Hypothesis Generation From Text Based On Co-Evolution Of Biomedical Concepts

Kishlay Jha[1], Guangxu Xun[1], Yaqing Wang[2], Aidong Zhang[1]
[1]University of Virginia, Charlottesville, Virginia
[2]State University of New York at Buffalo, Buffalo, New York
[1]{kj6ww, gx5bt, aidong}@virginia.edu, [2]{yaqingwa}@buffalo.edu

ABSTRACT

Hypothesis generation (HG) refers to the task of mining meaningful implicit association between disjoint biomedical concepts. The majority of prior studies have focused on uncovering these implicit linkages from static snapshots of the corpus, thereby largely ignoring the temporal dynamics of medical concepts. More recently, a few initial studies attempted to overcome this issue by modelling the temporal change of concepts from natural language text. However, they still fail to leverage the evolutionary features of concepts from contemporary knowledge-bases (KB's) such as semantic lexicons and ontologies. In practice such KB's contain *up-to-date* information that is important to incorporate, especially, in highly evolving domains such as biomedicine. Furthermore, considering the complementary strength of these sources of information - corpus and ontology - a few natural questions arise: Can joint modelling of (co)-evolutionary dynamics from these resources aid in encoding the temporal features at a granular level? Can the mutual evolution between these intertwined resources lead to better predictive effects? To answer these questions, in this study, we present a novel HG framework that unearths the latent associations between concepts by modeling their co-evolution across complementary sources of information. More specifically, the proposed approach adopts a shared temporal matrix factorization framework that models the co-evolution of concepts across both corpus and KB. Extensive experiments on the largest available biomedical corpus validates the effectiveness of the proposed approach.

CCS CONCEPTS

• **Information systems applications → Data mining**;

KEYWORDS

hypotheses generation, co-evolution, biomedical domain

ACM Reference Format: Kishlay Jha, Guangxu Xun, Yaqing Wang, Aidong Zhang. 2019. Hypothesis Generation From Text Based On Co-Evolution Of Biomedical Concepts. In *The 25th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3292500.3330977

KDD '19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08…$15.00
https://doi.org/10.1145/3292500.3330977

1 INTRODUCTION

The constant influx of scientific articles and their easy accessibility via the World Wide Web (WWW) has made medical informatics a fast growing field [11]. Practitioners in the field have thrived to make sense of huge number of academic publications, discovery notes, electronic medical records and other text materials (a.k.a "big biomedical data") leading to advancements of practical significance [5]. While this swift availability of scientific information has acted as an impetus for pacing research innovation, at the same time, it has also overwhelmed researchers trying to survey published studies and construct novel research hypotheses. For instance, consider a novice researcher attempting to formulate a new hypothesis for the cures of *Diabetes*. In doing so, at this point in time, one might have to survey tens of thousands of existing publications (more than 400,000 in PubMed [11] alone) already written on *Diabetes*. This overloaded amount of information presents a bottleneck, as it is almost impossible for one to process and analyze such a large volume of available material. Moreover, it introduces delays in scientific productivity, as biomedical researchers are faced with a daunting task of choosing postulates/hypotheses - based upon the manual inspection of literature - for possible *in-vitro* clinical trials. To mitigate these issues, there has been a growing research interest among data/text mining practitioners to develop computational models that are able to assist biomedical experts in forging analytically probable and medically sensible hypothesis. Towards this end, Hypotheses generation (HG), a sub-problem of biomedical text-mining, aims to discover cross-silo connections (also known as undiscovered public knowledge) by chaining together the already known and established scientific facts that remain dispersed across the disparate research fields [18]. Simply put, given an input concept of interest (e.g., disease or gene), HG attempts to find implicit links (e.g., potential drug target or novel indicator of disease's mechanism) that connects them in a previously unknown but semantically meaningful way. Finding such meaningful associations is the crux of the problem that this paper attempts to address.

Over the past few decades, numerous studies have been conducted to tackle this problem. Broadly, they can be categorized into three major groups: a) distributional approaches [15, 22], b) graph-based methods [2, 20] , and c) supervised machine learning based approaches [19]. Arguably, these studies made significant advances, however, they still contain a few inherent drawbacks. First, a majority of these preceding approaches rely on a pre-defined structure (e.g., graph) and hence possibly risk missing links that are not included in their route. Second, almost all of these studies assume that the domain is static. This is limiting because it is known that the biomedical domain is a highly evolving field with new facts being added and old ones being obsolete every single day [4].

To overcome these issues, more recently, a few studies [9, 21] attempted to formulate this problem in latent space and generated hypotheses by modeling the temporal evolution of concepts based on the diachronic biomedical corpora. While these studies substantiated the importance of leveraging the temporal component, they still neglected the evolutionary features of concepts present in contemporary biomedical ontologies. Such ontologies/taxonomies in biomedical domain are constantly updated by subject-matter-experts to reflect the *up-to-date* knowledge of the field. Thus, to gain a holistic understanding of temporal change, it is crucial to factor in the semantic change of medical concepts from these subject-matter-experts maintained KB too. Furthermore, in practice, a significant amount of information is also encoded in the (co)-evolutionary dynamics of medical concepts between these complementary sources of information (i.e., corpus and ontology). Considering the complementary strength of both these resources, a few natural questions arise: Would the joint modelling of co-evolutionary dynamics lead to the generation of robust temporal embeddings? Would the mutual interaction between these intertwined resources simulate better predictive effects and thus benefit tasks such as hypothesis generation? To answer these questions, in this study, we model the co-evolution of medical concepts driven by the complex interaction between concepts' linguistic usage (reflected in local context information) and their structural localities (reflected in domain ontology). More specifically, we achieve this by adopting a shared temporal matrix factorization framework, wherein the subspaces between multiple related matrices are jointly learned by sharing information between them. By collaboratively exploiting the evolutionary features of medical concepts from both corpus and domain knowledge, the proposed approach yields hypotheses that are medically sensible and of potential interest to the domain experts. In this study, our contributions can be summarized as:

(1) We propose a general framework for the task of hypothesis generation that is capable of inferring previously unknown but potentially interesting cross-silo connections by capturing the subtle cues manifested in the temporal drift.
(2) The proposed approach for capturing the temporal change models the (co)-evolutionary dynamics of medical concepts across both the complementary sources of information - corpus and domain knowledge - thereby generating temporal embeddings that are robust and useful for a variety of downstream biomedical text-mining tasks.
(3) We propose an effective technique to leverage the evolving topological properties of biomedical KB, resulting in vector representations that encode the temporal dynamics at a granular level.

