



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/196485/>

Version: Published Version

Proceedings Paper:

Newman-Griffis, D., Sivaraman, V., Perer, A. et al. (2021) TextEssence: a tool for interactive analysis of semantic shifts between corpora. In: Sil, A. and Lin, X.V., (eds.) Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations. The 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations, 06-11 Jun 2021, Online. Association for Computational Linguistics, pp. 106-115. ISBN: 9781954085480.

<https://doi.org/10.18653/v1/2021.naacl-demos.13>

© 2021 Association for Computational Linguistics. Licensed on a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

TextEssence: A Tool for Interactive Analysis of Semantic Shifts Between Corpora

Denis Newman-Griffis¹ Venkatesh Sivaraman² Adam Perer²
Eric Fosler-Lussier³ Harry Hochheiser^{1,4}

¹Department of Biomedical Informatics, University of Pittsburgh

²Human-Computer Interaction Institute, Carnegie Mellon University

³Department of Computer Science and Engineering, The Ohio State University

⁴Intelligent Systems Program, University of Pittsburgh

dnewmangriffis@pitt.edu

vsivaram@andrew.cmu.edu

Abstract

Embeddings of words and concepts capture syntactic and semantic regularities of language; however, they have seen limited use as tools to study characteristics of different corpora and how they relate to one another. We introduce TextEssence, an interactive system designed to enable comparative analysis of corpora using embeddings. TextEssence includes visual, neighbor-based, and similarity-based modes of embedding analysis in a lightweight, web-based interface. We further propose a new measure of embedding confidence based on nearest neighborhood overlap, to assist in identifying high-quality embeddings for corpus analysis. A case study on COVID-19 scientific literature illustrates the utility of the system. TextEssence can be found at <https://textessence.github.io>.

1 Introduction

Distributional representations of language, such as word and concept embeddings, provide powerful input features for NLP models in part because of their correlation with syntactic and semantic regularities in language use (Boleda, 2020). However, the use of embeddings as a lens to investigate those regularities, and what they reveal about different text corpora, has been fairly limited. Prior work using embeddings to study language shifts, such as the use of diachronic embeddings to measure semantic change in specific words over time (Hamilton et al., 2016; Schlechtweg et al., 2020), has focused primarily on quantitative measurement of change, rather than an interactive exploration of its qualitative aspects. On the other hand, prior work on interactive analysis of text collections has focused on analyzing individual corpora, rather than facilitating inter-corpus analysis (Liu et al., 2012; Weiss, 2014; Liu et al., 2019).

We introduce TextEssence, a novel tool that combines the strengths of these prior lines of research

by enabling interactive comparative analysis of different text corpora. TextEssence provides a multi-view web interface for users to explore the properties of and differences between multiple text corpora, all leveraging the statistical correlations captured by distributional embeddings. TextEssence can be used both for categorical analysis (i.e., comparing text of different genres or provenance) and diachronic analysis (i.e., investigating the change in a particular type of text over time).

Our paper makes the following contributions:

- We present TextEssence, a lightweight tool implemented in Python and the Svelte JavaScript framework, for interactive qualitative analysis of word and concept embeddings.
- We introduce a novel measure of *embedding confidence* to mitigate embedding instability and quantify the reliability of individual embedding results.
- We report on a case study using TextEssence to investigate diachronic shifts in the scientific literature related to COVID-19, and demonstrate that TextEssence captures meaningful month-to-month shifts in scientific discourse.

The remainder of the paper is organized as follows. §2 lays out the conceptual background behind TextEssence and its utility as a corpus analysis tool. In §3 and §4, we describe the nearest-neighbor analysis and user interface built into TextEssence. §5 describes our case study on scientific literature related to COVID-19, and §6 highlights key directions for future research.

2 Background

Computational analysis of text corpora can act as a lens into the social and cultural context in which those corpora were produced (Nguyen et al., 2020). Diachronic word embeddings have been shown to reflect important context behind the corpora they

are trained on, such as cultural shifts (Kulkarni et al., 2015; Hamilton et al., 2016; Garg et al., 2018), world events (Kutuzov et al., 2018), and changes in scientific and professional practice (Vylomova et al., 2019). However, these analyses have proceeded independently of work on interactive tools for exploring embeddings, which are typically limited to visual projections (Zhordaniya et al.; Warmerdam et al., 2020). TextEssence combines these directions into a single general-purpose tool for interactively studying differences between any set of corpora, whether categorical or diachronic.

