Exploring the Efficiency of Topic-Based Models in Computing Semantic Relatedness of Geographic Terms

Hossein Sadr
Department of Computer Engineering
Rasht Branch, Islamic Azad University
Rasht, Iran
Sadr@qiau.ac.ir

Mojdeh Nazari Soleimandarabi*
Young Researchers and Elite Club,
Rasht Branch, Islamic Azad University
Rasht, Iran
Mozhdeh_nazary@yahoo.com

Mir Mohsen Pedram
Department of Electrical and Computer Engineering
Faculty of Engineering, Kharazmi University
Tehran, Iran
Pedram@khu.ac.ir

Mohammad Teshnehlab
Industrial Control Center of Excellence, Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran
Teshnehlab@eetd.kntu.ac.ir

Received: 2020/04/10 Revised: 2020/06/30 Accepted: 2020/07/04

***Abstract*— Large number of semantic relatedness measures have been presented since the last decades. In spite of an extensive number of studies that have been conducted in this field, the understanding of their foundation is still limited in real world applications. In this paper, the state-of-the-art semantic relatedness measures are surveyed and in the following a unified topic-based models is proposed to highlight their equivalences and propose bridges between their theoretical bases. Presentation of a comprehensive unified approach of topic based models induces readers to have common understanding of them in spite of the complexities and differences between their architecture and configuration details. Moreover, it may underlie fundamental development of these models. Comprehensive experiments in application of semantic relatedness of geographic phrases have been conducted to evaluate topic based models in comparison to ontology-based models. Based on the obtained results, not only topic-based models in comparison to ontology-based models confront with fewer restrictions in real world, but also their performance in computing semantic relatedness of geographic phrases is significantly superior to ontology-based models.**

Keywords*— *Semantic Relatedness; Topic-based Models Latent Semantic Analysis; Latent Dirichlet Allocation; Explicit Semantic Analysis; Geographical Information Science Introduction.

1. INTRODUCTION

Semantic relatedness is a measure which is used to indicate degree to which words are related via any type of relations, including similarity and any possible semantic relationship [1]. Generally, two words are considered semantically related if they are about things that are related to each other in the real world. Human beings have an innate ability to determine if two words are related. For example, most would agree that *Island* and *Sea* are related while *Island* and *Car* are not. However, assigning a value that quantifies the degree to which two words are related still remains an insurmountable obstacle for computers [2].

Geographic knowledge is a crucial benefit in human activities. To share geographic information, it is necessary to extract concepts from the chaotic repository of implicit human minds' knowledge. Whereas up to 80 percent of human decisions affect space or are affected by spatial situations computing semantic relatedness of geographic terms plays a pivotal role in geographic information science [3, 4]. For example, people read news about important local events; search the web to find delicious restaurants; or find some information about their favourite celebrity. Actually, all words are more or less related to particular locations. Therefore, computing the semantic relatedness of geographic terms is an important technique in discovering the relationship among terms in geographic map, geographic information retrieval, explanatory search and geographic information science [5-7].

Background information about the particular concepts or terms is required to measure the lexical semantic relatedness. The primary techniques employed ontologies like WordNet to investigate relationships among terms [4]. Although these methods utilized explicit senses or concepts that humans can interpret and reason about, they could not provide enough coverage of terms in particular domain. Moreover, creating these knowledge bases is costly and time-consuming. Consequently, these measures are confronted with serious constraints in real world [2].

Considering the challenges that ontology-based models are confronted with yielded to the emergence of other approaches that are able to learn bags of related words from large corpora by analysing the terms co-occurrences without any supervision. These techniques are able to create low dimensional feature representation or concept space where terms are not considered independently. Noteworthy, these techniques can investigate hidden relations between a set of words. ESA [8], LSA [9] and LDA [10] are three indicators in this field. These methods are

facing two remarkable problems. First, in spite of the conceptual similarity between these models, each of them has its own architecture. Furthermore, each of them has been utilized and examined in a particular application. Second, architectural diversity of these models has led to difficult interpretation of them while a comprehensive comparison of their efficiency in a specific application has been unknown yet.

To this end, the state-of-the-art measures of semantic relatedness are surveyed in this paper and a unifying approach for conceptual representation of ESA [8], LSA [9] and LDA [10], known as *topic-based models*, is proposed in the following. It must be noted that *probabilistic topic models* [11, 12] were referred to LDA domain and its related improved methods in previous studies. However, *topic-based models* terminology is used in this paper which covers an extensive range of methods, which are precisely explained in section 3. The efficiency of these topic-based models has been comprehensively examined in application of computing semantic relatedness of geographic phrases. Moreover, to present the priority of topic-based models, these models have been compared to an extensive range of ontology-based models. To the best of our knowledge, no geo-semantic relatedness measure focused on topic-based models has ever been proposed.

The rest of this article is categorized as follows: Various computing semantic relatedness measures as well as the used knowledge sources are explained in section 2. A unifying approach for representing topic-based models is defined in section 3. Empirical results, evaluation framework and benchmark dataset in application of geographic semantic relatedness are described in section 4. Conclusions and directions for future research are mentioned in section 5.

2. Review of Literature

Due to geographic domain, computing relatedness is an important technique for discovering functional relation between places and constructing lexical resources. Over years, extensive researches have been carried out to study the methods of relatedness in geographic domain [13]. Accordingly, in this section, the state of the art in computing semantic relatedness is presented, and existing algorithms are categorized into two distinct types of measures, where each type exhibits unique properties. Computing semantic relatedness requires lexical and semantic information of terms and concepts usually encoded in some background knowledge. Various methods of semantic relatedness considering their background knowledge are divided into two distinct types: *ontology-based* and *topic-based*. Whereas, the emphasis of this paper is on computing semantic relatedness of geographic phrases, the scope of discussion will include studies from both the general and geographic domains.

2-1. Ontology-based models

The the term ontology refers to a taxonomic structure enriched with other semantic relationships such as antonymy and synonymy, and class properties or attribute. In ontology-based models, the semantic relations of concepts defined in ontology are used to compute semantic relatedness. Background knowledge resources are used for computing semantic relatedness in ontology-based models. Some methods, especially earlier ones, have leveraged dictionaries and thesaurus [14]. Over time WordNet has changed into one of the most popular ontologies for computing semantic relatedness. WordNet is a lexicalized ontology of English words. It groups nouns, verbs, adjectives, and adverbs into synsets, each expressing a distinct concept. Recently, Wikipedia has been employed as a strong knowledge resource for computing semantic relatedness [15]. In general, ontology-based models are categorized into two types: path based, information content based.

2-1-1. Path based measures

Path based measures determine the length of the path between nodes in an ontology. Most of these measures make use of the taxonomic links of an ontology for calculating the path length. Length of a path is obtained by counting the number of nodes or edges in a path. The shorter the path, the higher relatedness between concepts. Despite the simplicity of this method, it has proved a successful in initial application and various optimizations have been applied on it [16, 17].

Path was presented as a basic method which leveraged WordNet graph structure as ontology and considered inverse shortest path between two concepts [18]. The shorter the path from one node to another, the more related they are (Eq.1):

$$rel_{path}(c_1, c_2) = \frac{1}{max\ len\ (c_1,c_2)} \quad (1)$$

This method performed fairly well, but didn't take the graph depth into account. [19] normalized the path length using depth of the graph and solve Path's shortcoming (Eq.2):

$$rel_{Lch}(c_1, c_2) = -log\ \frac{len\ (c_1,c_2)}{2\times \underset{c\in wordnet}{depth(c)}} \quad (2)$$

Where $len\ (c_1, c_2)$ shows the path length between two concepts and depth is the length of the longest path from the root node of the taxonomy to a leaf node.