2 RELATED WORK

Discovering hidden, previously unknown and potentially useful associations between biomedical concepts is a problem of practical value in the research area of biomedical text-mining [14]. For a recent survey, please refer [3, 14]. The initial works [16, 18] in this area of study elucidated that the novel implicit links (e.g., *Fish Oils* $\xrightarrow{\text{treats}}$ *Raynaud's disease*) can be discovered by connecting independent nuggets of information remaining dispersed across the literature. While these pioneering studies laid the foundational groundwork, they were extremely time-consuming. Consequently, the subsequent studies focused on automating it. Primary studies such as [15, 22] applied statistical co-occurrence techniques (term frequency, inverse document frequency, record frequency and so on) to quantify the statistical strength between links. Similarly, [7, 22] adopted associate rule mining technique to estimate the strength of co-occurrences between concepts. While these purely co-occurrence based methods were progressive, a major drawback lies in their over-reliance on term frequencies. A greater statistical association implies strong but not necessarily semantically meaningful (real biological significance) association. To circumvent this drawback, we choose to model the problem of HG in latent space wherein the system is capable of capturing the implicit semantics between concepts, thereby finding connections that have greater semantic association.

Meanwhile, another line of research focused on modeling the problem of HG in a graph-based setting. Since graph based methods provide a natural way of representing concepts and their relationship, this line of research has attracted considerable attention. In [20], the authors presented a novel graph-based approach utilizing semantic predicates (subject-predicate-object), where subject/object refer to the entities (nodes) and predicates refer to the relationship (edge) between them. Another popular graph based HG system is *Obvio* [2]. Given a user input, *Obvio*, first constructs a graph on-the-fly and then uses the context information to automatically create semantically meaningful sub-graphs. One major contribution of this study is their ability to elucidate the meaning of complex associations between medical concepts along the multiple thematic dimensions. While graph-based approaches [2, 20] remain more successful than their distributional counterparts, they are still unable to find implicit connections. This is because the graph-based techniques still rely on a pre-defined structure/schema. More recently, some of the studies such as [19] applied supervised machine-learning based techniques to find the hidden connections. However, they require the domain expertise to generate the training data. This is both time-consuming and monetarily expensive. Despite important advances made, all of the aforementioned studies assumed the biomedical domain to be static. This is limiting because the domains in general (and in particular biomedicine) are usually dynamic with updates being made every now and then. To overcome this issue, recently, a few studies [9, 21] incorporated the temporal component by modelling the semantic evolution of medical concepts present in the historical biomedical corpus. However, these studies still neglect the semantic change of concepts from KB and thus fail to leverage the (co)-evolutionary dynamics of medical concepts.

Some of the motivation for this study stems from the research area of temporal network modelling [24, 25].While close in spirit, we differ from them in two aspects: a) Our focuses are different. b) Unlike modelling the temporal dynamics from multiple views of a network, in the current problem setting, our objective is to model the (co)-evolutionary features of medical concepts from their linguistic usage and structural localities in a concurrent manner.

3 METHODOLOGY

In this section, we describe our proposed framework in detail. Recall that the input to our hypothesis generation system is a topic of interest (A), date (d), and the goal is to predict previously unknown implicit links (C) at (d + 1). To tackle this problem, the key intuition behind our proposed approach is the following: If two medical concepts (A and C) are known to be primarily disjoint (i.e., no known relationship exists), and yet their implicit semantics continue to grow closer to each other over time, then these two terms have a higher chance of materializing a meaningful connection in the near future. In other words, our core objective is to capture the temporal 'proximity' between the medical concepts by modelling their semantic change over time. Generally speaking, this can be achieved by adopting a two-step solution: a) apply the temporal word embedding model [21] and generate the time-aware vector representations of concepts, b) quantify the degree of proximity between concepts by measuring the distance between their vector representations. While effective in practice, this class of techniques generate temporal embeddings in an isolated manner (e.g., corpus/ontology alone), and thus neglect the prevalent (co)-evolutionary features of medical concepts. To overcome this, in this study, we aim to generate the temporal embeddings that are infused with (co)-evolutionary dynamics generated due to the mutual influence of both complementary sources of information - corpus and ontology. Technically, we achieve this by adopting a shared temporal matrix factorization framework, wherein the sub-spaces between multiple related matrices are mutually learned by sharing the information between them. Further details on this are provided in the subsequent sub-sections. Section 3.1 and Section 3.2 introduce the two building blocks (modelling corpus-based and ontology-based evolutionary dynamics) of the proposed model. Then, in Section 3.3, we describe the joint co-evolution framework.

3.1 Corpus-Based Evolutionary Dynamics

To obtain the corpus-based temporal embeddings, we first need a text corpus collected across time (e.g., time-stamped scientific articles). Given this corpus, the objective is to generate the temporal word embeddings for each word present in the corpus. Traditionally, these temporal word embeddings could be generated by applying the neural network inspired language models such as Skip-gram (augmented with temporal component) [12] to the input sequential text. Simply put, the objective function of skip-gram is to predict the surrounding words within a fixed window, given a focus word. Following similar research direction, more recently in a related study [10], the authors proved that the objective function that the neural network attempts to solve in case of Skip-gram model with negative sampling is the same as the matrix factorization of Shifted Positive Point-wise Mutual Information (SPPMI) matrix obtained from the co-occurrence matrix of the corpus. As a result, the word and its corresponding context vectors can be obtained from the matrix decomposition of SPPMI matrix. This result is attractive as it enables the adoption/extension of techniques from the well-established area of matrix factorization. In this study, we utilize this equivalence result and propose a temporal matrix factorization based framework to obtain our temporal embeddings.

Formally, let us denote D_t as our time-stamped text corpus, where time-stamp t represents a discrete and ordered variable that varies from 1 to T. Given this corpus, we first collect all the concepts occurring in the corpus and prepare an overall vocabulary $V = \{w_1, ..., w_v\}$ of size $|V|$, where each w_i corresponds to an individual term. Note that this vocabulary is common to both the corpus and chosen ontology. Next, we construct a term-by-term $\mathbf{Y}(t)$ Pairwise Mutual-Information Matrix (PMI), whose i, j-th entry is:

$$PMI(i,j)_t = \log\left(\frac{\#(i,j)_t.|D_t|}{\#(i)_t.\#(j)_t}\right) \tag{1}$$

where $\#(i,j)_t$ counts the number of times the words w_i and w_j co-occurs within a document over the corpus D at time t, $\#(i)_t$ and $\#(j)_t$ denotes the total number of times words w_i and w_j occur in the corpus at time t alone. $|D_t|$ is the total number of word tokens in the corpus at time t. Following this, we compute the shifted positive point-wise mutual information matrix (SPPMI) specific to a corpus D at time t, whose (i,j)-th entry is:

$$SPPMI(i,j)_t = max(PMI(i,j)_t - \log k, 0) \tag{2}$$

where $\log k$ refers to a global constant. The constant $\log k$ acts as a prior on the probability of observing a positive example versus a negative example. A higher value of k implies that negative examples are more likely.

