Exploring Diachronic Changes of Biomedical Knowledge using Distributed Concept Representations

Gaurav Vashisth[1,2], **Jan-Niklas Voigt-Antons**[1,2], **Michael Mikhailov**[1] **and Roland Roller**[1]
[1]German Research Center for Artificial Intelligence (DFKI), Berlin, Germany
[2]Technische Universität Berlin
firstname.lastname@dfki.de

Abstract

In research best practices can change over time as new discoveries are made and novel methods are implemented. Scientific publications reporting about the latest facts and current state-of-the-art can be possibly outdated after some years or even proved to be false. A publication usually sheds light only on the knowledge of the period it has been published. Thus, the aspect of time can play an essential role in the reliability of the presented information. In Natural Language Processing many methods focus on information extraction from text, such as detecting entities and their relationship to each other. Those methods mostly focus on the facts presented in the text itself and not on the aspects of knowledge which changes over time.

This work instead examines the evolution in biomedical knowledge over time using scientific literature in terms of diachronic change. Mainly the usage of temporal and distributional concept representations are explored and evaluated by a proof-of-concept.

1 Introduction

Scientific literature presents knowledge for a particular time period it has been published. Various studies have been performed to explore such knowledge from scientific literature, where work by Swanson (1986) led to the discovery of a new drug to treat Raynaud's disease. Similarly, a study by Zhu et al. (2013) has concluded that drug discovery using scientific literature plays a pivotal role in the treatment of cancer, which can improve the quality of life of patients (Cummings et al., 2011). Although scientific literature is an excellent source of information, there has been an explosion in the number of publications each year. This poses a challenge for biomedical researchers and practitioners to keep themselves informed of recent developments. The increasing number at the same time provides an opportunity to automatically explore the data on how a change in knowledge has evolved. Some studies have tried to explore such changed knowledge by investigating temporal information (Zhou and Hripcsak, 2007; He and Chen, 2018), studying the diachronic change in the meaning of the word. A diachronic semantic change in language is associated with progression in the meaning of the word which is estimated by exploring its usage over time.

This work aims to automatically explore the advances in medical knowledge extracted from the abstracts of scientific research by using word/concept embeddings. Especially, we examine how treatments of pathological conditions have changed over time. For this reason we focus on concepts rather than words, as biomedical concepts can be mentioned in text in different ways (e.g. '*headache*', '*cephalgia*' or '*pain in the head*'). Moreover, biomedical concepts help to encapsulate noun phrases represented by more than one word, for example, '*eye lens*' or '*lung cancer*'. An analysis on word level instead would take all situations the single words occur into account, and therefore would be more general. To quantify such changes we measure how the usage of a biomedical concept has (semantically) changed over time by comparing different embedding periods.

The rest of the work is structured as follows: The next section presents related work in the context of diachronic changes in and outside the biomedical domain. Then, in Section 3 we present how the biomedical concept embeddings are generated and how the time aspect is taken into account. Section 4 shows the usage of our embeddings to explore diachronic changes as a proof-of-concept. Then we apply the temporal embeddings to explore some exemplary relational data of UMLS, followed by a conclusion.

Proceedings of the BioNLP 2019 workshop, pages 348–358
Florence, Italy, August 1, 2019. ©2019 Association for Computational Linguistics

2 Related Work

Human language is a complex system which has been evolving from the point of its origin whether it is because of social or cultural (Hamilton et al., 2016a,b) or technological (Phillips et al., 2017) reasons. Some words acquire new meaning much faster than other words (Blank, 1999) for example words like *broadcast*, *gay*, and *awful* have been used in a different context in the present time as compared to an earlier time.

To study the semantic change for words, initially, co-occurrence matrices (Sagi et al., 2009; Wijaya and Yeniterzi, 2011; Jatowt and Duh, 2014), K-means clustering (Wijaya and Yeniterzi, 2011), Frequency-based methods (Kulkarni et al., 2014) were used. Representations using co-occurrence matrices are based on the notion of word co-occurring in the same context. The co-occurrence matrix assumes that words occurring in same context tend to have the same meaning (Firth, 1957) and are represented by methods such pointwise mutual information (Turney and Pantel, 2010), Singular Value Decomposition matrices, and Latent Semantic Analysis.

Another popular method to represent words are distributed representations. Words are represented in a dense and continuous form, that enables us to capture the meaning in a condensed form. There are various methods such Word2Vec (Mikolov et al., 2013b,a) and Global Vectors for Word Representation (Glove) (Pennington et al., 2014) which create a distributed representation of words. Distributed methods consume less memory compared to co-occurrence matrices because of their compact size and ranges between 100 dimensions to 1000. Moreover, the distributed methods are robust baseline methods with their proven success in capturing linguistic meaning (Mikolov et al., 2013b).

Kim et al. (2014) explored the temporal changes in the meaning of word using Skip-gram negative sampling (SGNS) method. To generate word embedding for each time frame the embeddings from previous time frame was used to initialize the embedding for the next successive time frame. Hamilton et al. (2016b) try to answer two questions, first whether the frequency of a word affects the change in meaning, which has been long studied (Bybee et al., 2007; Pagel et al., 2007; Lieberman et al., 2007). Second, whether there is a relationship between a polysemous and semantic change of a word.

Also in the biomedical domain semantic changes in scientific abstracts have been explored (Yan and Zhu, 2018). In the study, the authors explored semantic changes for a set of words using their occurrence frequency and their distribution across different topics. Scientific literature has also motivated studies using biomedical concepts instead of free text; however, they only measure the similarity and relatedness between different concepts using different embedding methods (De Vine et al., 2014; Choi et al., 2016; Liu et al., 2018; Beam et al., 2018).