On the other hand, these basic path length methods did not take into account that the higher concepts in taxonomy are more abstract, i.e., concepts with length of one near top of taxonomy are more related than concepts with the same path length on the leaf level. To overcome this issue many measures have been proposed. [20] proposed a measure which leveraged the notion of lowest common subsumer (LCS) of two concepts. LCS is the first shared concept from the leaf to the root of hierarchy. It made difference between abstract and specific concepts of taxonomy by considering LCS (Eq.3):

$$rel_{wp}(c_1, c_2) = \frac{2\times depth\ (lcs)}{Depth\ (c_1)+depth(c_2\ )} \quad (3)$$

WordNet graph structure was also used for computing relatedness [21]. Unlike above method which only took *is-a* relations between concepts into consideration, HSO used all of existing relations in WordNet. The hypothesis stated that the strength of relatedness is correlated with path length and frequency of direction changes along the path, where change of directions and long path are penalized. i.e., two terms are related if they are in the same synset, they are antonym or one word is part of another (Eq.4):

$$rel_{HSO}(c_1, c_2) = C - len(c_1, c_2) - k.turns(c_1, c_2) \quad (4)$$

Where *C* and *K* are consonant, *len* is path length and *turn* is the frequency of direction changes between two concepts.

Although the frequency of changes is less, semantic relatedness between two concepts is more.

WikiRelate [22] was proposed as the first measure which used Wikipedia as ontology for computing semantic relatedness. In this study, they used Wikipedia category graph for computing the path length between two concepts. They applied WordNet path techniques on Wikipedia graphs.

Using Wikipedia as a knowledge resource has priorities compared to WordNet. Wikipedia is an encyclopaedia not a dictionary. It covers more nouns, concepts and domain specific terms. Moreover, the hierarchical structure of Wikipedia is a strict taxonomy and all relations such as hypernomy and mernonymy are defined in it [23]. Meanwhile, WordNet has some advantages compared to Wikipedia. Since Wikipedia does not name hyperlink between articles using semantic relations, the hierarchical structure of category tree rather represents a loose folksonomy than a strict taxonomy [24]. It is notable that in Wikipedia unlike WordNet all semantic relations are not explicitly specified.

These methods have been adapted in geographic domain with small modifications in a number of studies. Matching Distance Similarity Measure (MDSM), proposed by [25], was one of the first semantic relatedness measures, which has been specifically developed for geographic domain. Based on this method, asymmetric values for relatedness of spatial entity classes were achieved based on their degree of generalization within a hierarchical structure. In the other word, it compared entity classes in terms of their distances in the semantic structure that was defined by the semantic relations.

Following a similar line of research, [26] developed a method which used Wikipedia article graph for computing semantic relatedness. Based on their notion, the relatedness score was computed by assigning weight to spatial referred articles in Wikipedia article graph. It is worth noting that semantic networks which encode knowledge and meanings in the form of graphs, have been also used in computing semantic relatedness of geographic terms. Based on this notion [27] developed a method which was based on some forms of structural distance between nodes (e.g. edge counting) or on the topological comparison of sub graphs.

Recently [28] developed a method which leveraged Volunteered Geographic Information (VGI) for computing semantic relatedness. VGI is a large reusable unit of geographic knowledge generated by heterogeneous information communities. Using VGI information, they applied graph-based measures of semantic relatedness on Open Street Map (OSM) semantic network. Whereas, OSM semantic network consists of noisy and ambiguous data, in similar work they enriched the OSM semantic model with semantic web resources [29].

2-1-2. Information content based measures

Information content (IC) approaches are based on this hypothesis that the relatedness of two concepts depends on the amount of information that they share. Information content also presents the generality or specify of a concept. In taxonomy the shared properties of two concepts is often determined with respect to their LCS. These methods combine knowledge of a concept's hierarchical structure with statistics of its actual usage in text.

The first IC-based method is introduced by [30], which used WordNet as an ontology. Based on this measure the information content between two concepts was computed respect to their LCS. This means that if two pairs of terms have the same Lowest common subsumer, the semantic relatedness between them will be equal (Eq.5):

$$rel_{res}(c_1, c_2) = IC\left(lcs(c_1, c_2)\right) \qquad (5)$$

The information content of a concept is computed as (Eq.6):

$$IC(c) = -\log P(c) \qquad (6)$$

Where *p(c)* is the probability of encountering an instance of a concept *c* in a large corpus. Resnik's definition of IC is widely used by later methods. Most of the later

IC-based methods improved Resnik's in different ways to overcome a number of limitations. To address Resink's problems, [31] proposed a measure. In this method if a parent node is subsumer, leaf nodes are used to compute semantic distance. Since semantic relatedness is proportional to inverse semantic distance; by increasing semantic distance, semantic relatedness will decrease (Eq.7):

$$rel_{jcn}(c_1, c_2) = \frac{1}{IC(c_1)+IC(c_2)-2\,IC(lcs)} \qquad (7)$$

On the other hand, [32] proposed a universal measure derived from information content. In the beginning, it was only applied on taxonomic structures (Eq.8).

$$rel_{lin}(c_1, c_2) = 2.\frac{IC(lcs)}{IC(c_1)+IC(c_2)} \qquad (8)$$

According to Resink, any two pairs of concepts having the same lcs will receive the relatedness, which is not necessarily appropriate. To overcome this issue, Lin' measure aims to address the commonality of two concepts as well as their difference, both measured in terms of IC.

2-2. Topic- basd models

Topic-based models leverage statistical analysis on background corpus to build topics. In these approaches, unlike ontology-based models, relations among concepts are not organized structurally and their connections are unclear and they do not provide word senses and semantic relations among terms and terms' definitions explicitly. As a result, background knowledge is obtained at the term level and terms are connected implicitly. Based on these measures, term co-occurrences in an unstructured corpus are used for training topic and computing relatedness. Their background knowledge is extracted by performing statistical analysis on a large collection of unlabeled documents and their ultimate aim is to discover hidden structure between documents.

Latent semantic analysis (LSA) [9] is one of the most important measures in this field which uses vector presentation for computing relatedness and it is able to discover hidden structure among terms. It is a dimensional reduction approach which applies Singular Value Decomposition on term-document matrix in order to map terms to latent topics and generalize observed relations between terms and topics. Stevens et. al [33] proposed a measure which used Non-Negative Matrix Factorization for reducing the dimensions of term-document matrix. Following a similar line of research,

[34] used various global and local weighting method for constructing term-topic matrix in order to improve the performance of LSA.

Latent Dirichlet allocation (LDA) [10] is another existing technique in this field. It is a generative probabilistic model for collections of discrete data such as text corpora while LSA is an algebraic approach. Based on LDA, documents are presented as distributions over a set of topics and each term in a document is generated based on a distribution over terms which are specific to each topic. According to this, terms are related whether they have same topics or same topic distribution. LDA was rarely used for semantic relatedness. [35] proposed a method based on various distributions of LDA for computing relatedness.

Another approach was introduced by [8] and it was referred as Explicit Semantic Analysis (ESA). Based on ESA, vectors constructed from Wikipedia concepts are used for computing relatedness. The dimensions of concept vectors are equal to the number of Wikipedia articles and each element in the vector is weighted by TFIDF [36]. Semantic relatedness is computed using cosine similarity function.

Surprisingly, in GIScience topic-based models have been almost ignored except some notable exceptions. More recently, by considering spatial co-occurrences features, [37] extracted a relatedness measure directly from OpenStreetMaps vector data. Moreover, [38] proposed a method for computing relatedness of geographic terms which leveraged frequently used semantic measures for computing relatedness such as ESA [39] and generated human readable explanation by mining text hyperlinks and Wikipedia category graph.

Furthermore, LDA was extended to compute semantic relatedness in geographic domain. LDA adopts a probabilistic approach to cluster highly semantically related terms and include geographic dimension to Location Aware Topic Model (LATM) [40], which used as fully distributional approach for computing semantic relatedness. Recently, [41] proposed a hybrid method to quantify semantic relatedness of lexical definition. Based on their idea, related terms tend to be defined related terms. This measure combined existing WordNet and paraphrase detection techniques for computing semantic relatedness.

3. A UNIFYING APPROACH FOR TOPIC-BASED SEMANTIC MODELS

Terms that are appeared in a text are generally used as features in most machine learning methods. In fact, a vector in term space which is known as vector space representation is used to describe a document [42, 43]. The intuition behind topic modelling refers to the fact that textual documents can be expressed in terms of a limited number of underlying topics (so-called "concepts" in the literature of semantic information retrieval). To this end, topic-based models can be defined as a set of unsupervised algorithms that aim to find the hidden thematic topics in a large collection of documents. These models employ the occurrences of terms in documents for training the model and the topics that are extracted after training are generally consistent with human concepts [8].