Following this idea, now our objective is to obtain a dense, low-dimensional vector representation $\mathbf{V}'(t) = [\mathbf{v}'_{w_1}(t), \mathbf{v}'_{w_2}(t), ..., \mathbf{v}'_{w_v}(t)] \in R^{|V|\times n}$, $n \ll |V|$ for each word $w \in V$, at each time-period t. $\mathbf{v}'_{w_i}(t)$ denotes the embedding vector for the i-th word at time-stamp t, and n is the number of dimensions. To achieve this, we adopt a standard matrix factorization framework and set up a least square optimization problem, so that the PPMI matrix $\mathbf{Y}(t)$ matches $\mathbf{U}.\mathbf{V}'(t)^T$ as closely as possible. The formulated optimization is shown below:

$$\min_{\mathbf{U},\mathbf{V}'(t)\geq 0} \sum_{t=1}^{T} \frac{h(t)}{2} ||\mathbf{Y}(t) - \mathbf{U}\cdot\mathbf{V}'(t)^T||_F^2 \tag{3}$$

Both $\mathbf{U}$ and $\mathbf{V}'(t)$ are $|V| \times n$ matrices. The main difference between $\mathbf{U}$ and $\mathbf{V}'(t)$ is that $\mathbf{U}$ is a constant matrix and $\mathbf{V}'(t)$ is a time-dependent matrix. While it is possible to make both $\mathbf{U}$ and $\mathbf{V}'(t)$ time-dependent, as shown in [24], a simpler model can achieve good approximation and also avoid over-fitting. The function $\mathbf{V}'(t)$ can take on any canonical form, such as linear models, polynomial models and so on. $h(t)$ refers to a decay function that regulates the importance between current and historical snapshots. This acts as an smoothing. The exponential function is chosen as a decay function with parameter $\theta > 0$.

$$h(t) = e^{-\theta(T-t)} \tag{4}$$

One challenge in this setting is that the PPMI matrix $\mathbf{Y}(t)$ is large and difficult to fit into memory. However, as most of the real-world networks are usually sparse, the computation can be made efficient. In most of the real-world scenarios, the presence of a co-occurrence

conveys more significant information than the absence of a co-occurrence. This is because the absence of a co-occurrence could mean: a) either there exists no association between the two concepts. b) there might exist a possible association between them in the near future. The presence of co-occurrence is seemingly more meaningful and thus the aforementioned objective function is adjusted to prioritize the presence of co-occurrence rather than the absence of co-occurrence. However, a small number of negative co-occurrence is needed to properly train the model. Suppose $E(t)$ be the set of word-pairs (w_i, w_j) such that the value of $y_{ijt} = 0$, and $F(t)$ be the set of word-pairs (w_i, w_j) such that the value of $y_{ijt} > 0$. Then, total set of co-occurrences is shown below:

$$G(t) = E(t) \cup F(t) \tag{5}$$

Now, one can express the objective function as:

$$\min_{\mathbf{U},\mathbf{V}'(t)\geq 0} \sum_{t=1}^{T} \frac{h(t)}{2} \sum_{(w_i,w_j)\in G(t)} (y_{ijt} - (\mathbf{U}\cdot\mathbf{V}'(t)^T)_{ij})^2 \tag{6}$$

Note that non-negativity is imposed on the factors for the purpose of greater interpretability.

3.2 Ontology-Based Evolutionary Dynamics

Ontologies/Hierarchies usually represented as Trees are known to provide a natural way of categorizing the knowledge of a particular domain. Such ontologies, also referred to as knowledge-bases (KB's), are abundantly present in the biomedical domain. Some common examples include Medical Subject Headings (MeSH), Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), and International Classification of Diseases (ICD9). These KB's are periodically updated by the subject-matter-experts in order to reflect the contemporary knowledge of the field. Given that these KB's are manually curated and showcase the prevailing knowledge of the field, our speculation is that integrating the evolutionary features of concepts from these resources will result in more accurate temporal representation of biomedical concepts. In our present study, the KB chosen is hierarchical (i.e., IS-A relationships) in nature (further details in experiments). Basically, the edges between concepts in the Tree denotes "parent-child" relationship, and the depth of a concept from the root indicates its level of specificity. Note that greater the depth of a concept in the tree the greater is its semantic richness. To leverage this valuable information, we adopt a technique similar to Section 3.1, and later extend our objective function. More specifically, we first convert the given hierarchical KB into a semantic distance matrix $\mathbf{M}(t)$[1], and then approximate the semantic distance matrix by the product of two smaller matrix.

$$\mathbf{M}(t) \approx \mathbf{U}.\mathbf{V}''(t)^T \tag{7}$$

Here both $\mathbf{U}$ and $\mathbf{V}''(t)$ are $|V|\times n$ matrices with $n << |V|$, n denotes the number of dimensions. The semantic distance matrix $\mathbf{M}(t)$ between concepts is calculated based on two factors: a) shortest path between concepts, and b) the depth of least common subsumer (LCS). The LCS refers to the immediate common parent of two concepts. Given two concepts w_i, w_j at time t, the distance between them is calculated by the formula below:

$$\begin{aligned} l_{ij} = \log_2([\text{path}(w_i, w_j) + 1] \\ * [D' - \text{depth}(lcs(w_i, w_j))]) \end{aligned} \tag{8}$$

where $\text{path}(w_i, w_j)$ is the shortest distance between concept w_i, w_j at time t, $\text{depth}(lcs(w_i, w_j))$ is the depth of $lcs(w_i, w_j)$ at time t, D' is the maximum depth of the taxonomy, and $lcs(w_i, w_j)$ is the lowest common subsumer of w_i and w_j. Prior research studies [13] have shown that the exploitation of these two factors is an effective strategy to leverage the ontology specific features. Having obtained our semantic distance matrix $\mathbf{M}(t)$, our next step is to generate the ontology-specific temporal embeddings. To do so, similar to Equation 3, the optimization problem is formulated as shown below:

$$\min_{\mathbf{U},\mathbf{V}''(t)\geq 0} \sum_{t=1}^{T} \frac{h(t)}{2} ||\mathbf{M}(t) - \mathbf{U}\cdot\mathbf{V}''(t)^T||_F^2 \tag{9}$$

Though intuitive, in practice, this basic formulation does not fully leverage the typical topological properties of given hierarchical KB. To overcome this issue, we propose an enhanced strategy that exploits the topological properties of the available taxonomy in a more effective manner. Basically, we consider a practical assumption that in the hierarchical KB, *the meaning of a particular concept is particularly influenced by its ancestors in the following order: direct-parents (strongest), grand-parents (stronger), higher-ancestors (lower) and root (least).* As an example, consider the concept "Diabetes Mellitus, Lipoatrophic". This concept forms its semantics by inheriting the basic properties from its ancestor concepts ("Diabetes Mellitus, Type 2", "Diabetes Mellitus", "Endocrine System" and "root")[2], and also adds its own specific properties. Accordingly, the vector representation of a concept w_i should be modelled by quantifying the semantic contribution for each of its ancestor w_{ij}. We define the strategy to quantify semantic contribution by exploiting the principles of label propagation [1, 25], usually adopted in network modeling tasks. Simply put, the idea in label propagation is to preserve the local spatial consistency of network by nudging the neighbourhood concepts to have similar feature vectors. Much alike, we mould its principles to fit the current hierarchical structure of KB, and argue that the features of a concept should be particularly influenced by their ancestors in accordance to their level of specificity.

