. Proposed technique.

Step 4: Angle based neighborhood distance metric is used to find cluster center and datasets at closest distance with clusters are clustered to reduce squared distance.

Step 5: Once the clusters are formed without misclassified datasets, ontology is built and similarity measurement finds the similar seeds. The steps are shown in Fig. 1.

2.1. Ontology model

The Unified Medical Language System (UMLS) copy is stored in Microsoft SQL Server database that possesses ontology concepts. With the help of tables, root node and seed node is determined. This is similar to finding root for each ontology concept that may have more than single path for root node. Here, each path/route is assigned with unique distance from

Table 1
Angle based ontology measurement for dataset.

	Shortest distance	Weight matrix
NCBI GenBank (D_1)	4	$(4+1)^{-1} = 0.2$
Gene Ontology (D_2)	3	$(3+1)^{-1} = 0.25$
SwissProt (D_3)	2	$(2+1)^{-1} = 0.33$

root to seed. This is represented as a total number of available edges to the seed, which is used in assigning concept weight along path. This is inversely proportional to total edges between concept and seed along edges. If the result obtained has a varied concept, final weight is assigned high and then semantic vector representation for each dataset or document is formed. Finally, similarity distance is computed. In case of multiple paths or edges, maximum weights are assigned to each individual concept.

2.1.1. Ontology matching problem

In proposed work, ontology is defined in terms of $O=(C, P, I,\Lambda,\Gamma)$, where, C is the set of classes, P is the properties, I is the instances, Λ and Γ are the axioms and annotations. The alignment of ontology between two different ontologies is considered as a correspondence set and individual set consists of 4-tuples $(e, e', n, =)$, where e and e' represent two ontology entities and n is the confidence value that lies in [0, 1]. The correspondence sets between e and $e'$and = is defined as the equivalence relation between e and e'.

If the alignment is one on one, i.e. matching is done between a single entity in source ontology and single entity in destination ontology and vice versa. In the proposed method, the ontology matching is shown by the following Equation:

$$\begin{cases} \max & MF(x) \\ s.t. & X = \left(x_1, x_2, \ldots, x_{|O_1|},\right)^T \\ & x_i \in \{1, 2, ..., |O_1|\},\ i = 1, 2, \ldots, |O_1| \end{cases}$$

where $|O_1|$ represents entity sets cardinalities in ontology O_1 and $|O_2|$ represents entity sets cardinalities in ontology O_2, respectively and x_i is ith correspondence pair. The main objective of proposed method is to attain maximum MF function which is used to calculate X.

2.1.2. Asymmetric angle based similarity measurement (ASM)

The angle based neighborhood distance measurement is useful in finding the cluster heads with associated nearest nodes. The shortest distance is calculated using *is-a* relationship with angle based neighborhood distance measurement. The shortest distance or weight matrix is calculated using $(\text{shortest distance} + 1)^{-1}$. The shortest distance calculation is shown in Table 1. In addition, for a given query (x_q), the angle based neighborhood distance measurement using circumference angle value (α) with N training samples is given as:

$$\angle x_q x_i x_j = across \frac{\left(x_j - x_i\right)\left(x_q - x_i\right)^T}{\left\|x_j - x_i\right\| \cdot \left\|x_j - x_q\right\|} < \alpha, \forall j \neq i,\ j = 1, ..., N \tag{1}$$

Here, if $x_i \in S$, then the value of x_i is an angle based measurement function of x_q, iff $x_j (j \neq i,\ j = 1, \ldots, N)$lies outside the line (marked in red), which is shown in Fig. 2(a). The value of $x_q x_i x_j$ are considered as row vectors. Here, the size of boundary (red line) can be adjusted using a parameter, α.

Further, the center of line joining both circles is considered as a center of circle (Fig. 2(b)). The nodes within this region with high energy than other nodes are considered as cluster center. The other nodes that lie within red line are considered as neighborhood distance nodes. The angle based measurement is used to find distance of seed from cluster center. The seeds with similar values are noted, and shortest distance from cluster center is noted. When two or more query is generated, weight matrix calculation is made similar and maximum weighted seed is selected. Upon weight matrix calculation, semantic vector is assigned to selected seed with highest weight. Overall similarity distance between vectors is calculated using Eq. (1).