Various studies have already been conducted on this basis and each model has its own definition and domain specific applications. Considering the diversity and complexity of these models, providing a common interpretation of them can be valuable. Additionally, a comprehensive investigation of the performance of different models in the same application is not accessible.

Utilizing explored topics instead of term occurred in a document can result in an extensive set of semantic applications in the field of natural language processing and information retrieval. Generally, topic-based models have two primary components:

- **Topic representation**: Topic-based models generally include two types of representation as explicit and latent topics [44]. Explicit concepts show real world concepts and are based on human perception and knowledge. On the other hand, latent concepts are achieved by extracting the mathematical relations between terms and computing the terms occurrences probability.
- **Mapping terms to topics algorithm:** The goal of this mechanism is to map natural language term to the topic. It is worth mentioning that constructing the manual map is the most accurate one. i.e., a handcrafted ontology of terms is assigned to their corresponding concepts. This approach needs a lot of effort besides having a lot of complexity. In this regard, machine learning and information retrieval techniques are frequently used to perform mapping operation.

In all of topic-based models, each term appeared in a text is assigned to a set of topics by different degree of membership. The difference between these models refers to definition of topics and how each term is mapped to a corresponding set of topics. In this section, three indicators of this field are separately studied. Each model is designed with various goals and is based on various theories. Regardless of topic types and mapping term to document methods, various models can be described in a term of common definition.

Definition 1: Document- term occurrence matrix

For a set of documents $D = \{d_1, d_2, \dots, d_n\}$ and a set of terms $W = \{w_1, w_2, \dots, w_m\}$, each document is represented as a vector including terms that are appeared in it [36] (Eq.9):

$$\mathcal{D} = \begin{matrix} \\ d_1 \\ \vdots \\ d_n \end{matrix} \begin{matrix} \begin{matrix} w_1 & \dots & w_m \end{matrix} \\ \begin{bmatrix} \mathcal{D}_{1,1} & \dots & D_{1,m} \\ \vdots & \ddots & \vdots \\ \mathcal{D}_{n,1} & \dots & D_{n,m} \end{bmatrix}_{n*m} \end{matrix} \quad (9)$$

$\mathcal{D}$ is a co-occurrence matrix, where the rows of this matrix refer the documents of corpus *(D)* and its columns show the terms appeared in documents *(W)*. Each element $\mathcal{D}_{i,j}$ shows the importance of term w_j in document d_i. different methods have been proposed for weighting the terms of documents [36].

Definition 2: Topic model

Different topic models contain particular set of topics $\mathcal{T} = \{t_1, t_2, \dots, t_k\}$ and a matrix ϕ as follows (Eq.10):

$$\phi = \begin{matrix} \\ w_1 \\ \vdots \\ w_m \end{matrix} \begin{matrix} \begin{matrix} t_1 & \dots & t_k \end{matrix} \\ \begin{bmatrix} \varphi_{1,1} & \dots & \varphi_{1,k} \\ \vdots & \ddots & \vdots \\ \varphi_{m,1} & \dots & \varphi_{m,k} \end{bmatrix}_{m*k} \end{matrix} \quad (10)$$

The rows of this matrix refer tp the terms appeared in a corpus (W) and its columns show the set of topics ($\mathcal{T}$). The

values of this matrix are commonly trained according to the terms occurrences in documents of corpus.

Definition 3: Mapping text to topic space

The document space *(D)*, term space *(W)* and topic space *($\mathcal{T}$)* are three common spaces in all existing topic-based models that are able to map to each other. Mapping from one space to another is performed using matrix multiplication (Eq.11).

$$\mathcal{W}_{n\times k} = \mathcal{D}_{n\times m} \times \phi_{m\times k} \quad (11)$$

$\mathcal{W}_{n\times k}$ refers to the contribution of each document to each topic and it can be employed as a new representation of documents in topic space. In other word, in this matrix each document is defined by a vector of topics appeared in it. Based on the model, the process of mapping text to topic is known as inference [10] or interpretation [8].

Definition of topic space *($\mathcal{T}$)* and methods of topic training (ϕ) are various in different topic models. For instance, ESA [8] used Wikipedia articles as topic space and TFIDF [36] as weighting method for training topics. While in LSA [9], Eigenvectors and Singular Value Decomposition were respectively utilized to define topics and create a matrix. In the following, each model is explained separately.

3-1. Explicit Semantic Analysis

As previously mentioned, the differences between various topic-based models refer to the definition of topic space *($\mathcal{T}$)* and term to topic mapping method (ϕ). ESA [8] utilized Wikipedia articles as topic space ($\mathcal{T}$) and term frequency in different articles as a tool for mapping each term to topic space (ϕ). While set of Wikipedia articles are utilized as a topic space, term to topic mapping matrix can be shown as follows (Eq.12):

$$\phi = weighted\ inverted\ index = \begin{matrix} \\ w_1 \\ \vdots \\ w_m \end{matrix} \begin{matrix} t_1 & \dots & t_k \end{matrix} \begin{bmatrix} \varphi_{1,1} & \dots & \varphi_{1,k} \\ \vdots & \ddots & \vdots \\ \varphi_{m,1} & \dots & \varphi_{m,k} \end{bmatrix}_{m*k} \quad (12)$$

Where the rows of this matrix refer to the terms appeared in Wikipedia and its columns show the topic space (Wikipedia articles). Weight of each term w_i in each topic t_j is presented by $\phi_{i,j}$ and computed based on term frequency of term w_i in article t_j and inverse document frequency of term w_i in all Wikipedia articles as follows (Eq.13):

$$\phi_{i,j} = tf(w_i, t_j) \cdot \log \frac{n}{df_{w_i}} \quad (13)$$

Where *n* shows the number of Wikipedia articles. Matrix ϕ is reffered as weighted inverted index [8]. This matrix is responsible for mapping each text to semantic space and it is performed by semantic interpreter [8]. Finally, represented words in topic space are compared with standard vector comparison methods (e.g., cosine similarity measure [36]).

It can be said that the high performance of this model in a wide range of information retrieval applications can be due to the representation of documents based on Wikipedia topic space instead of terms occurred in a document .Particularly, this model can determine the relatedness of documents and words with correlation 0.72 and 0.75 in comparison to human judgment [8].

Although using human concepts as a set of topics makes this model more interpretable, the dependency between its topics leads to many problems in vector space like synonymy and polysemy. Moreover, unlike the statistical latent topic-based models [9, 45], the orthogonal property of topics that has a key role in performance is not properly addressed in this model. Based on ESA, each document representation is computed by its weighted relations with all topics. Therefore, each documents can be explained by a vector including thousands of related topics in Wikipedia. Therefore, document representation in this model is excessive and unlike many statistical latent topic-based models [9, 45] which try to reduce dimensions, is associated with increased dimensions. Enhancing number of dimensions leads to greater complexity in many algorithmic issues. In CHESA [46] method, it is tried to overcome this problems by defining a hierarchical structure among topics using Wikipedia category graph.

Considering the fact mapping terms to documents is only performed using term occurrences in various document, ESA generated topics for a piece of text can not only be noisy but also contain over specific topics [46]. This leads lower efficiency of this model in dealing with longer documents like *TREC* documents.

3-2. Latent Semantic Analysis

Unlike ESA [8] where topics are presented explicitly, LSA's topics are computational and latent. Principally, LSA [9] is a dimensional reduction method which can map a set of documents to topics by applying Singular Value Decomposition on $\mathcal{D}$ matrix and extracting k eigenvectors . According to Figure 1 and applying Singular Value Decomposition:

$$\acute{\mathcal{D}} = U\sum \quad (14)$$

Where $\sum$ is a diagonal matrix with equal dimensions to $\mathcal{D}$ which stores Eigenvectors. U and V are orthogonal matrices where their columns are equal to Eigenvectors corresponding to data in $\mathcal{D}$. Eigenvectors are able to store data in higher energy and lower dimensions. k Eigenvectors of U matrix can be utilized as data similarity matrix. By using this matrix, data can be mapped into a new k dimensional space (topic space). In fact, matrix multiplication $P = U_k\acute{U_k}$ is equal to a matrix which shows terms semantic relatedness using their latent topics. If the number of input documents is high, this matrix can be utilized in different applications of semantic relatedness [24].