Our study draws motivation from previous studies. However, different to other work we try to explore diachronic change using biomedical concepts. Particularly we would like to use diachronic change to assist the exploration of knowledge changes in the biomedical domain.

3 Temporal Concept Embeddings

In the following the generation of the biomedical temporal concept embeddings used to identify semantic changes is introduced.

3.1 Data Resources

The MEDLINE repository[1] is a bibliographic database from life sciences containing around 26 millions articles dating back to 1809. MEDLINE is quickly growing as the number of publications added to the repository each year are increasing (see Figure 1). Title and abstracts within the MEDLINE repository define the source to generate the embeddings in this work.

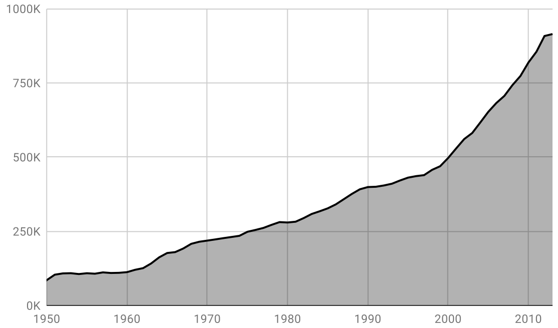


Figure 1: Number of MEDLINE abstracts published each year on PUBMED between 1950 and 2014.

Another relevant resource is the Unified Medical Language System (UMLS) (Bodenreider,

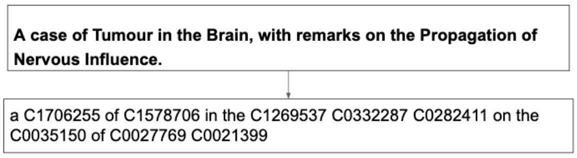
[1] https://mbr.nlm.nih.gov/

2004), a biomedical knowledge base which defines a large number of biomedical concepts and their relations to each other. Each concept is represented by a unique concept identifier known as CUI and includes word variations and synonyms. As we focus on the generation of concept embeddings, we normalize text mentions from MEDLINE abstracts to UMLS.

The concept normalization is carried out using MetaMap (Aronson, 2001), a popular named entity recognition system for biomedical text. However, to avoid processing millions of sentences with MetaMap, we use the MetaMapped 2015 MEDLINE Baseline Results, a MEDLINE subset already enriched by MetaMap Machine Output (MMO). In addition to that, we also use annual baseline files from the MEDLINE/PUBMED Baseline Repository (MBR) which contain meta-information about each publication such as publication ID, publication year and author name(s).

3.2 Data Preprocessing

First, publications from MMO are enriched with publication year (PubYear) by using the publication ID and the information from the MBR files. Then, the text occurrence of each medical abstract and its title are mapped and replaced with their concept ID, using the offset information provided in MMO (Figure 2). In this way, we create a text to train our embeddings. Since we do not consider character embeddings, we can treat concept IDs as words without any disadvantage.



A case of Tumour in the Brain, with remarks on the Propagation of Nervous Influence.

a C1706255 of C1578706 in the C1269537 C0332287 C0282411 on the C0035150 of C0027769 C0021399

Figure 2: Shows mapping of medical text to their corresponding concept ID for Publication ID:20895112.

To create temporal embeddings, the preprocessed MEDLINE abstracts are split into different time depended subsets using PubYear. Embeddings are then trained on those splits. Ideally, we would like to train models using equally sized time ranges, such as embeddings per year or decade. However, this is not easily possible for various reasons: Firstly, as seen in Figure 1 the number of publications is constantly increasing. A consistent split into equal time frames would result in highly unbalanced splits regarding the number of included abstracts. In addition to that PUBMED includes mainly titles and no abstracts before 1975, which further reduces the number of text for the lower represented period.

Period	# Publications
1809-1970	3,374,099
1971-1975	1,162,030
1976-1980	1,346,833
1981-1985	1,528,475
1986-1990	1,863,659
1991-1995	2,065,386
1996-2000	2,297,006
2001-2005	2,938,855
2006-2010	3,721,166
2011-2012	1,762,603
2013-2015	1,283,218

Table 1: Distribution of publications in each period

Conversely, the generation of equally sized splits (according to the number of abstracts/sentences) has the disadvantage that it will be more challenging to differentiate between particular years. Rounding up or down the number of included publications might also be not a satisfying solution, as the time ranges might differ too much. For this reason, we mainly focus on time range splits including 5 years of MEDLINE abstracts. As the number of publications is lower at the beginning of the 20th century and publications often do not contain any abstract, we combine the 'early' MEDLINE data into one big split (1809-1970). Moreover, as the number of publications steadily increase we create smaller splits from 2011. The final split into periods is presented in Table 1, including their corresponding number of abstracts.

3.3 Temporal Embeddings

To generate temporal embeddings, we use FastText (Bojanowski et al., 2016) in Skip-gram negative sampling (SGNS) mode, which predicts context words corresponding to a given target word occurring in its neighborhood. The values of the hyperparameters base on the recommendation of Levy et al. (2015). The authors did an extensive set of experiments using different representation methods and analyzed the effect of hyperparameters on the embeddings generated by them. We chose negative samples as 10, the minimum occurrence of concepts is 5, learning rate as .05, sampling threshold as .0001, dimension to 300 and context window to 10.