The nearest neighborhood distance is evaluated between data using clustering process. Here, angle based neighborhood distance estimation is implemented to find semantic similarity measurement between nearest neighbors. Here, angle based neighborhood distance measurement is chosen explicitly as a metric for similarity measurement, which can be applied to single or multi-class environment.

2.1.3. Ontology semantic measurement using clustering model–distance phase

The main aim of semantic similarity is to avoid limitations due to direct matching of concepts. The matching is done by checking similarity between same concepts in two different datasets. The ontology matching uses is-a relations for linking concepts in ontology (reference). The degree of similarity is calculated with available links between concepts. Semantic similarity is an inverse proportion over distance between neighborhood concepts. The distance between nearest neighbors is calculated with angle based neighborhood distance estimation.

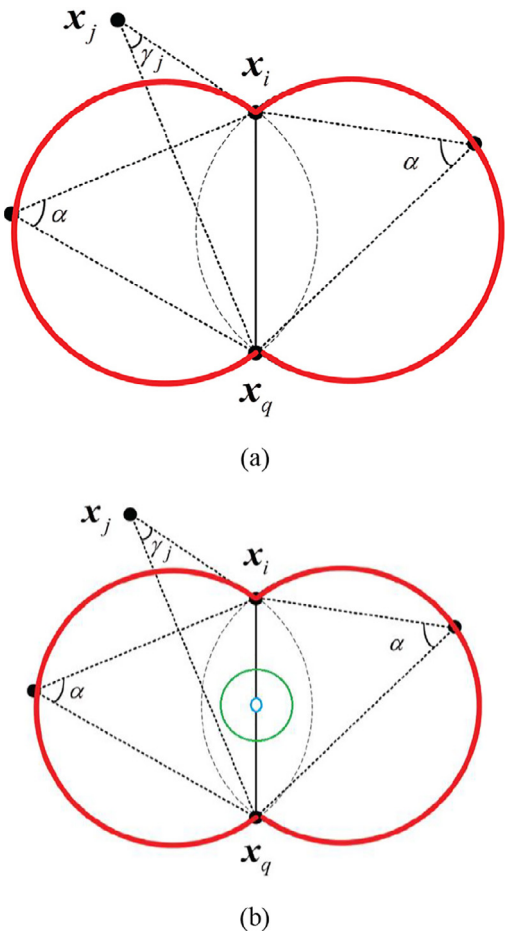


Fig. 2. Angle based neighborhood distance estimation.

In a vector space model, datasets are modelled as vectors over *N*- dimensional space and *N* denotes total available concepts. If vector is considered binary, 1 represents availability of concept and vice versa. Also, no variation in vector is found, since all information available in it is binary valued ones. The similarity measurement between two different datasets is computed using cosine difference between individual vectors. Another limitation associated with similarity comparisons is the matching between vectors with non-orthogonal component. Also, it discards information related to dimensions in which vectors are connected semantically.

The limitation is addressed using standard vector space model that generates vector in following way: The dataset is denoted by *N*-dimensional vector and concept (seed) extracted from related dataset is denoted as 1. During this allocation *is-a* relationship model is used and vector description is increased through inhibition of vectors with weights (non-zero). The dimension of seed corresponds to ancestor's dimensions. The weight (w) between concepts is related using one-to-one relationship between seed and ancestor. Here, we use minimum distance vector to find the distance and weight corresponds to non-orthogonal component, which is directly proportional to depth of common least ancestor. The semantic datasets are formed in such a way and semantic similarity measurement can be defined commonly using cosine pair relationship between two datasets a and b:

$$Similairity(a,b) = \frac{\sum_{i=1}^{N} a_i b_i}{\sqrt{\sum_{i=1}^{N} (a_i)^2}\sqrt{\sum_{i=1}^{N} (b_i)^2}} \tag{2}$$

where, n - union of entire concepts exist in a and b datasets.

2.1.4. Record linkage (RL) using Finite Mixture Model(FMM)–probabilistic phase

The probabilistic record linkage is used to measure the probabilities between relevant seeds, which are linked using similarity measurement in Eq. (1). The probabilistic agreements between matched and non-matched seeds are observed

Table 2
Weights of FMM for agreement and disagreement between the datasets of a specific class.