Notably, k Eigenvectors of U matrix are utilized as topic space in this model. Furthermore, U_k matrix, extracted by applying Singular Value Decomposition on $\mathcal{D}$ matrix, is responsible for mapping terms to topic space (Eq.15):

$$\underset{\mathcal{D}'}{\begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix}_{n*m}} = \underset{U_k}{\begin{bmatrix} \begin{bmatrix} \vdots \\ U_1 \\ \vdots \end{bmatrix} & \dots & \begin{bmatrix} \vdots \\ U_n \\ \vdots \end{bmatrix} \end{bmatrix}_{n*k}} \underset{\sum_k}{\begin{bmatrix} S_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & S_k \end{bmatrix}_{k*k}} \underset{V_k}{\begin{bmatrix} \begin{bmatrix} & V_1 & \end{bmatrix} \\ \vdots \\ \begin{bmatrix} & V_k & \end{bmatrix} \end{bmatrix}_{k*m}}$$

Fig 1. $\mathcal{D}$ matrix decomposition **u**sing SVD.

$$\phi = U_k = \begin{matrix} \\ w_1 \\ \vdots \\ w_m \end{matrix} \begin{matrix} \begin{matrix} t_1 & \dots & t_k \end{matrix} \\ \begin{bmatrix} \varphi_{1,1} & \cdots & \varphi_{1,k} \\ \vdots & \ddots & \vdots \\ \varphi_{m,1} & \dots & \varphi_{m,k} \end{bmatrix}_{m*k} \end{matrix} \quad (15)$$

Applying k vectors with higher energy not only yields to reduction in space but also causes enhancement in quality of future process [47]. While related terms have shared topics, the quality of future process ing dealing with synonyms is improved.

The value of various terms semantic mixture in creating latent topics is controlled by k parameter. This means that where the value of k is higher, the value of term mixture is lower and different terms are organized in more latent features. On the other hand, the numbers of computational dimension spaces are also higher.

Although extracted topics by LSA in comparison to ESA are not easily interpreted by humans, this set of statistical latent topics is uncorrelated. Moreover, the number of defined topics in this model compared to the number of terms in the source is very low. Topic independency along limited number of topics makes this method more flexible and provides it with the opportunity to have remarkable efficiency in various applications of machine learning [9, 47].

3-3. Latent Dirichlet Allocation

Many alternative matrix smoothing processes instead of Singular Value Decomposition have been proposed Since the work of Deerwester et al. [48]. LDA [10, 11], perhaps the most common currently in use matrix smoothing process, developed by Blei et al., allowing documents to have a mixture of topics. Unlike LSA [9], LDA is a fully generative model, where documents are assumed to be generated based on a per-document topic distribution (with a Dirichlet prior) and per-topic word distribution [10].

According to LDA [10], a probabilistic topic is a probability distribution over a collection of terms and a topic-based model is a formal statistical relation between a group of explicit and latent random variables that defines a probabilistic process generating topics.

Same as other topic-based models, each document is presented as mixture of various topic in LDA where it is assumed that the prior topic distribution in document is followed by Dirichlet allocation. Learning distribution of terms over topics is a Bayesian Inference and in principle, it is that term to topic mapping matrix (ϕ).

In LDA, by assuming documents are generated by a specific probabilistic model, the relations between term, topics and documents are learned. It is hypothesized that there are fixed set of topics that have been utilized throughout the corpus and each topic t_i is related to a multinomial distribution over terms, which is derived from Dirichlet prior. The generative model for each document d_i is expressed as follows [10]:

1. $Choose\ \ominus_i \sim Dir\ , a\ topic\ distributation\ for\ d_i$
2. $for\ each\ w_i \in d_i$

   $a) select\ a\ topic\ t_j \sim \ominus_i$

   $b) select\ the\ word\ w_i \in \emptyset_{t_j}$

Where $\ominus$ shows the probability distribution of topics over documents that are randomly selected using Dirichlet allocation. Moreover, ϕ shows the probability of terms being used for each topic, which is drawn from Bayesian inference. These two sets of distribution correspond to W and ϕ matrices. This model relies on α and β parameters for computing distribution over topics and terms. Consequently, term to topic mapping matrix is expressed as follow (Eq.16):

$$\phi = \begin{matrix} \\ w_1 \\ \vdots \\ w_m \end{matrix} \begin{matrix} \begin{matrix} t_1 & \dots & t_k \end{matrix} \\ \begin{bmatrix} \varphi_{1,1} & \cdots & \varphi_{1,k} \\ \vdots & \ddots & \vdots \\ \varphi_{k,1} & \dots & \varphi_{m,k} \end{bmatrix}_{m*k} \end{matrix} \quad (16)$$

Here each row shows the distribution of particular term over a specific set of topics and each column shows a fixed set of randomly selected topics. The weight of each term in each topic $\phi_{i,j}$ is measured using Bayesian inference. Finally, semantic relatedness measure between terms is expressed as dot product between their corresponding vectors obtained for topic space [35].

It is worth mentioning that each topic-based model has its potential and drawbacks. Although extracted topics from LDA [10, 11] similar to LSA [9] are latent, it has conceptual advantages over LSA. It is able to make difference between meanings of terms, i.e. it handles polysemy [35]. In LDA, each topic is a set of terms that express a meaning together, while LSA has a particular representation for each term. It means that all meanings of a term are represented by the same vector making it hard to distinguish which meaning is being referred.

On the other hand, LDA is a probabilistic model with interpretable topic. Although LDA's computational complexity and execution time are higher than LSA [9] and ESA [8], its topics are more interpretable than LSA. Major disadvantages of LDA is that it is difficult to know when LDA is working since topics are soft-clusters and there is no objective metric to say "this is the best choice" of hyper parameters. If learning is working, metrics like perplexity can be analysed. It is notable that they are very poor indicators of the overall quality of the model, for example, it is possible have a model with very low perplexity; however, whose topics are not very informative.

4. EXPERIMENTS

All of topic-based models have one thing in common. All of them map each term of a text to a set of topics. Whereas the types of generated topics in various models are inconsistent with each other, evaluating the quality of generated topics in different methods is not possible. For example, topics generated in LSA [9] are mathematical topics while ESA [8] generated topics are Wikipedia articles. Although a plenty of measures have been introduced for evaluating the generated topics [33], previous researches have shown that the quality of generated topics according to topic coherence measures did not necessarily lead to better performance [33]. All of the methods proposed for evaluating topic-based models discussed in section 3 were adhoc and application specific. In this section a comprehensive evaluation of all three methods; LSA [9], ESA [8] and LDA [10] in an extensive framework of proposed approach and in application of geographic semantic relatedness is presented.

4-1. Evaluation method

Evaluating of semantic relatedness measures is a crucial factor in examining the strengths and weakness of different approaches and provides a transparent view of their performance. In addition, evaluation results can be used to strengthen and restructure the model. Semantic relatedness measures are typically evaluated by two types of approaches*: in-vivo* and *in-vitro*.

4-1-1. In-vivo evaluation

Due to in-vivo experiments a relatedness measure is evaluated directly by assessing performance of a specific application. There are several applications in natural language processing that require computing semantic relatedness such as: Word sense disambiguation [49], named entity recognition [50] and information retrieval [51, 52]. By applying various relatedness methods on these frameworks and evaluating their performance, different methods can be analysed in a specific application.

It should be taken into consideration that the performance of each computing semantic relatedness framework is based on its background knowledge. Moreover, in-vivo evaluation entails the influence of application-specific parameters beside the measures of semantic relatedness. Accordingly, this evaluation is not always precise and accurate and may generate different results in different application. As a result, in-vivo cannot be used as a comprehensive measure for evaluating semantic relatedness methods.

4-1-2. In-vitro evaluation

The main contribution of topic-based models is mapping terms to topic space. Therefore, a method for evaluating these models is leveraging extracted topics for text representation and comparing them with scores of semantic relatedness introduced by human judgment. In other word, superior models should be able to represent each term in a way that computing relatedness using this representation has more agreement with relatedness scores determined by humans.