$$b_{ij}^{(t)} = \frac{1}{\sqrt{\lambda}} \tag{10}$$

λ denotes the depth of ancestor concept (w_{ij}) in the tree. Note that the semantic contribution value of each concept changes over time based on their evolving structural locality. Having calculated the semantic contribution value, now, each concept in the tree adjusts (updates) its feature vectors based on its ancestors. Suppose that the initial feature vector of concept w_i is $\mathbf{v}_i(t)$, and the updated vector is $\mathbf{v}''(t)$ at timestamp t. Then, the feature vector update process from $\mathbf{v}_i(t)$ to $\mathbf{v}''(t)$ can be modeled by the following optimization problem.

[1]Note that the hierarchical KB is released every year and thus evolves over time.

[2]https://meshb.nlm.nih.gov/record/ui?ui=D003920

$$\min_{\mathbf{v}''(t)} \alpha \sum_i ||\mathbf{v}_i''(t) - \mathbf{v}_i(t)||^2 + (1-\alpha) \sum_{j \in Ancestors(w_i)} b_{jj}^{(t)} ||\mathbf{v}_i''(t) - \mathbf{v}_{ij}(t)||^2 \tag{11}$$

In the above Equation 11, the first term is known as the fitting constraint. This constraint penalizes large deviation from the initial feature vectors. The second term ensures that the feature vectors of concepts are updated in accordance to the semantic contribution of its ancestors. α balances the contribution of each part of the equation. As the formulation in Equation 11 is convex, its solution can be found by solving a system of linear equations. The closed updates are give below:

$$\mathbf{v}_i''(t) = (1-\alpha)(\mathbf{I} - \alpha \mathbf{B}(t))^{-1} \mathbf{v}_i(t) \tag{12}$$

where $\mathbf{I} \in R^{|V| \times |V|}$ is an identity matrix. $\mathbf{B}(t)$ is defined as the depth matrix. Next, we substitute the analytical solution of Equation 11 in Equation 9.

$$\mathbf{S}(t) = (1-\alpha)(\mathbf{I} - \alpha \mathbf{B}(t))^{-1} \tag{13}$$

$$\mathbf{V}''(t) = \mathbf{S}(t)\mathbf{V}(t) \tag{14}$$

$$\min_{\mathbf{U}, \mathbf{V}''(t) \geq 0} \sum_{t=1}^{T} \frac{h(t)}{2} ||\mathbf{M}(t) - \mathbf{U} \cdot \mathbf{S}(t) \cdot \mathbf{V}(t)^T||_F^2 \tag{15}$$

In this regard, one might ask: What is the necessity of adopting this route when the semantic distance matrix $\mathbf{M}(t)$ already captures the global hierarchical information? In our research we found two reasons for it: a) the strategy to exploit the typical ancestral property of a given concept acts as a "local regularization" and thus aids to leverage the taxonomic features in a more effective way. b) it provides a good initialization (generates basis vectors that are much closer to the best basis vectors found) for the Non-negative matrix factorization (NMF) formulation, resulting in improved convergence speed and accuracy.

3.3 Corpus-Ontology Based (Co)-Evolutionary Dynamics

Both Section 3.1 and Section 3.2 can obtain the temporal embeddings for biomedical concepts. The former exploits the local context information from natural language text and the later leverages upon the topological properties of given taxonomy. However, these two components should not be isolated from one another as they provide complementary sources of information. Furthermore, a significant amount of information is encoded in their (co)-evolutionary dynamics with respect to one another. To address this, we propose to jointly model the co-evolution of biomedical concepts from these interdependent sources of information. The objective function to be optimized is shown below:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}(t), \mathbf{V}'(t) \geq 0} \sum_{t=1}^{T} \frac{h(t)}{2} ||\mathbf{Y}(t) - \mathbf{U} \cdot \mathbf{V}'(t)^T||_F^2 \\ + ||\mathbf{M}(t) - \mathbf{U} \cdot \mathbf{S}(t) \cdot \mathbf{V}(t)^T||_F^2 \end{aligned} \tag{16}$$

As it can be observed, the first and second part of objective function models the temporal change of concepts from natural language text and ontology respectively. To facilitate the joint learning and mutual sharing of information, the latent factor $\mathbf{U}$ is shared by both parts of the objective function. As mentioned before, both $\mathbf{V}$, $\mathbf{V}'$ can take any canonical form (e.g., linear, polynomial and so on). For simplicity of the model, we choose a linear function. For instance: $\mathbf{V}(t) = \mathbf{X}t + \mathbf{Y}$. As $\mathbf{V}(t) \geq 0$, both $\mathbf{X} \geq 0$ and $\mathbf{Y} \geq 0$. Now, after adding regularization terms the expanded form of Equation 16 becomes:

$$\begin{aligned} J(\mathbf{U}, \mathbf{X}', \mathbf{Y}', \mathbf{X}, \mathbf{Y}) = \sum_{t=1}^{T} \frac{h(t)}{2} \sum_{(w_i, w_j) \in G(t)} (y_{ijt} - \mathbf{U} \cdot (\mathbf{X}'t + \mathbf{Y}')^T)_{ij} + \\ \sum_{(w_i, w_j) \in G(t)} (m_{ijt} - \mathbf{U} \cdot \mathbf{S}(t) \cdot (\mathbf{X}t + \mathbf{Y})^T)_{ij} + \frac{\beta}{2} ||\mathbf{U}||^2 \\ + \frac{\gamma_1}{2} ||\mathbf{X}||^2 + \frac{\omega_1}{2} ||\mathbf{Y}||^2 + \frac{\gamma_2}{2} ||\mathbf{X}'||^2 + \frac{\omega_2}{2} ||\mathbf{Y}'||^2 \end{aligned} \tag{17}$$

where $G(t)$ refers to the set of co-occurrence set as defined in Equation 6. The bound-constraint formulation of the above objective function is shown below:

$$\begin{aligned} &\min_{\mathbf{U}, \mathbf{X}', \mathbf{Y}', \mathbf{X}, \mathbf{Y}} J(\mathbf{U}, \mathbf{X}', \mathbf{Y}', \mathbf{X}, \mathbf{Y}) \\ &\text{subject to } \mathbf{U}, \mathbf{X}', \mathbf{Y}', \mathbf{X}, \mathbf{Y} \geq \mathbf{0} \end{aligned} \tag{18}$$