The different temporal embeddings were trained sequentially, starting from the first period (1809-1970) and ending with (2013-2015). We started the training of the first period with random initialization of the embeddings. All other embeddings were then initialized by the values of the former time embedding. This incremental training process has been applied as the training of a particular time period can build on the knowledge seen in earlier periods. Incremental training can be seen as an analogy of how human knowledge evolves over time. The temporal concept embeddings used in this work can be downloaded here[2].

3.4 Measuring Semantic Changes

To measure the semantic change between a concept pair we use cosine distance (similarity) at different periods as also described in Hamilton et al. (2016b). A cosine distance closer to 1 shows a stronger similarity/relation between the two concepts than a distance closer to 0. In this work, however, we are particularly interested in examining whether the semantic shift can be used to explore how treatments (of particular diseases) evolved. Therefore, we selected particular concept pairs and explore how their similarity score evolves.

In addition to cosine similarity we use Positive Pointwise Mutual Information (PPMI) matrix (Levy and Goldberg, 2014), as reference measure. A PPMI matrix is a variant of Pointwise Mutual Information (PMI) and provides an association between two words occurring together in a corpus and how strongly they are related to each other (Church and Hanks, 1990). When a specific word pair co-occurs more frequently they have a higher PPMI score and vice-versa. PPMI is still widely used co-occurrence matrix method and in this work we have used a normalized PPMI score which ranges between 0 to 1, whereas 1 indicates more frequent pairs.

4 Exploring Biomedical Knowledge Changes

In this section, we examine the usage of temporal concept embeddings to detect diachronic changes in the context of altering knowledge in biomedical literature. Particularly, we explore whether the embeddings reveal known changes in treatments in biomedical history, as a proof-of-concept. For instance, we would like to know whether it is possible to see a relative change in terms of cosine similarity, i.e., if a preferred treatment for some *Disease X* changes at time *t* from one medication to a new one (see example in Figure 3). Our assumption is that the usage of temporal concept embeddings reveal a similar pattern. Before time *t* we assume, that the old treatment has got a higher cosine similarity compared to the new treatment. And then after some decrease the new medication outperforms the other one. In the following, we will explore this phenomenon based on various examples.

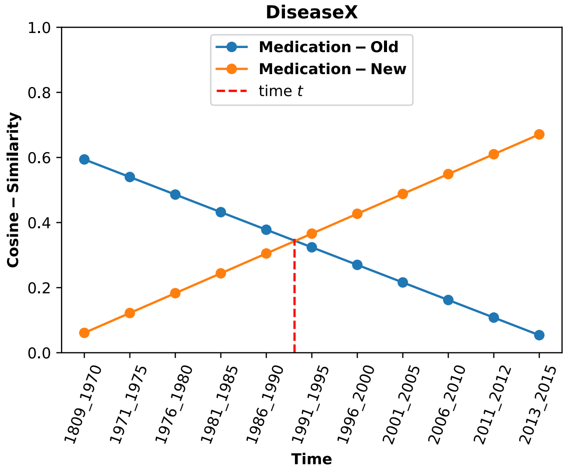


Figure 3: Shows a treatment change of some Disease X from *Medication-old* to *Medication-new* at time *t*.

4.1 Proof-of-Concept

In this section, different examples are presented to explore the usage of temporal concept embeddings to detect knowledge changes. We use those examples as a proof-of-concept. Each example includes a high-level introduction, followed by an investigation of the similarity scores over time and an explanation of the presented results.

In order to provide reliable insights, presented results are supported through a significance test (Welch's T-test) using a confidence interval of 99% (p_value $<$ 0.01). The significance test relies on 15 different complete sets of temporal embeddings (all periods) which were trained from scratch.

[2] http://biomedical.dfki.de/

4.1.1 Minoxidil

Minoxidil (CUI=C0026196) is a medication, initially used for treating high blood pressure (Hypertension) (Stoehr et al., 2019) . Nowadays Minoxidil is still used as a drug of last resort for treatment of resistant hypertension (remains above a target level, in spite of being prescribed three or more anti-hypertensive drugs simultaneously with different mechanisms of action). However, in 1988 FDA approved the medication also for treating hair loss problems. Presently, Minoxidil is used mainly to treat early baldness pattern such as *Androgenic Alopecia* (C0162311) and *Scalp Hair Loss* (C0574769).

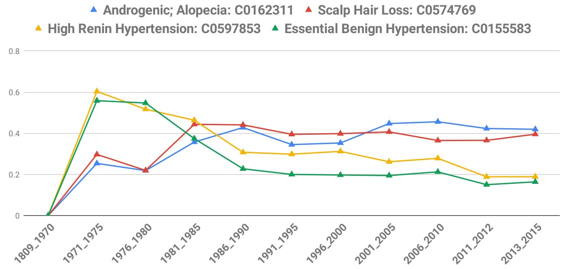


Figure 4: The similarity score of minoxidil with medical conditions from 1809 to 2015. Change in usage ocurs in 1986. Where **Old Usage** was *High Blood Pressure* and **New Usage** is *hair fall*

The exploration of the cosine similarity for minoxidil and its change in treatment is presented in Figure 4. The figure depicts a high similarity to hypertension in the '70s, which is significantly higher than the high blood pressure. However, after 1980 the similarity slowly decreases in the next following years. Around 1985 we can see a big drop. At the same time the similarity of *alopecia* and *scalp hair loss* strongly increase around 1985. From the following period, the similarity score of both concepts outperforms hypertension and are significantly higher than hypertension.