Briefly, in-vitro evaluation is done by mathematical analysis and correlation with human judgment. Typically, a dataset containing a set of pairs of terms are given to human judges, who estimate their relatedness within a certain scale. Different methods of semantic relatedness are applied in this dataset. Finally, the results are correlated against the human judgement to derive an indication of the accuracy of the method. However, in-vitro evaluation does not assess how well the method performs on real data [53].

It is worth noting that in-vitro evaluations are independent of background knowledge and various parameters of the underlying application. Moreover, based on them different approaches are compared in the same situation. According to these privileges, they are referred as gold standard for evaluating semantic relatedness methods. In order to obtain a deeper insight about semantic relatedness measures and compare different methods in the same situation without any regard to background knowledge, in-vitro evaluation is used in the following experiments for evaluating different topic-based models and comparing them to other existing methods.

4-2. Benchmark dataset

Over the past years, several datasets focused on semantic relatedness and similarity have been created. The first gold standard was introduced by Robertson and Goodenough in 1965 containing 65 word pairs ranked by their synonymy [54]. Following the similar line of research, Miller and Charles presented a similar dataset containing 30 word pairs in 1991 [54]. More recently, WordSimilarity-353 was published by Finkelstein [55] in 2002 which contains 353 word pairs ranked by their semantic relatedness.

A smaller number of geographic semantic relatedness datasets have been created in the field of geographical information science and geographic information retrieval. In this field, Rodriguez and Egenhofer [25] published their dataset MDSM for evaluating their Matching Distance Similarity Measure in 2004. This dataset contained human judgments about relatedness among 33 geographic terms. In 2008, Janowicz et al. [56] published a dataset for evaluating Sim-DL measure. It was a small dataset containing six geographic terms related to bodies and water. Whereas the size of all existing dataset tends to be rather small and they only focused on semantic similarity, [54] published Geo-Relatedness and Similarity Dataset (GeReSiD) in 2014. Before GeReSiD, the MDSM dataset was the largest similarity gold standard for geographic terms.

GeReSiD includes 97 geographic terms combined into 50 phrase pairs. This dataset is larger than the other existing datasets of this field and contains both natural and man-made phrases. Unlike other existing datasets, it focused on geographic phrases. The judgments of this dataset were collected form 203 Native English speakers through an online survey [54]. GeReSiD is available online and it can be used as a valuable resource of evaluating geographic models and defining correlation of experimental results with human judgment in GIScience.

In this dataset the notion of geo-semantic relatedness is discussed based on Lehrer's semantic field [54]. Semantic field is a set of lexemes which cover a certain conceptual domain and bear certain specifiable relation to another. Therefore, geo-semantic relatedness is defined as specific sub-domain of semantic relatedness focusing on relations ground in geographic dimension, i.e., relations in which at least on pair has spatial dimension. For example, *stadium* and *basketball* are geographically related, where stadium is a strong geographic component that grounds the other term geographically.

In order to increase the usability and clarity of this dataset, its terms have been mapped to their corresponding sysnsets in WordNet. Unlike other existing datasets [25], this dataset has a uniform distribution, i.e., the number of high relatedness pairs (16), middle relatedness pairs (18) and low relatedness pairs (16) comply the same distribution. It is notable that choosing term with uniform distribution can be challenging for different methods of semantic relatedness and can analyze their accuracy in various fields. For example, if the semantic relatedness measure determines low relatedness for each pair and the majority of samples used in dataset are irrelevant, the correlation between experiment results and human judgments will always be high. As a result, the weaknesses of this method will not become characterized.

Moreover, GeReSiD specifies the explicit differences between geo-semantic relatedness and similarity. Several pairs in this dataset are related but not similar and their scores confirm this. More specially, <*speed bump, car park*> obtained a relatedness scores of 0.54 and a similarity of 0.38 and this clarified that these phrases were related but not similar. Based on results, semantic similarity is generally lower than semantic relatedness and this clearly reflects that semantic similarity is a specific kind of semantic relatedness [28, 41, 54].

As the results of this dataset have been collected online, they may not be accurate and reliable. For this purpose, the semantic judgments on the phrase pairs were analysed with respect to Interrater agreement (IRA) and interrater reliability (IRR) [57]. IRA and IRR identify the compatibility and homogeneity of the results provided by raters. The human judgments have IRA and IRR in the interval of [0.61, 0.67]. Considering the type psychologist test, it is a fair agreement and it can be used as a gold standard for evaluating geo-semantic relatedness and similarity [54]. Due to advantages listed for this dataset and being open source, this dataset has been used in the following experiments.

4-3. Experimental setup

In order to evaluate the effectiveness of different topic-based models in the task of geo-semantic relatedness, these models have been evaluated and compared to state-of-the-art ontology-based models. In this section, the implementation details and configurations of each model used in our experiments are described individually. As there are multiple configurations for each model, it has been tried to choose the most efficient configuration according to previous researches.

As previously discussed in section 3, topic-based models perform based on processing background corpus and extracting topics according to terms occurrences in documents. Previous studies [8, 24] have indicated the importance of background corpus in performance of models. In recent years, Wikipedia as a comprehensive background corpus has attracted the attention of researchers in this field [23]. Therefore, in order to have a comprehensive evaluation between topic-based models, Wikipedia has been applied as background corpus in our experiments.

We leveraged an early spring 2013 of Wikipedia, containing about 4 million articles as the background corpus in the following experiments. Wikipedia is publicly available for download in an XML format[1]. This file is parsed using Wikipedia-Miner toolkit [58]. Upon removing small and overly specific concepts (those having fewer than 100 words and fewer than 5 incoming or outgoing links), the rest of articles were served for representing topics. The texts of these articles were processed by first tokenizing them, removing stop words and rare words (occurring in fewer than 3 articles), and stemming the remaining words using Porter algorithm [58]. The remaindered distinct terms served for representing each term as topic vectors. As described in Section 3.1, the importance of each term in each article has been calculated using TFIDF [36] weighting schema.

GenSim [59] was leveraged for implementing LSA and LDA models. By employing incremental algorithms, this tool makes these models able to be applicable on giant resources such as Wikipedia [23]. In order to compute SVD using Genism, Brand's algorithm [60] for fast incremental SVD updates has been used. Additionally, for the purpose of inferring topic distributions in LDA model, variational Bayes approximation [10] has been utilized.

As it was previously mentioned in Section 3, the number of topics in both LSA [9, 48] and LDA [10] must be determined in advance. According to previous successful researches [45], $k = 300$ topics have been selected for implementing each model.

4-4. Experimental results

In this section extensive experiments for evaluating various topic-based models in computing semantic relatedness of geographic phrases have been conducted. In order to present the priorities of topic-based models in comparison to other existing models, the performance of them have been compared to a wide range of ontology-based models.

The correlations between experimental results obtained by each model and scores determined by human judgment are presented in Table 1 based on Spearman (ρ) and Pearson (r) correlation coefficient. Previous researches [8, 54] revealed the importance of Spearman correlation in the task of word relatedness. Empirical results indicate that the performance of topic-based models is significantly better than ontology-based models in application of semantic relatedness of geographic phrases. Considering that the average correlation score between topic-based measure with human judgment ($\rho^*_{topic} = 0.64$) is significantly higher than the average correlation score between ontology-based models with human judgment ($\rho^*_{ontology} = 0.39$). The best ontology-based model has the correlation of $\rho_{path} = 0.45$ with human judgments, while the best topic-based model has the correlation of$\rho_{lsa} = 0.72$. The lowest correlation refers to Resink [30] and the highest one refers to LSA [9, 48].

It is notable that for computing semantic relatedness of geographic terms using ontology-based models, each phrase should be mapped into its corresponding concepts in WordNet. It is referred as entity linking and faced several challenges [50]. As it was presented in section 4.2, in GeReSiD [54] each phrase has been mapped to its corresponding concept in WordNet. For evaluating ontology-based models, these concepts have been used. In other word, considering challenges in automated methods for entity linking [50]; the efficiency of ontology-based models in real world will be lower than values presented in Table 1. While topic-based models do not require linked concepts and their results are more consistent to real world.