Database	a_i	b_i	Agreement	Disagreement
NCBI GenBank (D_1)	0.9987	0.0004	8.9260	−5.1974
Gene Ontology (D_2)	0.9821	0.0010	8.6227	−5.1783
SwissProt (D_3)	0.9728	0.0020	8.2324	−4.7285

Table 3
Evaluation of similarity between the databases.

Dataset Combinations	Similarity between the variables
$D_1(a_i)$, $D_2(b_i)$	0.1034
$D_1(a_i)$, $D_1(b_i)$	0.9342
$D_3(a_i)$, $D_2(b_i)$	0.0923
$D_2(a_i)$, $D_4(b_i)$	0.1223
$D_2(a_i)$, $D_2(b_i)$	0.9423
$D_3(a_i)$, $D_3(b_i)$	0.8923
$D_4(a_i)$, $D_4(b_i)$	0.9143

using agreement and disagreement exist between seeds. Next step is to form Latent class model from the observed pairs using clustering process, where correlated attributes are present within clusters. Hence, probabilities are assigned over each clusters using Finite Mixture Model or latent class model. Here, clusters are described in terms of probabilistic model that defines probabilistic distribution of seeds within clusters. Here, similar probabilities arise between relevant seeds within clusters. The main advantage of FMM uses chi-square significance test, which is superior to clustering process. This performs top down approach, while previous section uses bottom up approach.

In FMM, unknown status of pairs are linked with probabilities of a and b datasets. To achieve this, agreement and disagreement weight for a_i and b_i probabilities of ith variable are calculated using:

$$\text{Agreement weight } (A_i) = \log_2\left(\frac{a_i}{b_i}\right) \tag{3}$$

$$\text{Disagreement weight } (D_i) = \log_2\left(\frac{1-a_i}{1-b_i}\right) \tag{4}$$

Probabilistic linking pair is used for calculating individual link weights through summation of total link weight of seed pairs. The weight of seed pair above threshold value is regarded as a link by FMM. The threshold value obtained through estimated match pairs are sorted in a descending order based on total link weights and then it is counted backwards.

3. Performance evaluation

This paper acts as a hybrid model between bottom up and top down approach. It combines correlation and probabilistic approaches to provide best record linkage. The evaluation on data classification in search space is carried out using angle based neighborhood estimation. Matlab simulation tool is used to calculate the neighborhood estimation for finding nearest neighbors. During classification process, test datasets are excluded from its training documents or its potential neighbors. Angle based neighborhood is applied to compare datasets vector against available training datasets. Classification process is handled using angle based neighborhood estimation process with k nearest neighbors. During each iteration, test datasets are returned from training datasets and subsequent dataset is selected. Further, the process continues till iteration ends and hence returned datasets are thus evaluated. The database includes:

1. NCBI GenBank(D_1),
2. Gene Ontology(D_2),
3. SwissProt (D_3) and
4. Human Activity Recognition Using Smartphones Datasets(D_4).

Similarity between similar and different dataset is evaluated to find similarity difference in proposed system. The shortest distance measurement for different datasets is found using weight matrix for different datasets is shown in Table 1. It is seen that shortest distance between the datasets of D_3 database is lesser than other databases. This leads significant weight matrix for D_3 database than other databases (Table 2).

Using weights of two different datasets (a_i, b_i) of an individual database, agreement and disagreement between datasets are found. Similarly, for all datasets, agreement/disagreement between them is found to evaluate similarity between variables in database, which is shown in Table 3, where the dataset variable a_i in dataset D_1 is compared with dataset variable b_i in dataset D_2.