Next, we find the update rules for our cost function $J(\mathbf{U}, \mathbf{X}', \mathbf{Y}', \mathbf{X}, \mathbf{Y})$ with respect to each of the model parameters $\{ \mathbf{U}, \mathbf{X}', \mathbf{Y}', \mathbf{X}, \mathbf{Y} \}$ and run the stochastic gradient descent. The choice of optimization method is agnostic to the model and thus anything that successfully solves Equation 18 should generate quality temporal vector representations. Note that the update requires calculating inverse of a matrix (Refer Equation 13). This step is computationally expensive. Thus, to overcome this, we adopt an iterative approach (See below) similar to [1] and obtain our solution.

$$\mathbf{S}(t) = (1-\alpha) \sum_{b=1}^{B} (\alpha \mathbf{B}(t))^{b-1} \tag{19}$$

where B refers to the number of iterations. Once the iterative algorithm converges, we can obtain our time-aware embeddings as $\mathbf{V}'(t) = \mathbf{X}'t + \mathbf{Y}'$. As our vector representations are parameterized with time, it allows us to predict the future co-occurrence matrix $\mathbf{Y}(t+1) \approx \mathbf{U}\mathbf{V}'(t+1)^T$. The entry values in $\mathbf{Y}(t+1)$ quantify the likelihood of future association (hypothesis) between biomedical concepts. Now, given an input concept of interest (A), the candidate concepts (C) are ranked based on their predicted future co-occurrence value and then presented to the user for further analysis and investigation. Having described the nuances of our methodology, in the next section we describe our experimental protocol and perform extensive analysis to validate the effectiveness of proposed approach.

4 EXPERIMENTS

In this section, we demonstrate the efficacy of our proposed framework. Towards this end, we perform both qualitative and quantitative evaluations. The qualitative evaluation determines the extent to which our approach is capable of rediscovering the already known knowledge (and potentially new knowledge), whereas the quantitative evaluation is intended to analyze the overall quality of predictions/discoveries made by the system.

Dataset Description: MEDLINE[3], the largest available scientific repository, is used as the primary source of information for performing experiments. At present, it provides access to more than 24 million time-stamped articles primarily from the domain of life-sciences and bio-medicine. Among others, each article in MEDLINE contains the following attributes: a) unique identifier known as PMID, b) title, c) abstract, d) publication date and e) Medical Subject Headings (MeSH) terms. Previous studies [22] have shown that using concepts from raw title/abstract may introduce noise to the system and prove computationally expensive. To circumvent this problem, a majority of studies [7, 15, 23] conduct their investigation studies by choosing MeSH terms as their unit of analysis. MeSH terms in MEDLINE refer to a set of special keywords that are assigned to each article by the subject-matter-experts. As the experts annotate these terms based on the full-content of the article, they can be assumed to represent the conceptual meaning of an article. Being manually curated, they are highly accurate and find their utility in a multitude of downstream biomedical applications. Considering its high input quality and broader applicability, in this study, we use MeSH terms as our unit of analysis[4]. Fortunately, these MeSH terms are also arranged in a hierarchical/taxonomic structure[5]. In our study, this taxonomic structure of MeSH terms serve as our Knowledge-base. As of year 2018, there are approximately 28,000 MeSH terms (V). For our experiments, we generate the temporal embeddings for these medical concepts. As recommended in some of the prior studies [10, 12], we set the dimensionality of our temporal embeddings to $n = 200$. The hyper-parameter for exponential decay function is set to $\theta = 0.3$. The regularization weights $\beta = \gamma_1 = \gamma_2 = \omega_1 = \omega_2 =$ is 0.01. The value of α in Equation 11 is empirically set to 0.5. Finally, the number of iteration for model and the value of B in Equation 19 are both set to 200.

4.1 Qualitative evaluation

To perform qualitative assessment, we borrow experimental settings from the hypotheses generation literature [15, 23]. A common way of performing evaluation is to replicate the five golden test-cases (enumerated below) reported by the pioneers in this area of study. For the sake of uniformity, we adopt the same setting and run the proposed model on these test-cases and probe for the results.

(1) Raynaud's Disease (RD) and Fish Oils (FO) (**1985**)
(2) Migraine Disorder (MIG) and Magnesium (MG) (**1988**)
(3) Arginine (ARG) and Somatomedin C (IGF1) (**1994**)
(4) Alzheimer Disease (AD) Indomethacin (INN) (**1989**)
(5) Schizophrenia (SZ) and Calcium - Independent Phospholipase A2 (PA2) (**1997**)

To recapitulate our problem statement, the input to our hypothesis generation algorithm is a topic of interest (A) (e.g., Raynaud's disease), date (d) (e.g., 1985) and the goal is to find new biological relationships (C) (e.g., Fish Oils). The date (d) in the input acts as a cut-off threshold. Both the proposed model and baseline algorithms are run on the *pre-cut-off* segment (before date d) and the obtained results (predicted connections) are evaluated in the *post-cut-off* segment (after date d). To analyze the predicted results, we need a ground truth. However, there is no standard ground truth available and creating one remains an open problem [23]. Therefore, for the purpose of quantitative analysis, a supposedly ground truth is constructed. All those connections that co-occur with the input concept of interest in the post-cut-off segment but not in the pre-cut-off segment are assumed to be valid connections. These valid connections are ranked based on their TF-IDF co-occurrence score with the input concept of interest. The candidate set for target 'C' terms are all the concepts present in vocabulary besides - A and *Co-occur(A)*. *Co-occur(A)* refers to the set of terms that have co-occurred with A before the threshold date d. All the possible target terms are ranked based on their predicted co-occurrence score with the input concept of interest. Then, the top-k results are presented to the user (Refer Table 1). We would like to note the readers that the results in Table 1 are present both with/without pre-defined semantic filter. Semantic filters are needed because in the biomedical domain practitioners have a diverse range of interest. Some experts working in a specific area (ex: Genes or Drugs) might be interested only in those terms that have a possible genetic linkages or posses certain chemical properties. On the other hand, a novice biomedical scientist might have a general interest and is possibly looking for a surprising (or radical) connection. To handle this broad range of interest, Table 1 reports the target terms both with/without semantic filtering. Note that the semantic category information for biomedical concepts can be obtained from Unified Medical Language Systems (UMLS) [6]. To emphasize our focus on finding potential therapeutic preventions (and in the interest of space), we report results only for the semantic category "Drugs". Now, in the rest of this section, we discuss the ability of proposed model to rediscover the already known knowledge.

Raynaud's Disease (RD) and Fish Oils (FO): To replicate this knowledge, we seeded our HG system with input concept (A) as "Raynaud disease" and a date (d) as "1985". The objective is to find possible treatments (e.g, "Fish Oils") or other terms of biological significance in the top-k results. The top-k results for this and all other test cases are reported in Table 1, along with the evidences in the form of PMIDS. As it can be observed from Table 1, the target term "Fish Oils" in ranked 3. If we filter the terms by Semantic category "Drug", the term "Fish Oils" obtain rank 1.