4.1.2 Microprolactinoma

Microprolactinoma (Prolactinoma)[3] is a type of benign tumor that occurs in the pituitary gland of the brain (Casanueva et al., 2006; Glezer and Bronstein, 2015). Its treatment has changed notably over time. Until the 1970's this tumor was removed by a surgical method known as *Transethmoidal Hypophysectomy* (C0405509) (Richards et al., 1974). Beginning from the late 1970's a new class of medical therapy with *Dopamine Agonists* was introduced to treat Microprolactinoma (C0344452) without having to undergo a surgery. *Dopamine Agonists* is a class of drugs that activate dopamine receptors. The treatment using *Dopamine Agonists* has a cure rate of more than 80%. The most effective *Dopamine Agonists* used as a main treatment drugs are *Cabergoline* (C0107994) and *bromocriptine* (C0006230) (Tirosh and Shimon, 2016; Glezer and Bronstein, 2015) which are D2 dopamine agonists that inhibit prolactin secretion. Only if patients do not respond to medications, a surgical method called *Transsphenoidal surgery*[4] (C2985562) is used (Tirosh and Shimon, 2016).

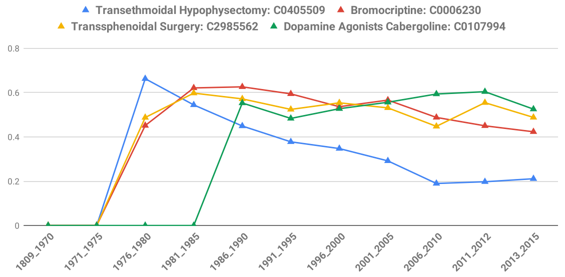


Figure 5: Similarity Score of Microprolactinoma with different treatment methods from 1809 until 2015. Change in the medication occurs after 1976. Where **Old Method** was *Transethmoidal hypophysectomy* and **New Methods** are *Transsphenoidal surgery*, *Bromocriptine*, *dopamine agonists cabergoline*.

Figure 5 presents the semantic shift in the use of different treatment methods for Microprolactinoma. The first embedding point is seen from the period 1976-1980. Before that period none of the concepts occurred frequently enough to be considered in the embedding. Within the first period of occurrence (1976-1980) we have a significantly higher similarity score of Microprolactinoma with *Transethmoidal Hypophysectomy* in comparison to *Bromocriptine* concepts and *Transsphenoidal surgery* which starts decreasing in the next following years. After 1980, we see an increase in the similarity for all the *bromocriptine* concepts along with *Transsphenoidal surgery*, which shows a change in the treatment method for Microprolactinoma. The similarity score of both the *Bromocriptine* and *Transsphenoidal surgery* concepts have significantly higher similarity score than *Transethmoidal Hypophysectomy* from 1981. Whereas

[3] Microadenoma of a pituitary gland

[4] A surgical method used to remove tumors of pituitary glands.

from 1986, after the induction of *cabergoline*, both of the *dopamine agonists* and *Transsphenoidal surgery* have a higher similarity score than *Transethmoidal Hypophysectomy*. Also *Cabergoline* is getting more popular after 2006 and is then significantly higher than other treatments.

4.1.3 White Blood Cell Cancer

A subtype of cancer of white blood cells known as chronic myeloid leukemia (CML) or *Chronic Myelosis* (C0023473) is a medical condition. In this condition there is an abnormal increase in the number of white blood cells (WBC) compared to red blood cells (RBC). WBC are responsible for protecting the body against infections, but when produced in large numbers, they start accumulating in blood and bone marrow. This prohibits the growth of RBC and causes weight loss, spleen enlargement and bone pain (Radich et al., 2018).

Before 2001, *Chronic Myelosis* was treated predominantly by chemotherapy using alkylating antineoplastic agents, such as *Mitobronitol* (C0026236) and *Myelobromol* (C0700014). The introduction of targeted therapy led to the improved survival rate of patients compared to the earlier generation of medication. The new targeted therapy method includes a class of drugs called *Tyrosine Kinase Inhibitors* (TKI) (C1268567), whereas *Imatinib* (C0935989) is one of the most important representatives of this class. *Tyrosine Kinase Inhibitors* were first synthesized in 1998 (Yaish et al., 1988), and *Imatinib* was first approved in 2001 to treat this type of blood cancer.

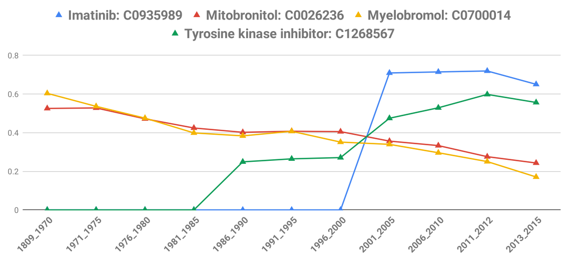


Figure 6: Similarity Score of White blood cell cancer with different treatment methods from 1809 until 2015. Change in the medication occurs in 2001. **Old Methods** were *Mitobronitol*, *Myelobromol*and **New Methods** are *Tyrosine Kinase Inhibitor* and *Imatinib*.

Figure 6 depicts different treatments used for white blood cell cancer. The similarity score for both *Mitobronitol* and *Myelobromol* is high in '70s. However, after '70s their score starts decreasing but are still significantly higher than *Tyrosine Kinase Inhibitor* from 1990's to 2000. From 2001 there is a significantly higher similarity for both *Tyrosine Kinase Inhibitor* and *Imatinib* as compared to both *Mitobronitol* and *Myelobromol*.