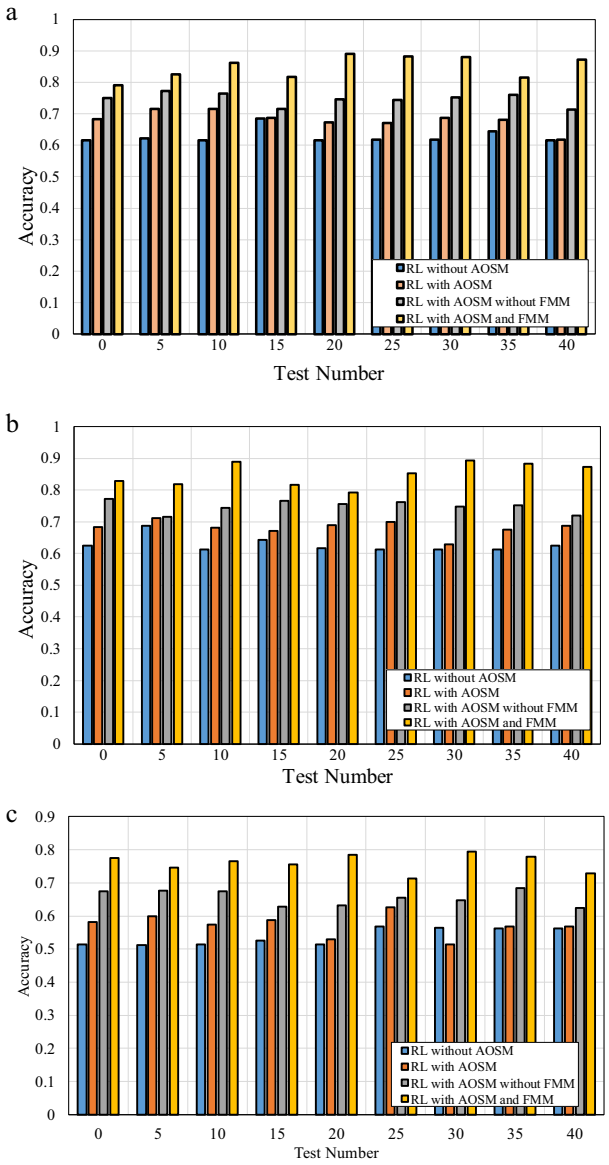


Fig. 3. (a): Evaluation of accuracy between proposed and other techniques for D_1 dataset. (b): Evaluation of accuracy between proposed and other techniques for D_2 dataset. (c): Evaluation of accuracy between proposed and other techniques for D_3 dataset.

It is seen that there is vast declination in similarity measurement between non-agreement datasets. Also, it is seen that agreement between similar datasets poses high-level similarity agreement between seeds. The ratio between agreement and non-agreement samples is in the form of 1:9 ratio. Further, similarity between the datasets involves agreement and non-agreement strategies and accuracy between different modes of proposed method is calculated in Fig. 3.

The accuracy between different modes of proposed method is implemented over the dataset D_1 (Fig. 3). In non-semantic record linkage, accuracy reduces with inaccurate training sets to satisfyrequired output. The required output describes similar datasets with high relevance using reference dataset. Similarly, accuracy of other modes without AOSM and FMM is reduced to a greater extent due to the presence of improper training samples. This is totally reversed in case of proposed technique, where accuracy maintains between 0.80–0.91% over all datasets.

Further, proposed method is compared with other conventional methods like probabilistic record linkage [2], supervised record linkage [3] and graph-based record linkage [7]. The proposed method is compared against precision, recall, and F-measure. The results from Table 4 shows that precision, recall and f-measure for three different datasets attains a better rate in proposed method with FMM. The other methods like probabilistic record linkage, supervised record linkage and graph-based record linkage method underperforms in its precision, recall, and F-measure against the proposed method.

Table 4
Comparison between proposed and existing methods in terms of precision, recall, and *F*-measure between different databases.

Record linkage methods	NCBI GenBank			Gene Ontology			SwissProt		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	*F*-measure
RL + AOSM + FMM	0.96	0.97	0.96	0.97	0.94	0.95	0.69	0.6	0.64
RL with AOSM	0.93	0.97	0.95	0.98	0.92	0.95	0.58	0.74	0.65
Graph based record linkage	0.92	0.77	0.84	0.75	0.98	0.85	0.45	0.74	0.57
Probabilistic record linkage	0.98	0.75	0.93	0.9	0.6	0.72	0.36	0.86	0.5
Supervised record linkage	0.97	0.9	0.93	0.98	0.88	0.93	0.43	0.75	0.55

4. Conclusions

In this paper, proposed algorithm is implemented on both real-world and synthetic datasets. Classical graphbased algorithms for record linkage rated poor performance when compared with angle based neighborhood. Additionally, ontology is used for measuring the similarity between records, which improved the accuracy of similarity measurement between distant variables in datasets. Further, the application of ontology for record linkage improved the searching pattern for datasets at various places in search space. Hence, probabilistic record linkage is combined with ontology based angle neighborhood distance estimation. Finally, comparison with conventional algorithms proved its accuracy, sensitivity and specificity against conventional methods. This work can further be applied on big data analytics, where linkage of datasets from different sources can be considered as a preliminary task. Also, heuristic based approaches could improve the linkage pattern in large datasets with repeated iterations.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compeleceng.2018.01.026.