Migraine Disorder (MIG) and Magnesium (MG): In 1988, the authors in [17] studied the possible linkage between "Migraine Disorder" and "Magnesium". In their conclusion, the authors reported eleven previously unknown connections. In our results (Refer Table 1), we found the target term at rank 5 (overall) and rank 2 (semantic filter - Drug) respectively.

[3] https://www.nlm.nih.gov/bsd/medline.html
[4] https://github.com/kishlayjha/hypotheses-generation-coEvolution
[5] https://www.nlm.nih.gov/mesh/intro_trees.html
[6] https://semanticnetwork.nlm.nih.gov

FO-RD	eicosapentaenoic acid (PMID: 2403828)	lipoproteins, vldl (PMID: 2536517)	fish oils (PMID: 1626282)	lipoproteins, hdl (PMID: 1626282)	glycerides (PMID: 2536517)	fish oils (PMID: 2536517)	oils (PMID: 2536517)	linolenic acids (PMID:19410347)	cardiolipins (PMID: 11080015)	lipid peroxides (No evidence)
MG-MI	tritium (PMID: 27140442)	adioisotopes (PMID: 26160074)	AMPA receptors (PMID: 25916335)	Technetium (PMID: 25533715)	magnesium (PMID: 24512583)	nicergoline (PMID: 24182946)	magnesium (PMID: 23808884)	benzamides (No evidence)	blood preeclampsia (PMID: 23634460)	iodine radioisotope (No evidence)
AD-INN	Amyloid beta-Peptides (PMID: 23028221)	Sulfones (PMID: 22746060)	Sulindac (PMID: 19486646)	Lactones (PMID: 18213383)	Diclofenac (PMID: 17627676)	receptors, prostaglandin (PMID: 16914866)	Cyclooxygenase Inhibitors (PMID: 16763096)	Sulindac (PMID: 16195368)	Ibuprofen (PMID: 16169124)	Nanocapsules (No evidence)
IGFI-ARG	Biomarkers (PMID: 29217318)	Somatomedin C (PMID: 29150243)	Indomethacin (PMID: 29098662)	IGF1 protein (PMID: 28302957)	Multiple Sclerosis (PMID: 28295976)	Interferon beta-1b (PMID: 28177374)	Galactosemias	Indomethacin (PMID: 27982500)	Fibroblasts (PMID: 27272689)	Galactosephosphate (PMID: 27225493)
SZ-PA2	Antipsychotic Agents (PMID: 29164477)	biomarkers (PMID: 28651698)	Phospholipase A2 (PMID: 28549837)	PLA2G6 protein (PMID: 27434078)	phosphatidic acid (PMID: 27333658)	PLA2G6 protein (PMID: 27095818)	Phospholipase A2 (PMID: 26938821)	Fatty Acids, Unsaturated (PMID: 26894921)	phosphatidylserines (PMID: 26160611)	Erythrocytes (PMID: 23587695)

Table 1: The terms in left and right block of the above table are top-5 terms with and without semantic filter ("Drugs")

Arginine (ARG) and Somatomedin C (IGF1): In this test-case, the authors [18] explored the relationship between a growth-regulating peptide (i.e., Somatomedin C) and an amino acid (i.e., Arginine). In our results, we found the target concept *Somatomedin C* at rank 2 (overall) and 3 (semantic filter - Drug) respectively.

Alzheimer Disease (AD) Indomethacin (INN): The objective of this case-study was to find a possible connection between Indomethacin (an anti-inflammatory agent) and Alzheimer Disease (a progressive disorder that cause memory loss and other mental issues) [18]. The target term "Indomethacin" is ranked 5 (Overall) and 2 (Semantic filter - Drug) respectively.

Schizophrenia (SZ) and Calcium - Independent Phospholipase A2 (PA2): Schizophrenia is a chronic disorder that affects person's ability to think, feel and reason clearly [18]. In our results, the target term Phospholipase A2 (PA2) was ranked 3 (Overall) and 2 (Semantic - Filter) respectively.

Discovery example for the case of Autism: In our experiments, we tried to analyze the results of proposed approach on new test cases. To do so, we choose a disease of biomedical significance: *Autism*. Autism is a serious development disorder found in children that impairs the ability to communicate and interact. We seeded our algorithm with input as "Autism", date (d) as "2014" and analyzed the top-k results. The top term found was "calcineurin" (a protein phosphate). Upon manually inspecting the medical literature, we found that there might exist an indirect link between the calcineurin and autism via terms such as "Bcl-2", "calmodulin" and "synaptic plasticity". Although clinical trails are needed to corroborate any hypothesis, several recent studies [8] suggest that these terms are of potential clinical interest.

From the results of above qualitative analysis, one can infer that the proposed HG system is able to successfully replicate the known knowledge and potentially discover new practical knowledge. While this form of evaluation provides insight into the quality of top-ranked results, a quantitative form of evaluation is necessary to gain an understanding of overall results.

4.2 Quantitative evaluation

The objective of this section is to examine the overall quality of prediction/discoveries generated. To achieve this, we split the corpus into pre-segment/post-segment (Refer Section 4.1), and obtain the ranked set for both generated connections and ground truth. Then, *Spearman coefficient* is used to measure the performance. As a post-processing step, all the trivial connections (check-tags [11] such as "humans", "male", "female" and so on) are removed from both the ground truth and predicted set. Next in this section, we report the quantitative results and discuss our findings on all the five test-cases enumerated in Section 4.1. In this regard, one might question: How is the performance of HG systems in test-cases other than the traditional five test-cases? To answer this, we choose 200 diseases of biomedical significance and conducted experiments using the same timeslicing scheme. Specifically, for each of these 200 diseases, we set the cut-off date to January 1, 2014, which resulted in a pre-cut-off set composed of 19,895,212 million documents published before January 1, 2014 and a post-cut-off set composed of 4,587,929 documents published after January 1, 2014. The results obtained are reported and analyzed later in this section.

Evaluation baselines for quantitative evaluation: To compare the performance of proposed model with existing hypothesis generation systems, the following six baseline algorithms are implemented.

(1) *Jaccard*: Jaccard is a popular link prediction technique. The formula to calculate the strength of association between two concepts is given below:
$Association(A, C) = \mid Count_A \cap Count_C \mid / \mid Count_A \cup Count_C \mid$,
where $Count_i$ refers the set of terms that co-occur with i.
(2) *Preferential Attachment*: Preferential Attachment is another classical link prediction technique. The formula to calculate preferential attachment is given below:
$Association(A, C) = \mid Count_A \mid + \mid Count_C \mid$, where $Count_i$ refers the set of terms that co-occur with i.
(3) *Arrowsmith*: Arrowsmith is a popular hypothesis generation system proposed in [18].
(4) *BITOLA*: BITOLA is a recent hypothesis generation algorithm proposed in [6].
(5) *Static Embeddings*: Static embeddings refers to the word embeddings generated from given corpus without incorporating any temporal component. The static embeddings are generated by training the standard CBOW [12] model on the entire MEDLINE corpus. All the hyper-parameters for CBOW are chosen as suggested in the study [12].
(6) *Dynamic MeSH Embedding* [21]: DME refers to a recent HG algorithm that models the semantic evolution of medical concepts from the diachornic biomedical corpora alone. It does not incorporate the (co)-evolving features of medical concepts from contemporary knowledge bases.