4.1.4 Hepatitis-C

Hepatitis-C (C0220847) is an infectious blood-borne disease which is caused by the hepatitis C virus (HCV). Hepatitis-C mainly affects the liver which can cause liver diseases and eventually lead to liver failure. HCV spreads mostly through infected blood transfusions or poorly sterilized injection needles, also during intravenous injection of drugs. (Maheshwari and Thuluvath, 2010).

Presently there is no vaccine to prevent HCV virus, however chronic infections are treated by antiviral medications (Webster and Klenerman, 2015). Until 2011, *Polyethylene Interferon Alpha-2a* (C0391001), *Polyethylene Interferon Alpha-2b* (C0796545) in combination with *Ribavirin*[5] (C0035525) were used to treat hepatitis-C and had a cure rate of less than 50%. From 2011, the second generation of antiviral medication known as Direct Antiviral Agents (DAA) was approved by the FDA. DAA directly interfere with the machinery of Hepatitis-C virus, thus inhibiting its growth and transmission. There are several classes of DAA that are used at different stages in the treatment of Hepatitis-C such as *Telaprevir*, *Boceprevir*, *Daclatasvir*. However, for current work we just show *Telaprevir* (C1876229). This DAA is used in combination with *Ribavirin* which have a cure rate of more than 90% (Rivett and Alexander, 2019).

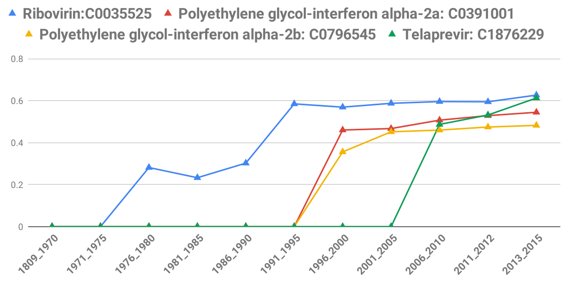


Figure 7: Similarity Score of Hepatitis-C with different treatment methods from 1809 to 2015, change in the medication occurs in 2011.**Old Methods** were *Polyethylene interferon alpha-(2a and 2b)* and **New Methods** is *Telaprevir*.

[5] First generation of antivirals.

Figure 7 shows a rise in the similarity of second generation of antivirals (*Telaprevir*) from 2011 as compared with first generations (*Polyethylene Interferon Alpha-(2a,2b)*) where there is a decrease in the similarity. From 2011 the similarity score of *Telaprevir* is significantly higher than the both *Polyethylene interferon alpha-(2a,2b)*, respectively. We can also notice, that the similarity score *Ribavirin* is high this is because it is still used in combination with the new generation of antiviral medications as well. Before 1976 the occurrence of any antivirals medication concepts that appears close to Hepatitis-C is not high enough as such concepts are not present.

4.2 Concept Embeddings vs. Co-occurrence

As seen in the examples above, temporal concept embeddings can be used to identify diachronic changes. In comparison to that, those changes can be also identified using a simple co-occurence metrics, such as PPMI. Figure 8 shows an example for White Blood Cell Cancer. However, in comparison to the example in Section 4.1.3, changes can be much stronger and values can quickly decrease to zero, if the co-occurrence of two concepts suddenly decreases.

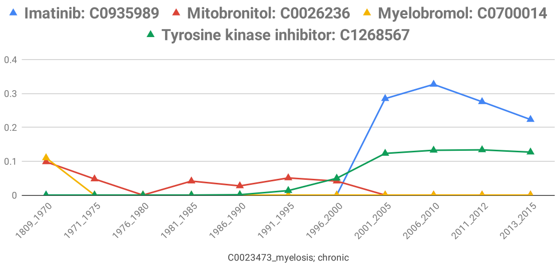


Figure 8: Similarity Score using PPMI matrix for WBC with different treatment methods from 1809 to 2015.

The score of *Mitobronitol* for instance suddenly drops to 0 in 1976-1980 and then increases in 1981. Conversely, the concept embeddings show a slow decrease in similarity for same pair at 1976-1980. This is can have several reasons: Firstly, even if concepts do not occur together within the same context window, they might occur within the same context which is considered by concept embedding. Moreover, the initialization of embeddings for 1976-1980 build on top of the previous period (1970-1975). The incremental learning mechanism helps concept embeddings to overcome the drawback of sudden drop in the similarity of a concept-pair if they do not co-occur in a specific period.

4.3 Discussion

The previous examples showed that we can use diachronic semantic changes to identify medical knowledge change. To measure the change in treatment of some disease from an old medication to a new medication was not as simple as our initial assumption was. Originally, examples were provided by a medical student on a rather high level. Given these examples the corresponding concepts and concept IDs had to be identified within UMLS. In various cases those concepts were ambiguous and the most appropriate concept had to be selected, e.g. Hepatitis-C in UMLS is represented as Hepatitis C virus (C0220847) as well as Hepatitis C (C0019196).

It also happened that a concept mention did not show the effect we were interested in (no occurrence, low similarity scores, no increase/decrease). This caused a more detailed manual analysis to find out why. In some cases, if a concept did not show the effect we were searching for, it turned out that a more specific concept instead showed the expected effect. For instance we found that particular derivatives of *dopamine agonists* such *Cabergoline* and *Bromocriptine* were more talked about in the context of Microprolactinoma than *dopamine agonists*. This is an interesting aspect of how information are connected and which are actually mentioned in the scientific text. Unfortunately this is difficult to solve given our high level examples and a method solely based on general literature.