2.1 From words to domain concepts

When corpora of interest are drawn from specialized domains, such as medicine, it is often necessary to shift analysis from individual words to *domain concepts*, which serve to reify the shared knowledge that underpins discourse within these communities. Reified domain concepts may be referred to by multi-word surface forms (e.g., “Lou Gehrig’s disease”) and multiple distinct surface forms (e.g., “Lou Gehrig’s disease” and “amyotrophic lateral sclerosis”), making them more semantically powerful but also posing distinct challenges from traditional word-level representations.

A variety of embedding algorithms have been developed for learning representations of domain concepts and real-world entities from text, including weakly-supervised methods requiring only a terminology (Newman-Griffis et al., 2018); methods using pre-trained NER models for noisy annotation (De Vine et al., 2014; Chen et al., 2020); and methods leveraging explicit annotations of concept mentions (as in Wikipedia) (Yamada et al., 2020).¹ These algorithms capture valuable patterns about concept types and relationships that can inform corpus analysis (Runge and Hovy, 2020).

TextEssence only requires pre-trained embeddings as input, so it can accommodate any embedding algorithm suiting the needs and characteristics of specific corpora (e.g. availability of annotations or knowledge graph resources). Furthermore, while the remainder of this paper primarily refers to concepts, TextEssence can easily be used for word-level embeddings in addition to concepts.

2.2 Why static embeddings?

Contextualized, language model-based embeddings can provide more discriminative features for

¹The significant literature on learning embeddings from knowledge graph structure is omitted here for brevity.

NLP than static (i.e., non-contextualized) embeddings. However, static embeddings have several advantages for this comparative use case. First, they are less resource-intensive than contextualized models, and can be efficiently trained several times without pre-training to focus entirely on the characteristics of a given corpus. Second, the scope of what static embedding methods are able to capture from a given corpus has been well-established in the literature, but is an area of current investigation for contextualized models (Jawahar et al., 2019; Zhao and Bethard, 2020). Finally, the nature of contextualized representations makes them best suited for context-sensitive tasks, while static embeddings capture aggregate patterns that lend themselves to corpus-level analysis. Nevertheless, as work on qualitative and visual analysis of contextualized models grows (Hoover et al., 2020), new opportunities for comparative analysis of local contexts will provide fascinating future research.

3 Identifying Stable Embeddings for Analysis

While embeddings are a well-established means of capturing syntax and semantics from natural language text (Boleda, 2020), the problem of comparing multiple sets of embeddings remains an active area of research. The typical approach is to consider the nearest neighbors of specific points, consistent with the “similar items have similar representations” intuition of embeddings. This method also avoids the conceptual difficulties and low replicability of comparing embedding spaces numerically (e.g. by cosine distances) (Gonen et al., 2020). However, even nearest neighborhoods are often unstable, and vary dramatically across runs of the same embedding algorithm on the same corpus (Wendlandt et al., 2018; Antoniak and Mimno, 2018). In a setting such as our case study, the relatively small sub-corpora we use (typically less than 100 million tokens each) exacerbate this instability. Therefore, to quantify variation across embedding replicates and identify informative concepts, we introduce a measure of *embedding confidence*.²

We define embedding confidence as the mean overlap between the top k nearest neighbors of a

²An *embedding replicate* here refers to the embedding matrix output by running a specific embedding training algorithm on a specific corpus. Ten runs of word2vec on a given Wikipedia dump produce ten replicates; using different Wikipedia dumps would produce one replicate each of ten different sets of embeddings.