The correlation between the results generated by various topic-based models in comparison to each other is illustrated in Figure 2. The distribution of semantic relatedness estimated by each model is presented. As previously mentioned, the number of high/middle/low relatedness pairs in GeReSiD [54] comply the same distribution. Therefore, the histogram of value distribution determined by human judgments has a uniform distribution. In contrast, the histogram of values estimated by ESA[8], LSA [9] and LDA [10] models has skewness. This confirms that unlike the high correlation between semantic

[1] https://dumps.wikimedia.org/

relatedness values estimated by topic-based models and values estimated by humans; topic-based models are still unable to identify the semantic relatedness between some pairs. This can have an important role in development of these models in the future.

On the other hand, as illustrated in Figure 2, the values of semantic relatedness determined by various topic-based models follow same distribution and have high correlation to each other. So the values determined by different topic-based models often coordinate with each other. These empirical results imply the need to integrate these models into a unifying approach.

The values of semantic relatedness estimated by topic-based models with the values estimated by humans for each pair of phrases in GeReSiD [54] are shown in Figure 3. As it is illustrated in the chart, the values generated by topic-based models are lower than the actual values determined by human judgments. However, duo to the distribution of values of semantic relatedness estimated by topic-based models (Figure 2), it is predictable.

In order to show that the results are not likely to occur randomly, but rather are likely to be attributable to a specific cause, statistical hypothesis testing was employed to determine if the results are statistically significance or not. Statistical significance tests were calculated between the dependent Spearman correlations coefficients produced by topic-based models and ontology-based models. One tailed hypothesis test (topic-based models are better than ontology-based models) was used as an alternative test for assessing the difference between two paired correlations. It is worth nothing that the calculations rely on the tests implemented in the package *cocor*[2]. Statistical tests revealed that the results are considered statistically significant and the null hypothesis is rejected. Therefore, topic-based models are superior to ontology-based models and the results are caused by something other than mere random chance.

TABLE 1. THE CORRELATIONS AMONG DIFFERENT ALGORITHMS AND HUMAN JUDGMENTS BASED ON SPEARMAN (P) AND PEARSON (R) CORRELATION COEFFICIENT ON GERESID DATASET.

Type	Algorithm	Spearman's Correlation (ρ)	p-value	Pearson's Correlation (r)	p-value
Topic-based	ESA[8]	**0.68**	0.0000	**0.50**	0.0002
	LSA[9]	**0.72**	0.0000	**0.59**	0.0000
	LDA [35]	**0.52**	0.0001	0.39	0.0057
Ontology-based	HSO [21]	0.41	0.0033	0.39	0.0056
	Resnik [30]	0.26	0.0739	0.32	0.0224
	Lin [32]	0.39	0.0056	**0.45**	0.0011
	Jcn [31]	0.31	0.0266	0.38	0.0065
	Lch [19]	0.37	0.0087	0.37	0.0074
	Wup [20]	0.33	0.0183	0.34	0.0163
	Path [18]	**0.45**	0.0010	0.35	0.0125

[1] http://www.comparingcorrelations.org/

5. CONCLUSION

Different methods have been for computing semantic relatedness of term in recent years. These models are commonly unsupervised and employ the analysis of terms occurrences distribution in documents of a large corpus. In spite of the conceptual correlation between all these models, each model has its own architecture, potential and drawbacks and it has been also evaluated in particular domain application. According to these facts, (1) studying these models individually can be hard and challenging owing to the diversity of the architecture and the conceptual relatedness among them, (2) based on the evaluations of various models in different applications, comparing their efficiency is not simply possible.

ESA [8], LSA [9] and LDA [10] are known as three major indicators in this filed. To this end, all these models are presented in a unifying approach for the first time in this paper.

This unified approach provides readers with a common interpretation of these models in spite of the fundamental differences in their details. Additionally, to have comprehensive evaluations of these models, extensive experiments have been conducted in the field of geo-semantic relatedness. Moreover, the obtained results have been compared to an extensive range of ontology-based models. Based on the results, it can be concluded that:

Topic-based models are not only confront with fewer limitations in comparison to ontology-based models but also they are more applicable in real world problems. For instance, all ontology-based models require to automatically map each term to its related concepts in ontology, which yields to various challenges. In addition, creating and maintaining ontology in a particular domain costly and timely. On the other hand, topic-based models are not confronted with any particular limitations and they are more applicable to real world problems.

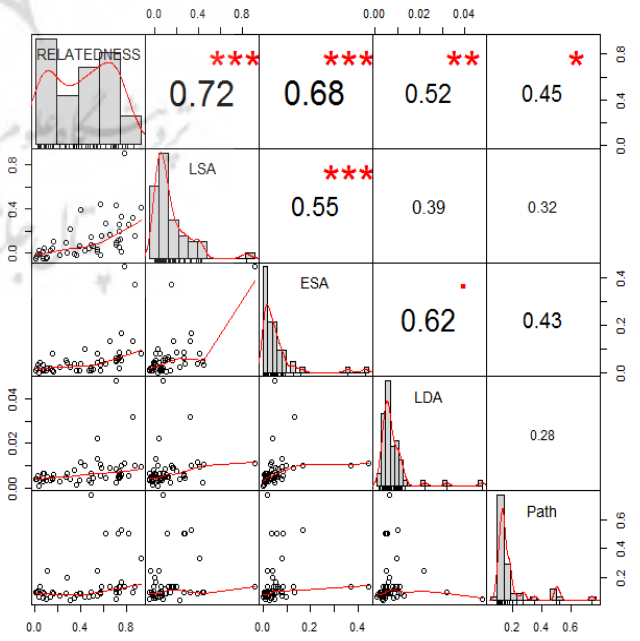


Fig 2. Scatterplot matrix of different topic-based models, with histograms, kernel density overlays, absolute Spearman's correlations, and significance asterisks (0.05, 0.01, 0.001

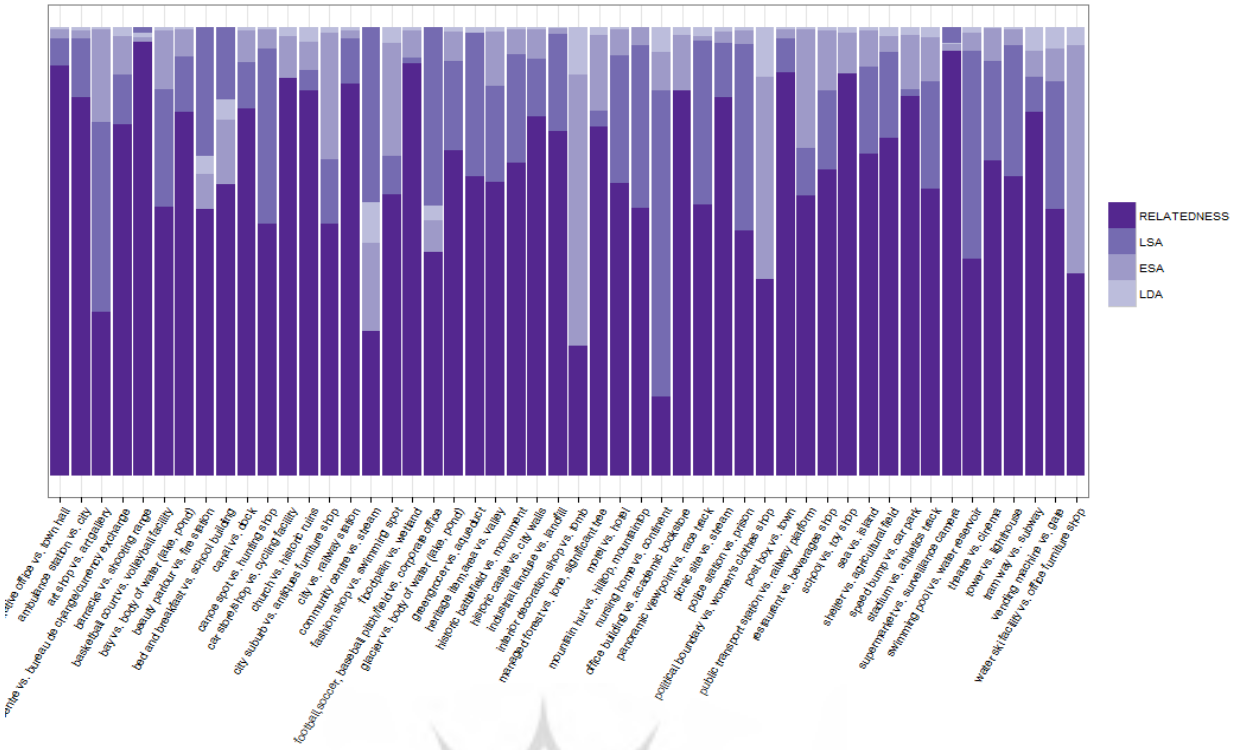


Fig 3: comparision of semantic relatedness values determined by humans and values estimated by topic-based models (ESA, LSA and LDA) for each pair of phrases in GeReSiD dataset

- Topic-based models have significantly better performance in computing semantic relatedness of geographic phrases compared to ontology-based models.
- The distribution of semantic relatedness values estimated by different topic-based models is similar to each other and there is significant correlation among them. Experimental results also confirmed the requirement for proposing a comprehensive unified approach for integrating these models.