However, even though the examples above were manually selected with a lot of domain knowledge, we can clearly show that knowledge changes are present in our temporal concept embeddings. In order to address possible concerns, the next section explores knowledge changes of known UMLS pairs.

5 Exploring Existing Medical Knowledge

In the previous section, we showed that changes in biomedical knowledge and particularly changes of treatments could be reflected within temporal concept embeddings. However, those examples were manually selected by a medical expert. In this section instead we apply the technique to explore known drug-disease pair relations of the UMLS

		Concept Embeddings				**Co-occurrence**		
Period	**#**	**POS (MAX)**	**#**	**NEG (MAX)**	**#**	**POS (MAX)**	**#**	**NEG (MAX)**
1809-1970	7	0.330 (0.754)	56	0.208 (0.588)	225	0.026 (0.286)	573	0.003 (0.171)
1971-1975	2	0.318 (0.721)	52	0.197 (0.596)	644	0.022 (0.354)	984	0.003 (0.086)
1976-1980	11	0.307 (0.742)	106	0.168 (0.564)	360	0.026 (0.325)	680	0.003 (0.134)
1981-1985	10	0.310 (0.711)	137	0.157 (0.530)	355	0.029 (0.330)	663	0.003 (0.150)
1986-1990	12	0.304 (0.681)	135	0.155 (0.553)	388	0.028 (0.310)	729	0.002 (0.139)
1991-1995	16	0.301 (0.672)	150	0.149 (0.505)	527	0.026 (0.266)	761	0.002 (0.073)
1996-2000	12	0.297 (0.680)	157	0.149 (0.510)	566	0.025 (0.337)	780	0.002 (0.149)
2001-2005	7	0.287 (0.689)	147	0.146 (0.499)	536	0.024 (0.309)	767	0.002 (0.121)
2006-2010	10	0.271 (0.695)	177	0.144 (0.476)	655	0.021 (0.300)	832	0.002 (0.077)
2011-2012	13	0.272 (0.730)	146	0.153 (0.467)	957	0.017 (0.355)	1178	0.002 (0.088)
2013-2015	15	0.265 (0.696)	136	0.152 (0.425)	1158	0.015 (0.246)	1264	0.002 (0.077)

Table 2: Exploration of known (positive) and unknown (negative) drug-disease concept pairs of UMLS across different time periods. The table shows the mean and its maximum scores below POS and NEG in terms of cosine similarity and PPMI. In addition to that, that table shows the number of concept pairs (#) which do not occur together within the set of 3,000 drug-disease pairs.

Metathesaurus. First we explore known concept pairs with cosine similarity for concept embeddings in comparison to PPMI. After that we examine selected relations of UMLS and track their similarity across different periods.

5.1 Exploring known Drug-Disease Pairs: Concept Embeddings vs. Co-occurrence

In the following we examine concept embeddings using cosine similarity in comparison to the co-occurrence metric PPMI on known drug-disease relations of UMLS. To do so, we use *may-treat* and *may-prevent* relations of UMLS and selected randomly for each time period a set of 3,000 concept pairs. We made sure, that both concepts occurred within that time slice. Then we randomly generated a set of negative concept pairs (unknown according to UMLS) with the same size. Next we use both sets (positive and negative) to calculate cosine similarity using concept embedding and PPMI matrix .

The results are presented in Table 2 and show, that the average score is higher for the known relations pairs (positive) in comparison to the randomly generated negative pairs. This is valid for cosine similarity and PPMI. Moreover we can see, that the average cosine score for concept embedding is above the PPMI, as well as for the corresponding MAX scores. However, both scores can not be directly compared.

Interestingly, the table shows a varying number of concept pairs which are not covered by a metric (lower than .05 for concept embedding and zero for PPMI). Particularly the co-occurrence metric PPMI has fewer information about various concept pairs in comparison to concept embedding. For instance, in period 2013-2015 while the cosine similarity for concept embedding score for only 15 positive concept pairs is below .05, 1158 concept pairs are not considered by co-occurrence, as concepts do not occur together frequent enough. Note, the low PPMI scores might be related to the sparseness of the PPMI matrix.

Overall, the results show, that the incremental temporal concept embeddings have got an advantage over the co-occurrence metric PPMI. As the concept embedding uses knowledge from previous time slices and considers contextual information it is able to better cope with the situation if concept pairs do not frequently together.

5.2 Exploring Drug-Disease Pairs across different Time Periods

In the following we use temporal concept embeddings to explore changes in biomedical knowledge. We apply the technique to explore known drug-disease pair relations *may_treat* and *may_prevent* of the UMLS Metathesaurus. An increase over time might indicate[6] a higher use of drug against the corresponding disease in present time as compared to previous periods; whereas a decrease can indicate new treatment therapy for the disease from disease-drug pair. This might be interesting as often various treatments exist for a disease. In this way, it might be possible to identify a more popular treatment (according to similarity score) which is at the same time also encoded within the embeddings.

[6] Of course it could also mean something different.