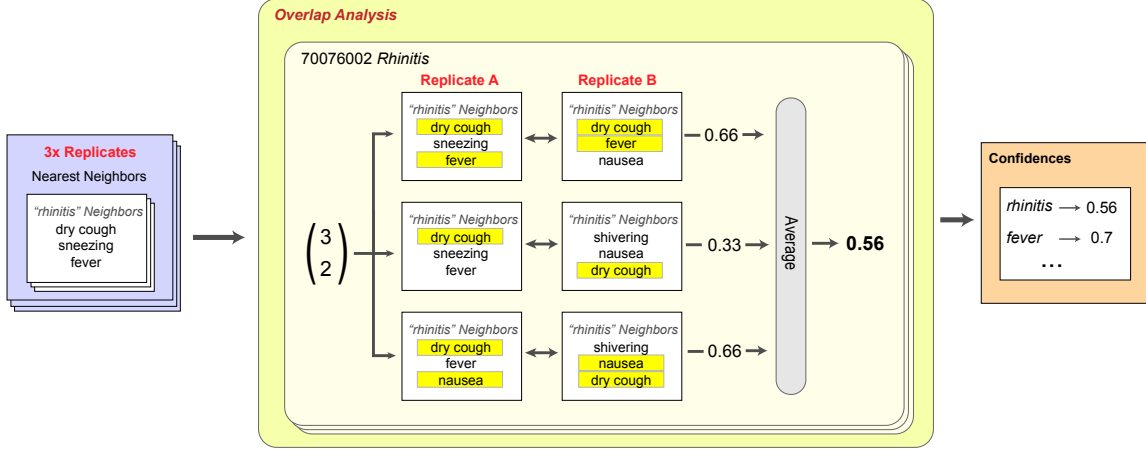


Figure 1: Illustration of embedding confidence calculation, using 3 embedding replicates. Our experiments on CORP-19 used this same process with 10 replicates.

given item between multiple embedding replicates. Formally, let $E^1 \dots E^m$ be m embedding replicates trained on a given corpus, and let $kNN^i(c)$ be the set of k nearest neighbors by cosine similarity of concept c in replicate E^i . Then, the embedding confidence $EC@k$ is defined as:

$$EC@k(c, E^1 \dots E^m) = \frac{1}{m(m-1)} \sum_i^m \sum_{j \neq i}^m \frac{|kNN^i(c) \cap kNN^j(c)|}{k}$$

An example of this calculation is illustrated in Figure 1.

We can then define the set of *high-confidence* concepts for the given corpus as the set of all concepts with an embedding confidence above a given threshold. A higher threshold will restrict to highly-stable concepts only, but exclude the majority of embeddings. We recommend an initial threshold of 0.5, which can be configured based on observed quality of the filtered embeddings.

3.1 Computing aggregate nearest neighbors

After filtering for high-confidence concepts, we summarize nearest neighbors across replicates by computing *aggregate nearest neighbors*. The aggregate neighbor set of a concept c is the set of high-confidence concepts with highest average cosine similarity to c over the embedding replicates. More precisely, let $D^i(c)$ be the vector of pairwise similarities between concept c and all concepts in the embedding vocabulary,³ as calculated in replicate E^i . Then, we calculate the aggregate pairwise

³All embedding replicates trained on a given corpus will share the same vocabulary.

distance vector for c as:

$$D_{Agg}(c) = \frac{1}{m} \sum_i^m D^i(c)$$

The k aggregate nearest neighbors of c , $kNN_{Agg}(c)$, are then the k concepts with lowest values in $D_{Agg}(c)$. This helps to provide a more reliable picture of the concept’s nearest neighbors, while excluding concepts whose neighbor sets are uncertain.

4 The TextEssence Interface

The workflow for using TextEssence to compare different corpora is illustrated in Figure 2. Given the set of corpora to compare, the user (1) trains embedding replicates on each corpus; (2) identifies the high-confidence set of embeddings for each corpus; and (3) provides these as input to TextEssence. TextEssence then offers three modalities for interactively exploring their learned representations: (1) Browse, an interactive visualization of the embedding space; (2) Inspect, a detailed comparison of a given concept’s neighbor sets across corpora; and (3) Compare, a tool for analyzing the pairwise relationships between two or more concepts.