Note that the first two algorithms (Jaccard and Preferential Attachment) are from the link prediction literature. As we formulated the current task into a weighted link prediction problem, it is of

Table 2: Spearman's Correlation for FO-RD

Algorithm	k=200	k=800	k=1000	k=1500
Jaccard	0.012	0.011	0.017	0.102
Preferntial attachment	0.004	0.006	0.009	0.101
Arrowsmith	0.018	0.013	0.012	0.106
BITOLA	0.019	0.021	0.018	0.119
Static (No evolution)	0.027	0.031	0.019	0.127
DME (No co-evolution)	0.068	0.081	0.101	0.189
Proposed	**0.189**	**0.205**	**0.301**	**0.407**

Table 3: Spearman's Correlation for MG-MIG

Algorithm	k=200	k=800	k=1000	k=1500
Jaccard	0.017	0.023	0.009	0.109
Preferntial attachment	0.019	0.026	0.011	0.112
Arrowsmith	0.021	0.041	0.017	0.115
BITOLA	0.023	0.042	0.019	0.127
Static (No evolution)	0.034	0.061	0.027	0.136
DME (No co-evolution)	0.078	0.092	0.109	0.193
Proposed	**0.179**	**0.275**	**0.389**	**0.469**

Table 4: Spearman's Correlation for AD-INN

Algorithm	k=200	k=800	k=1000	k=1500
Jaccard	0.012	0.014	0.018	0.100
Preferential attachment	0.011	0.013	0.017	0.112
Arrowsmith	0.014	0.023	0.038	0.118
BITOLA	0.027	0.032	0.047	0.124
Static (No evolution)	0.036	0.045	0.101	0.137
DME (No co-evolution)	0.058	0.079	0.112	0.187
Proposed	**0.197**	**0.292**	**0.362**	**0.447**

Table 5: Spearman's Correlation for IGF1-ARG

Algorithm	k=200	k=800	k=1000	k=1500
Jaccard	0.018	0.026	0.013	0.101
Preferntial attachment	0.022	0.012	0.017	0.103
Arrowsmith	0.022	0.031	0.017	0.104
BITOLA	0.026	0.032	0.018	0.119
Static (No evolution)	0.033	0.082	0.028	0.157
DME (No co-evolution)	0.092	0.097	0.125	0.194
Proposed	**0.280**	**0.385**	**0.425**	**0.487**

Table 6: Spearman's Correlation for SZ-PA2

Algorithm	k=200	k=800	k=1000	k=1500
Jaccard	0.024	0.014	0.095	0.112
Preferntial attachment	0.023	0.015	0.017	0.121
Arrowsmith	0.089	0.029	0.102	0.136
BITOLA	0.092	0.031	0.108	0.143
Static (No evolution)	0.017	0.095	0.129	0.195
DME (No co-evolution)	0.098	0.164	0.157	0.278
Proposed	**0.187**	**0.224**	**0.384**	**0.416**

Table 7: Mean Average Precision@k for 200 disease

Algorithm	k=200	k=800	k=1000	k=1500
Jaccard	0.012	0.013	0.017	0.102
Preferntial attachment	0.011	0.012	0.015	0.103
Arrowsmith	0.018	0.011	0.012	0.106
BITOLA	0.019	0.021	0.018	0.119
Static (No evolution)	0.027	0.031	0.019	0.127
DME (No co-evolution)	0.068	0.081	0.101	0.189
Proposed	**0.185**	**0.262**	**0.392**	**0.435**

interest to compare the results with classical link prediction techniques.

Evaluation metrics for quantitative evaluation: Two evaluation metrics are used to quantify our results: 1) Spearman Coefficient@k and 2) Mean Average Precision (MAP@k).

Results: Table 2, 3, 4, 5, 6 reports the Spearman-Coefficient@k for each of the five golden datasets enumerated in Section 4.1. The value of K is gradually increased from top 200 to 1500 and results are reported. Table 7 reports the MAP@K by consolidating numbers across 200 diseases (excluding the five golden test-cases) of biomedical significance.

Discussion: From Tables 2, 3, 4, 5, 6 and 7 it can be observed that the proposed model consistently outperforms all the existing baselines in terms of both Spearman-Coefficient@K and MAP@K. This result indicates the ability of proposed framework to find semantically meaningful connections at top ranks. Analyzing the overall results from different perspectives, we detect various trends. First, the contemporary HG systems - ARROWSMITH and BITOLA - perform better than classical link prediction techniques. This highlights the challenges that are unique to HG task and encourages us to develop solutions tailored to HG. Second, we notice that though the contemporary HG algorithms perform better than link prediction techniques, they fall behind the Static embedding approach. Upon manual inspection of results, we found that this is mainly due to two factors: a) over reliance on co-occurrence statistics, b) failing to capture the implicit semantics of medical concepts. To elaborate, the baseline HG algorithms (Number 3 and 4) are purely distributional in nature. This results in promoting those terms that are "contextually generic". Contextually generic terms are those terms that co-occur frequently with the input concept of interest but have meager semantic meaning associated to them. For instance, consider the example of "Migraine Disorder". Some of the related terms that frequently co-occur with Migraine are "headache", "pain". While these terms are statistically associated to "Migraine", they have poor semantic association. As baseline HG algorithms rely strongly on statistical co-occurrence, these contextually generic terms are ranked higher. This proves counter-intuitive as these same terms are ranked lower in the ground truth. Another point we wish to highlight is that, as embeddings based approaches are capable of capturing the implicit semantics, they successfully promote those terms that have functional relationship with input concept of interest. Recall that the word embeddings can capture special features such as linear analogical relationships

vec("ibuprofen") - vec("headache") ≈ vec("treats"). This special feature provides leverage to embedding based techniques over other approaches. Third, we observe that a recent temporal embedding based approach [21] performs better than Static embedding [12]. This result highlights the importance of leveraging the semantic change of concepts for predictive tasks such as Hypothesis generation. Lastly, we would like to highlight that the proposed model outperforms the existing temporal embedding approach. This is because the existing temporal embedding based approach [21] fails to leverage the (co)-evolutionary features of medical concepts from contemporary KB's. In our experiments, we found that such subject-matter-expert maintained KB's have invaluable information and their incorporation is important to generate robust temporal embeddings. Furthermore, we noticed that the collaborative exploitation of semantics from natural language text and KB's proved particularly helpful for domain-specific (rare) words. As an illustration, consider the medical concept "Adioisotopes". This concept rarely co-occurs with "Magnesium" but is known to have strong semantic association with it. The recent temporal word embedding approach [21] (without external knowledge) fails to identify this term (and such domain-specific words in general) in top-ranks, due to the lack of sufficient statistical information. While such domain-specific words lack local-context information, their semantics can be mined from human curated KB's. As the proposed framework effectively leverages the KB's, such domain-specific terms are successfully promoted to higher ranks in our predicted set, thereby resulting in improved performance. In summary, from our both qualitative and quantitative experiments, we conclude that jointly leveraging the local-context information from natural language text and topological features from knowledge-base aids to generate temporal embeddings that are both robust and posses better predictive power, thereby, generating effective hypothesis.