Drug	Disease	1809-1970	1971-1975	1991-1995	1996-2000	2011-2012	2013-2015
Oxymetholone Epoetin Alfa Recombinant	Anemias	0.36 0.00	0.43 0.00	0.23 0.43	0.18 0.37	0.23 0.35	0.23 0.42
Sodium Cromoglycate Aalmeterol	Bronchitic Asthma	0.58 0.00	0.60 0.00	0.51 0.56	0.45 0.55	0.45 0.50	0.29 0.56
Tolazamide Sitagliptin	Type 2 Diabetes Mellitus	0.63 0.00	0.46 0.00	0.48 0.00	0.45 0.00	0.28 0.54	0.27 0.62
Pramipexole Amantadine Hydrochloride	Syndrome Parkinson's	0.00 0.41	0.00 0.51	0.39 0.41	0.48 0.36	0.46 0.17	0.46 0.20
Risperidone Acetophenazine Maleate	Type Schizophrenia	0.00 0.46	0.00 0.41	0.53 0.35	0.60 0.32	0.58 0.20	0.59 0.20
Tamoxifen Testolactone	Tumor of Breast	0.00 0.43	0.28 0.38	0.51 0.25	0.48 0.22	0.49 0.12	0.49 0.10

Table 3: Decrease (upper part) and increase (lower part) in similarity for *may-treat* and *may-prevent* drug-disease pairs across different time periods

Table 3 presents results for particular diseases in terms of increasing and decreasing similarity scores for known *may-treat* and *may-prevent* drug-disease pairs. The similarity scores shown here are for the first two periods (1809-1970, 1971-1975), two periods from the middle (1991-1995, 1996-2000) and the last two ones (2011-2012, 2013-2015). Each row contains a two different known drugs related to a disease. The upper part presents a scenario with a decreasing similarity score (relative to the disease) and the lower part an increasing score. For example, the table shows that the similarity between *Tolazamide* and *Type 2 Diabetes Mellitus* is .63 in 1809-1970. With each succeeding period the value decreases and eventually reaches .27 in 2013-2015. On the other hand, the similarity between the *Sitagliptin* with *Type 2 Diabetes Mellitus* is 0 until 1996-2000 due to its absence in this period. However, from 2011 we see a sudden and strong increase.

The table shows that we can detect changes of known relational facts. The results are also in line with our original hypothesis that scientific journals reflect the change in medical knowledge since each journal provide current medical facts. As scientific research around these fact evolves, we witness a change in medical knowledge which is present in the scientific journals.

6 Conclusion

In the present work, we have successfully shown that it is possible to explore the diachronic semantic change on a biomedical concept level. The automatic exploration of knowledge changes might be particularly useful to extend structured knowledge, such as UMLS potentially. For instance, UMLS often includes an extensive range of different treatments or preventions for a disease. However, all relations have the same importance and the same weighting. Thus it is not necessarily obvious which one is the treatment of choice (also depending on time, but also co-morbidities or other symptoms). Our proposed method could be a first (and simplistic) step to highlight particular concept pairs. For instance, temporal concept embeddings could be used to support (distantly supervised) relation extraction (Roller and Stevenson, 2014) or to spot particular trends automatically (Chen et al., 2007).

However, our current approach has got some limitations as it is unable to detect the negative polarity between the pairs. In terms of this we assume that a higher similarity is correlated with a stronger use, which is not necessarily correct. Future work could take this into account.

Finally, as mentioned in Section 4.3, it would be interesting to address the problem that sometimes only particular child concepts show an effect we are interested in. It might be possible to overcome this by including graph embeddings in addition to the text based temporal ones.

Acknowledgments

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 780495 (BigMedilytics). In addition to that we would like to thank our colleagues for their feedback and suggestions.

References

Alan R Aronson. 2001. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. In *Proceedings of the AMIA Symposium*, page 17. American Medical Informatics Association.

Andrew L. Beam, Benjamin Kompa, Inbar Fried, Nathan P. Palmer, Xu Shi, Tianxi Cai, and Isaac S. Kohane. 2018. Clinical Concept Embeddings Learned from Massive Sources of Medical Data. *CoRR*, abs/1804.01486.

Andreas Blank. 1999. Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. *Historical semantics and cognition*, 13:6.

Olivier Bodenreider. 2004. The Unified Medical Language System (UMLS): Integrating Biomedical Terminology. *Nucleic acids research*, 32(Database issue):D267–70.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *arXiv preprint arXiv:1607.04606*.

Joan Bybee et al. 2007. *Frequency of Use and the Organization of Language*. Oxford University Press on Demand.

Felipe F Casanueva, Mark E Molitch, Janet A Schlechte, Roger Abs, Vivien Bonert, Marcello D Bronstein, Thierry Brue, Paolo Cappabianca, Annamaria Colao, Rudolf Fahlbusch, et al. 2006. Guidelines of the Pituitary Society for the diagnosis and management of prolactinomas. *Clinical endocrinology*, 65(2):265–273.

Elizabeth S Chen, Peter D Stetson, Yves A Lussier, Marianthi Markatou, George Hripcsak, and Carol Friedman. 2007. Detection of practice pattern trends through natural language processing of clinical narratives and biomedical literature. In *AMIA Annual Symposium Proceedings*, volume 2007, page 120. American Medical Informatics Association.

Youngduck Choi, Chill Yi-I Chiu, and David Sontag. 2016. Learning Low-Dimensional Representations of Medical Concepts. *AMIA Summits on Translational Science Proceedings*, 2016:41.

Kenneth Ward Church and Patrick Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational linguistics*, 16(1):22–29.

Greta G. Cummings, Susan Armijo Olivo, Patricia D. Biondo, Carla R. Stiles, Ozden Yurtseven, Robin L. Fainsinger, and Neil A. Hagen. 2011. Effectiveness of Knowledge Translation Interventions to Improve Cancer Pain Management. *J. Pain Symptom Manage.*, 41(5):915–939.