References

[1] Abril D, Navarro-Arribas G, Torra V. Improving record linkage with supervised learning for disclosure risk assessment. Inf Fusion 2012;13(4):274–84. Available at http://www.sciencedirect.com/science/article/pii/S1566253511000352.
[2] Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. J Clin Epidemiol 2011;64(5):565–72.
[3] Short J, Caulfield B. Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. J Saf Res 2016;58:1–14.
[4] Vatsalan D, Christen P, Verykios VS. A taxonomy of privacy-preserving record linkage techniques. Inf Syst 2013;38(6):946–69. Available at http://www.sciencedirect.com/science/article/pii/S0306437912001470.
[5] Ong TC, Mannino MV, Schilling LM, Kahn MG. Improving record linkage performance in the presence of missing linkage data. J Biomed Inf 2014;52:43–54.
[6] Randall SM, Ferrante AM, Boyd JH, Bauer JK, Semmens JB. Privacy-preserving record linkage on large real world datasets. J Biomed Inf 2014;50:205–12.
[7] Randall SM, Boyd JH, Ferrante AM, Bauer JK, Semmens JB. Use of graph theory measures to identify errors in record linkage. Comput Methods Programs Biomed 2014;115(2):55–63.
[8] Herranz J, Nin J, Rodríguez P, Tassa T. Revisiting distance-based record linkage for privacy-preserving release of statistical datasets. Data Knowl Eng 2015;100:78–93. Available at http://www.sciencedirect.com/science/article/pii/S0169023X15000543.
[9] Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. J Biomed Inf 2015;56:80–6.
[10] Li T, Zhang L, Lu W, Hou H, Liu X, Pedrycz W, Zhong C. Interval kernel fuzzy c-means clustering of incomplete data. Neurocomputing 2017. Available at http://www.sciencedirect.com/science/article/pii/S0925231217300486.
[11] Mandacaru PMP, Andrade AL, Rocha MS, Aguiar FP, Nogueira MSM, Girodo AM, Malta DC. Qualifying information on deaths and serious injuries caused by road traffic in five Brazilian capitals using record linkage. Accid Anal Prev 2017;106:392–8.
[12] Chi Y, Hong J, Jurek A, Liu W, O'Reilly D. Privacy preserving record linkage in the presence of missing values. Inf Syst 2017;71:199–210.
[13] Jurek A, Hong J, Chi Y, Liu W. A novel ensemble learning approach to unsupervised record linkage. Inf Syst 2017;71:40–54.
[14] Smith D. Secure pseudonymisation for privacy-preserving probabilistic record linkage. J Inf Secur Appl 2017.
[15] Abril D, Torra V, Navarro-Arribas G. Supervised learning using a symmetric bilinear form for record linkage. Inf Fusion 2015;26:144–53. Available at http://www.sciencedirect.com/science/article/pii/S1566253514001286.
[16] Lu Y, Sinnott RO. Semantic privacy-preserving framework for electronic health record linkage. Telematics Inf 2017. Available at http://www.sciencedirect.com/science/article/pii/S0736585316304300.

Mr. T. PRABHU is working as Associate Professor, Department of CSE, for Dr. MGR Educational and Research Institute University, Chennai. He has completed B.Sc. degree in Kamaraj College, Tuticorin, MCA in Madurai Kamaraj University, Madurai, M.E in Madha Engineering College, Chennai. Pursuing Ph.D. in Computer Science and Engineering, Manonmaniam Sundarana University, Thirunelveli. His research interest - Data Mining, Distributed Computing, and Information Security.

Dr. C. Suresh Gnana Dhas is working as a Professor & Head, Department of CSE, for Vivekanadha College of Engineering for Women, Tamilnadu. He has more than 25 years of teaching and academic experience in various reputed engineering Universities/ Instutions. His interest areas are Cryptography, Network Security, Network Routing, Wireless Networks, Key Management and distributed systems. He has published 70 reputed international journals and 32 National and 45 conference proceedings. He is the technical program committee, organizer/co-organizer of several conferences and reviewer of international journals.