5 CONCLUSIONS

In this study, we proposed a general framework for hypothesis generation that models the temporal (co)-evolution of biomedial concepts from two complementary sources of information - corpus and domain knowledge. By synthesizing the mutual evolution of concepts from these intertwined resources, the proposed model generates temporal embeddings that are both robust and posses higher predictive effects. Technically, the model achieves this by adopting a temporal co-factorization framework wherein the subspaces between multiple related matrices are learned by sharing a constant factor. Both qualitative and quantitative experiments conducted on the largest biomedical corpora validates the efficacy of the proposed approach, and suggests that the proposed framework has potential for generating new practical knowledge.

ACKNOWLEDGMENT

This work was supported in part by the US National Science Foundation (NSF) under grant IIS-1514204. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

[1] Yoshua Bengio, Olivier Delalleau, and Nicolas Le Roux. 2006. 11 Label Propagation and Quadratic Criterion. (2006).
[2] D. Cameron, R. Kavuluru, T. C. Rindflesch, A. P. Sheth, K. Thirunarayan, and O. Bodenreider. 2015. Context-driven automatic subgraph creation for literature-based discovery. *J Biomed Inform* 54 (Apr 2015), 141–57.
[3] Vishrawas Gopalakrishnan, Kishlay Jha, Wei Jin, and Aidong Zhang. 2019. A Survey on Literature Based Discovery Approaches in Biomedical Domain. *Journal of biomedical informatics* (2019), 103141.
[4] Anika Groß, Cédric Pruski, and Erhard Rahm. 2016. Evolution of biomedical ontologies and mappings: Overview of recent approaches. *Computational and structural biotechnology journal* 14 (2016), 333–340.
[5] Doug Howe, Maria Costanzo, Petra Fey, Takashi Gojobori, Linda Hannick, Winston Hide, David P Hill, Renate Kania, Mary Schaeffer, Susan St Pierre, et al. 2008. Big data: The future of biocuration. *Nature* 455, 7209 (2008), 47–50.
[6] Dimitar Hristovski, Carol Friedman, Thomas C Rindflesch, and Borut Peterlin. 2006. Exploiting semantic relations for literature-based discovery. In *AMIA annual symposium proceedings*, Vol. 2006. American Medical Informatics Association, 349.
[7] Xiaohua Hu, Xiaodan Zhang, Illhoi Yoo, Xiaofeng Wang, and Jiali Feng. 2010. Mining hidden connections among biomedical concepts from disjoint biomedical literature sets through semantic-based association rule. *International Journal of Intelligent Systems* 25, 2 (2010), 207–23.
[8] James Humble, Kazuhiro Hiratsuka, Haruo Kasai, and Taro Toyoizumi. 2019. Intrinsic spine dynamics are critical for recurrent network learning in models with and without autism spectrum disorder. *bioRxiv* (2019), 525980.
[9] Kishlay Jha, Guangxu Xun, Yaqing Wang, Vishrawas Gopalakrishnan, and Aidong Zhang. 2018. Concepts-Bridges: Uncovering Conceptual Bridges Based on Biomedical Concept Evolution. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. ACM, New York, NY, USA, 1599–1607. https://doi.org/10.1145/3219819.3220071
[10] Omer Levy, Yoav Goldberg, and Ido Dagan. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics* 3 (2015), 211–225.
[11] Zhiyong Lu. 2011. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database* 2011 (2011), baq036.
[12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[13] Hoa A Nguyen and Hoa Al-Mubaid. 2006. New ontology-based semantic similarity measure for the biomedical domain. In *Granular Computing, 2006 IEEE International Conference on*. IEEE, 623–628.
[14] Yakub Sebastian, Eu-Gene Siew, and Sylvester O Orimaye. 2017. Emerging approaches in literature-based discovery: techniques and performance review. *The Knowledge Engineering Review* 32 (2017).
[15] Padmini Srinivasan. 2004. Text mining: Generating Hypotheses from MEDLINE. *J. Assoc. Inf. Sci. Technol.* 55, 5 (2004), 396–413.
[16] Don R Swanson. 1986. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in biology and medicine* 30, 1 (1986), 7–18.
[17] Don R Swanson. 1988. Migraine and magnesium: eleven neglected connections. *Perspectives in biology and medicine* 31, 4 (1988), 526–557.
[18] Don R Swanson and Neil R Smalheiser. 1997. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial intelligence* 91, 2 (1997), 183–203.
[19] D. Weissenborn, M. Schroeder, and G. Tsatsaronis. 2015. Discovering relations between indirectly connected biomedical concepts. *J Biomed Semantics* 6 (2015), 28.
[20] B. Wilkowski, M. Fiszman, C. M. Miller, D. Hristovski, S. Arabandi, G. Rosemblat, and T. C. Rindflesch. 2011. Graph-based methods for discovery browsing with semantic predications. *AMIA Annu Symp Proc* 2011 (2011), 1514–23.
[21] Guangxu Xun, Kishlay Jha, Vishrawas Gopalakrishnan, Yaliang Li, and Aidong Zhang. 2017. Generating Medical Hypotheses Based on Evolutionary Medical Concepts. In *IEEE 17th International Conference on Data Mining, ICDM 2017, December 18-21, 2017, New Orleans, USA*.
[22] Meliha Yetisgen-Yildiz and Wanda Pratt. 2006. Using statistical and knowledge-based approaches for literature-based discovery. *Journal of biomedical informatics* 39, 6 (2006), 600–611.
[23] Meliha Yetisgen-Yildiz and Wanda Pratt. 2009. A new evaluation methodology for literature-based discovery systems. *Journal of biomedical informatics* 42, 4 (2009), 633–643.
[24] Wenchao Yu, Charu C Aggarwal, and Wei Wang. 2017. Temporally factorized network modeling for evolutionary network analysis. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 455–464.
[25] Wenchao Yu, Wei Cheng, Charu C Aggarwal, Haifeng Chen, and Wei Wang. 2017. Link Prediction with Spatial and Temporal Consistency in Dynamic Networks.. In *IJCAI*. 3343–3349.