4.1 Browse: visualizing embedding changes

The first interface presented to the user is an overview visualization of one of the embedding spaces, projected into 2-D using t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten and Hinton, 2008). High-confidence concepts are depicted as points in a scatter plot and may be color-coded based on pre-existing groupings; for example, in our case study (§5), we color-coded

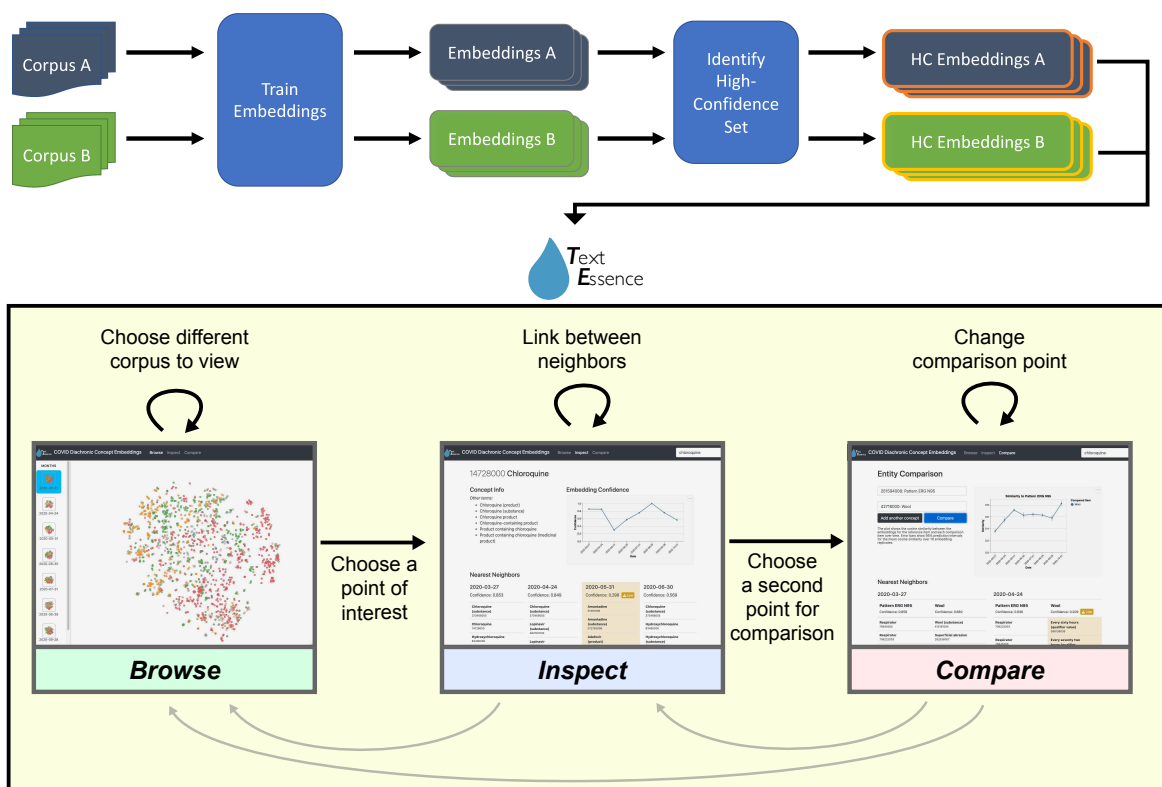


Figure 2: Workflow for comparing corpus embeddings with TextEssence. The system enables three different kinds of interactions: (1) Browse the embedding space for each corpus; (2) Inspect a single concept in each corpus; and (3) Compare two or more concepts across corpora. Each view transitions to the others using the current concept.

medical concepts based on their semantic groups in the UMLS (McCray et al., 2001), such as “Chemicals & Drugs” and “Disorders.” The user can select a point to highlight its aggregated nearest neighbors in the high-dimensional space, an interaction similar to TensorFlow’s Embedding Projector (Smilkov et al., 2016) that helps distinguish true neighbors from artifacts of the dimensionality reduction process.

The Browse interface also expands upon existing dimensionality reduction tools by enabling visual comparison of multiple corpora (e.g., embeddings from individual months). This is challenging because the embedding spaces are trained separately, and can therefore differ greatly in both high-dimensional and reduced representations. While previous work on comparing projected data has focused on algorithmically aligning projections (Liu et al., 2020; Chen et al., 2018) and adding new comparison-focused visualizations (Cutura et al., 2020), we chose to align the projections using a simple Procrustes transformation and enable the user to compare them using animation.