Previous researches have shown that computing semantic relatedness of phrases is applicable in an extensive range of real world problems. Using semantic relatedness measures in solving different problems in the field of Geographic Information Science is one of the remarkable research areas. The results of this paper confirmed that not only topic-based models are confronted with less constraint compared to ontology-based models but also their performance in computing semantic relatedness of geographic phrases is significantly superior. Consequently, using topic–based models for computing sematic relatedness and their application in the field of Geographic Information Science can yield to an extensive range of future researches in this research area.

REFERENCES

[1] M. Kokla and E. Guilbert, "A Review of Geospatial Semantic Information Modeling and Elicitation Approaches," *ISPRS International Journal of Geo-Information*, vol. 9, no. 3, p. 146, 2020.

[2] Z. Chen and Y. Yang, "Semantic relatedness algorithm for keyword sets of geographic metadata," *Cartography and Geographic Information Science*, vol. 47, no. 2, pp. 125-140, 2020.

[3] H. Sadr, M. N. Soleimandarabi, M. Pedram, and M. Teshnelab, "Unified Topic-Based Semantic Models: A Study in Computing the Semantic Relatedness of Geographic Terms," in *2019 5th International Conference on Web Research (ICWR)*, IEEE, 2019, pp. 134-140.

[4] A. Mehler, R. Gleim, R. Gaitsch, W. Hemati, and T. Uslu, "From Topic Networks to Distributed Cognitive Maps: Zipfian Topic Universes in the Area of Volunteered Geographic Information," *Complexity*, 2020.

[5] M. N. Soleimandarabi, S. A. Mirroshandel, and H. Sadr, "A Survey of semantic relatedness measures," *International Journal of Computer Science and Network Solutions*, vol. 3, no. 2, pp. 1-11, 2015.

[6] M. N. Soleimandarabi and S. A. Mirroshandel, "A novel approach for computing semantic relatedness of geographic terms," *Indian Journal of Science and Technology*, vol. 8, no. 27, pp. 1-11, 2015.

[7] M. N. Soleimandarabi, S. A. Mirroshandel, and H. Sadr, "The Significance of Semantic Relatedness and Similarity measures in Geographic Information Science," *International Journal of Computer Science and Network Solutions*, vol. 3, no. 2, 2015.

[8] E. Gabrilovich and S. Markovitch, "Wikipedia-based Semantic Interpretation for Natural Language Processing," *Journal of Artificial Intelligence Research*, vol. 34, pp. 443-498, 2009.

[9] S. T. Dumais, "Latent semantic analysis," *Annual Review of Information Science and Technology*, vol. 38, no. 1, pp. 188-230, 2004.

[10] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993-1022, March 2003.

[11] D. M. Blei, "Probabilistic Topic Models," *Commun. ACM*, vol. 55, no. 4, pp. 77-84, April 2012.

[12] H. Sadr, M. Nazari Solimandarabi, and M. Mirhosseini Moghadam, "Categorization of Persian Detached Handwritten Letters Using Intelligent Combinations of Classifiers," *Journal of Advances in Computer Research*, vol. 8, no. 4, pp. 13-21, 2017.

[13] M. A. H. Taieb, T. Zesch, and M. B. Aouicha, "A survey of semantic relatedness evaluation datasets and procedures," *Artificial Intelligence Review*, pp. 1-42, 2019.

[14] S. A. Salloum, R. Khan, and K. Shaalan, "A Survey of Semantic Analysis Approaches," in *Joint European-US Workshop on Applications of Invariance in Computer Vision*, Springer, 2020, pp. 61-70.

[15] D. Degl'Innocenti, D. De Nart, M. Helmy, and C. Tasso, "Fast, accurate, multilingual semantic relatedness measurement using wikipedia links," in *Intelligent Natural Language Processing: Trends and Applications*: Springer, 2018, pp. 571-584.

[16] H. Sadr and M. Nazari Solimandarabi, "Presentation of an efficient automatic short answer grading model based on combination of pseudo relevance feedback and semantic relatedness measures," *Journal of Advances in Computer Research*, vol. 10, no. 2, pp. 17-30, 2019.

[17] A. H. Jadidinejad and H. Sadr, "Improving weak queries using local cluster analysis as a preliminary framework," *Indian Journal of Science and Technology*, vol. 8, no. 5, pp. 495-510, 2015.

[18] T. Pedersen, S. Patwardhan, and J. Michelizzi, "WordNet::Similarity: Measuring the Relatedness of Concepts," in *Demonstration Papers at HLT-NAACL 2004*, Stroudsburg, PA, USA, 2004, Association for Computational Linguistics, in HLT-NAACL--Demonstrations '04, pp. 38-41.

[19] C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," *WordNet: An electronic lexical database*, vol. 49, no. 2, pp. 265-283, 1998.

[20] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 1994, pp. 133-138.

[21] G. Hirst and D. St-Onge, "Lexical chains as representations of context for the detection and correction of malapropisms," *WordNet: An electronic lexical database*, vol. 305, pp. 305-332, 1998.

[22] S. P. Ponzetto and M. Strube, "Knowledge derived from wikipedia for computing semantic relatedness," *J. Artif. Int. Res.*, vol. 30, no. 1, pp. 181-212, October 2007.

[23] O. Medelyan, D. Milne, C. Legg, and I. H. Witten, "Mining meaning from Wikipedia," *Int. J. Hum.-Comput. Stud.*, vol. 67, no. 9, pp. 716-754, September 2009.

[24] Z. Zhang, A. L. Gentile, and F. Ciravegna, "Recent advances in methods of lexical semantic relatedness – a survey," *Natural Language Engineering*, vol. 19, no. 4, pp. 411-479, 2013.

[25] M. Andrea Rodriguez and M. J. Egenhofer, "Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure," *International Journal of Geographical Information Science*, vol. 18, no. 3, pp. 229-256, 2004.

[26] B. Hecht and M. Raubal, "GeoSR: Geographically explore semantic relations in world knowledge," in *The European Information Society*, Springer, 2008, pp. 95-113.

[27] Z. Lin, M. R. Lyu, and I. King, "MatchSim: a novel similarity measure based on maximum neighborhood matching," *Knowledge and information systems*, vol. 32, no. 1, pp. 141-166, 2012.

[28] A. Ballatore, M. Bertolotto, and D. Wilson, "Geographic knowledge extraction and semantic similarity in OpenStreetMap," *Knowledge and Information Systems*, vol. 37, no. 1, pp. 61-81, 2013.

[29] A. Ballatore and M. Bertolotto, "Semantically enriching VGI in support of implicit feedback analysis," in *Web and Wireless Geographical Information Systems*, Springer, 2011, pp. 78-93.

[30] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *arXiv preprint cmp-lg/9511007*, 1995.

[31] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," *arXiv preprint cmp-lg/9709008*, 1997.

[32] D. Lin, "An Information-Theoretic Definition of Similarity," in *Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco, CA, USA, 1998, in ICML '98, pp. 296-304.