Lance De Vine, Guido Zuccon, Bevan Koopman, Laurianne Sitbon, and Peter Bruza. 2014. Medical Semantic Similarity with a Neural Language Model. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1819–1822. ACM.

John R Firth. 1957. A Synopsis of Linguistic Theory, 1930-1955. *Studies in linguistic analysis*.

Andrea Glezer and Marcello D. Bronstein. 2015. Prolactinomas. *Endocrinol. Metab. Clin. North Am.*, 44(1):71–78.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proc. Assoc. Comput. Ling. (ACL)*.

Jiangen He and Chaomei Chen. 2018. Predictive Effects of Novelty Measured by Temporal Embeddings on the Growth of Scientific Literature. *Frontiers in Research Metrics and Analytics*, 3:9.

Adam Jatowt and Kevin Duh. 2014. A Framework for Analyzing Semantic Change of Words across Time. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 229–238. IEEE Press.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal Analysis of Language through Neural Language Models. In *Proceedings of the Workshop on Language Technologies and Computational Social Science@ACL 2014, Baltimore, MD, USA, June 26, 2014*, pages 61–65. Association for Computational Linguistics.

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2014. Statistically Significant Detection of Linguistic Change. *CoRR*, abs/1411.3315.

Omer Levy and Yoav Goldberg. 2014. Neural Word Embedding as Implicit Matrix Factorization. In *Advances in Neural Information Processing Systems*, pages 2177–2185.

Omer Levy, Yoav Goldberg, and Ido Dagan. 2015. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225.

Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang, and Martin A Nowak. 2007. Quantifying the Evolutionary Dynamics of Language. *Nature*, 449(7163):713.

Yue Liu, Tao Ge, Kusum S Mathews, Heng Ji, and Deborah L McGuinness. 2018. Exploiting Task-Oriented Resources to Learn Word Embeddings for Clinical Abbreviation Expansion. *arXiv preprint arXiv:1804.04225*.

Anurag Maheshwari and Paul J. Thuluvath. 2010. Management of acute hepatitis C. *Clin. Liver Dis.*, 14(1):169–176.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013b. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

Mark Pagel, Quentin D Atkinson, and Andrew Meade. 2007. Frequency of Word-Use Predicts Rates of Lexical Evolution throughout Indo-European History. *Nature*, 449(7163):717.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A Meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543. ACL.

Lawrence Phillips, Kyle Shaffer, Dustin Arendt, Nathan Hodas, and Svitlana Volkova. 2017. Intrinsic and extrinsic evaluation of spatiotemporal text representations in Twitter streams. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 201–210.

Jerald P Radich, Michael Deininger, Camille N Abboud, Jessica K Altman, Ellin Berman, Ravi Bhatia, Bhavana Bhatnagar, Peter Curtin, Daniel J DeAngelo, Jason Gotlib, et al. 2018. Chronic myeloid leukemia, version 1.2019, nccn clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*, 16(9):1108–1135.

S H Richards, J P Thomas, and D Kilby. 1974. Transethmoidal hypophysectomy for pituitary tumours. *Proceedings of the Royal Society of Medicine*, 67(9):889–892.

Lucy Rivett and Graeme Alexander. 2019. Is the conquest of hepatitis c imminent? *Expert reviews in molecular medicine*, 21.

Roland Roller and Mark Stevenson. 2014. Self-supervised Relation Extraction Using UMLS. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, pages 116–127, Cham. Springer International Publishing.

Eyal Sagi, Stefan Kaufmann, and Brady Clark. 2009. Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 104–111. Association for Computational Linguistics.

Jenna R Stoehr, Jennifer N Choi, Maria Colavincenzo, and Stefan Vanderweil. 2019. Off-label use of topical minoxidil in alopecia: A review. *American journal of clinical dermatology*, pages 1–14.

Don R Swanson. 1986. Fish oil, raynaud's syndrome, and undiscovered public knowledge. *Perspectives in biology and medicine*, 30(1):7–18.

Amit Tirosh and Ilan Shimon. 2016. Current approach to treatments for prolactinomas. *Minerva Endocrinol.*, 41(3):316–323.

Peter D Turney and Patrick Pantel. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of artificial intelligence research*, 37:141–188.

Daniel P Webster and Paul Klenerman. 2015. Hepatitis c. *Hepatitis C. Lancet*, 385(9973):1124–1135.

Derry Tanti Wijaya and Reyyan Yeniterzi. 2011. Understanding Semantic Change of Words over Centuries. In *Proceedings of the 2011 International Workshop on DETecting and Exploiting Cultural diversiTy on the Social Web*, pages 35–40. ACM.

P. Yaish, A. Gazit, C. Gilon, and A. Levitzki. 1988. Blocking of EGF-dependent cell proliferation by EGF receptor kinase inhibitors. *Science*, 242(4880):933–935.

Erjia Yan and Yongjun Zhu. 2018. Tracking Word Semantic Change in Biomedical Literature. *International journal of medical informatics*, 109:76–86.

Li Zhou and George Hripcsak. 2007. Temporal Reasoning with Medical Data—a Review with Emphasis on Medical Natural Language Processing. *Journal of biomedical informatics*, 40(2):183–202.

Fei Zhu, Preecha Patumcharoenpol, Cheng Zhang, Yang Yang, Jonathan Chan, Asawin Meechai, Wanwipa Vongsangnak, and Bairong Shen. 2013. Biomedical Text Mining and Its Applications in Cancer Research. *Journal of biomedical informatics*, 46(2):200–211.