[33] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler, "Exploring Topic Coherence over Many Models and Many Topics," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Stroudsburg, PA, USA, Association for Computational Linguistics, 2012, pp. 952-961.

[34] M. C. Lintean, C. Moldovan, V. Rus, and D. S. McNamara, "The Role of Local and Global Weighting in Assessing the Semantic Similarity of Texts Using Latent Semantic Analysis," in *FLAIRS Conference*, 2010. pp. 235-240.

[35] V. Rus, N. Niraula, and R. Banjade, "Similarity measures based on latent dirichlet allocation," in *Computational Linguistics and Intelligent Text Processing*: Springer, 2013, pp. 459-470.

[36] P. D. Turney and P. Pantel, "From frequency to meaning: vector space models of semantics," *Journal of Artificial Intelligence Research*, vol. 37, no. 1, pp. 141-188, January 2010.

[37] C. Mülligann, K. Janowicz, M. Ye, and W.-C. Lee, "Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information," in *Spatial Information Theory*, Springer, 2011, pp. 350-370.

[38] B. Hecht *et al.*, "Explanatory semantic relatedness and explicit spatialization for exploratory search," in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 2012, pp. 415-424.

[39] E. Gabrilovich and S. Markovitch, "Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis," in *Proceedings of The Twentieth International Joint Conference for Artificial Intelligence*, Hyderabad, India, 2007, pp. 1606-1611.

[40] C. Wang, J. Wang, X. Xie, and W.-Y. Ma, "Mining geographic knowledge using location aware topic model," in *Proceedings of the 4th ACM Workshop on Geographical Information Retrieval*, ACM, 2007, pp. 65-70.

[41] A. Ballatore, D. C. Wilson, and M. Bertolotto, "Computing the semantic similarity of geographic terms using volunteered lexical definitions," *International Journal of Geographical Information Science*, vol. 27, no. 10, pp. 2099-2118, 2013.

[42] H. Sadr, M. M. Pedram, and M. Teshnehlab, "A Robust Sentiment Analysis Method Based on Sequential Combination of Convolutional and Recursive Neural

Networks," *Neural Processing Letters*, vol. 50, no. 3, pp. 2745-2761 , 2019.

[43] H. Sadr, R. Atani, and M. Yamaghani, "The Significance of Normalization Factor of Documents to Enhance the Quality of Search in Information Retrieval Systems," *International Journal of Computer Science and Network Solutions*, vol. 2, no. 5, pp. 91-97, 2014.

[44] P. Zhang, S. Wang, D. Li, X. Li, and Z. Xu, "Combine Topic Modeling with Semantic Embedding: Embedding Enhanced Topic Model," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[45] D. Ștefănescu, R. Banjade, and V. Rus, "Latent Semantic Analysis Models on Wikipedia and TASA," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, N. C. Chair, Ed., European Language Resources Association (ELRA), May 2014, pp. 1417-1422.

[46] S. Liberman and S. Markovitch, "Wikipedia-based Compact Hierarchical Semantics with Application to Semantic Relatedness," Technion - Israel Institute of Technology, Technical Report CS-2010-06, March 2010.

[47] I. Assent, "Clustering high dimensional data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 4, pp. 340-350, 2012.

[48] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391-407, 1990.

[49] E. Agirre, O. López de Lacalle, and A. Soroa, "Random Walks for Knowledge-Based Word Sense Disambiguation," *Computational Linguistics*, vol. 40, no. 1, pp. 57-84, April 2013.

[50] J. Wang and J. Han, "Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 99, no. PrePrints, p. 1, 2014.

[51] O. Egozi, S. Markovitch, and E. Gabrilovich, "Concept-Based Information Retrieval using Explicit Semantic Analysis," *ACM Transactions on Information {S}ystems*, vol. 29, no. 2, pp. 1-34, 2011.

[52] H. Sadr, M. M. Pedram, and M. Teshnehlab, "Multi-View Deep Network: A Deep Model Based on Learning Features From Heterogeneous Neural Networks for Sentiment Analysis," *IEEE Access*, vol. 8, pp. 86984-86997, 2020.

[53] T. Zesch and I. Gurevych, "Wisdom of crowds versus wisdom of linguists–measuring the semantic relatedness of words," *Natural Language Engineering*, vol. 16, no. 01, pp. 25-59, 2010.

[54] A. Ballatore, M. Bertolotto, and D. Wilson, "An evaluative baseline for geo-semantic relatedness and similarity," *GeoInformatica*, pp. 1-21, 2014.

[55] L. Finkelstein *et al.*, "Placing Search in Context: The Concept Revisited," *ACM Transactions on Information Systems*, vol. 20, no. 1, pp. 116-131, January 2002.

[56] K. Janowicz, C. Keßler, I. Panov, M. Wilkes, M. Espeter, and M. Schwarz, "A study on the cognitive plausibility of SIM-DL similarity rankings for geographic feature types," in *The European Information Society*, Springer, 2008, pp. 115-134.

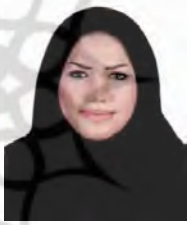
[57] J. M. LeBreton and J. L. Senter, "Answers to 20 questions about interrater reliability and interrater agreement," *Organizational Research Methods*, vol. 11, no. 4, pp. 815-852, 2008.

[58] D. Milne and I. H. Witten, "An open-source toolkit for mining Wikipedia," *Artificial Intelligence*, vol. 194, no. 0, pp. 222-239, 2013.

[59] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*, Valletta, Malta, 2010, pp. 46-50.

[60] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and Its Applications*, vol. 415, no. 1, pp. 20-30, 2006.

Hossein Sadr received his Ph.D. degree in Computer software Engineering from Islamic Azad University, Iran, in 2020, and his M.Sc. degree in the same filed from Islamic Azad University, Science and Research Branch, Iran, in 2013. He is also a member of Intelligent Systems Scientific Society of Iran (ISSSI). He is currently a lecturer in the Department of Computer Engineering at Islamic Azad University and is actively involved in the organization of a number of flagship conferences and workshops as well as cooperating as a reviewer with reputable journals, such as IEEE Access and Neural Processing Letters. His main areas of research are Natural Language Processing, Information Retrieval, Machine Learning, Deep Neural Networks, and Cognitive Science.
Email: Sadr@qiau.ac.ir



Mozhdeh Nazari Solimandarabi received her B.Sc. in computer software engineering and his M.Sc. degree in the same filed from Science and Research University. She is currently a Ph.D. student in the Department of Electrical and Computer Engineering at Qazvin Islamic Azad University. Her main areas of research are Natural Language Processing, Sentiment Analysis, Information Retrieval, and Machine Learning.
Email: Mozhdeh_nazary@yahoo.com

Mir Mohsen Pedram received his Ph.D. degree in Electrical Engineering from Tarbiat Modarres University, Tehran, Iran, 2003, his M.Sc. degree in Electrical Engineering from Tarbiat Modarres University, Tehran, Iran, 1994 and his B.Sc. degree in Electrical Engineering from Isfahan University of Technology, Isfahan, Iran, 1990. He is currently an Associate Professor in the Department of Electrical and Computer Engineering at Kharazmi University. His main areas of research are Intelligent Systems, Machine Learning, Data Mining, and Cognitive Science.
Email: Pedram@khu.ac.ir

Mohammad Teshnehlab received the B.Sc. degree from Stony Brook University, USA, in 1980, the M.Sc. degree from Oita University, Japan, in 1990, and the Ph.D. degree from Saga University, Japan, in 1994. He is a faculty member of Electrical Eng. Department of K. N. Toosi University of Technology. Professor Teshnehlab is a member of the Industrial Control Center of Excellence and founder of Intelligent Systems Laboratory (ISLab). He is also a co-founder and member of Intelligent Systems Scientific Society of Iran (ISSSI) and a member of the editorial board of the Iranian Journal of Fuzzy Systems (IJFS), International Journal of Information & Communication Technology Research (IJICTR), and Scientific Journal of Computational Intelligence in Electrical Engineering. His research areas are Artificial Rough and Deep Neural Networks, Fuzzy Systems and Neural Nets, Optimization, and Expert Systems.
Email: Teshnehlab@eetd.kntu.ac.ir