When the user hovers on a corpus thumbnail, *pre-*

view lines are shown between the positions of each concept in the current and destination corpora. The direction of each line is disambiguated by increasing its width as the line approaches its destination. In addition, the width and opacity of each point’s preview line are proportional to the fraction of the point’s aggregate nearest neighbors that differ between the source and destination corpora. This serves to draw attention to the concepts that shift the most. Upon clicking the corpus thumbnail, the points smoothly follow their trajectory lines to form the destination plot. In addition, when a concept is selected, the user can opt to *center* the visualization on that point and then transition between corpora, revealing how neighboring concepts move relative to the selected one.

4.2 Inspect: tracking individual concept change

Once a particular concept of interest has been identified, the Inspect view presents an interactive table depicting how that concept’s aggregated nearest neighbors have changed over time. This view also displays other contextualizing informa-

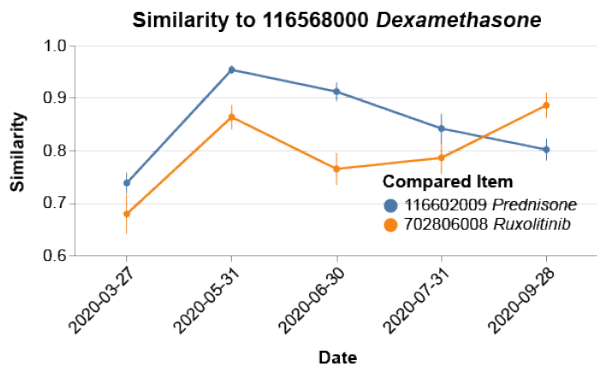


Figure 3: Similarity over time of two drugs to 116568000 *Dexamethasone* in our case study. April, August, and October are omitted as *Dexamethasone* was not high confidence for these months. Similarity values are mean over embedding replicates within each month; error bars indicate standard deviations.

tion about the concept, including its definitions (derived from the UMLS (Bodenreider, 2004) for our case study⁴), the terms used to refer to the concept (limited to SNOMED CT for our case study), and a visualization of the concept’s embedding confidence over the sub-corpora analyzed. For information completeness, we display nearest neighbors from every corpus analyzed, even in corpora where the concept was not designated high-confidence (note that a concept must be high-confidence in at least one corpus to be selectable in the interface). In these cases, a warning is shown that the concept itself is not high-confidence in that corpus; the neighbors themselves are still exclusively drawn from the high-confidence set.

4.3 Compare: tracking pair similarity

The Compare view facilitates analysis of the changing relationship between two or more concepts across corpora (e.g. from month to month). This view displays paired nearest neighbor tables, one per corpus, showing the aggregate nearest neighbors of each of the concepts being compared. An adjacent line graph depicts the similarity between the concepts in each corpus, with one concept specified as the reference item and the others serving as comparison items (similar to Figure 3). Similarity between two concepts for a specific corpus is calculated by averaging the cosine similarity between the corresponding embeddings in each replicate.

⁴We included definitions from all English-language sources in the UMLS, as SNOMED CT includes definitions only for a small subset of concepts.

Month	Docs	Words	Entities	Hi-Conf.
March	41,750	158M	38,451	15,100
April	10,738	41M	25,142	1,851
May	73,444	125M	40,297	5,051
June	24,813	34M	19,749	2,729
July	24,786	35M	19,334	2,800
August	28,642	31M	19,134	2,407
September	33,732	38M	20,947	4,381
October	38,866	44M	21,470	1,990

Table 1: 2020 monthly snapshots of CORD-19 dataset (documents added each month only; not cumulative). Entities denotes the number of SNOMED CT codes for which embeddings were learned; Hi-Conf. is the subset of these that had confidence above the 0.5 threshold. The high document count in March 2020 included all COVID-19 literature published prior to March 13, 2020 (beginning of CORD-19 dataset); the spike in May 2020 was due to adding arXiv and Medline as data sources for CORD-19.

5 Case Study: Diachronic Change in CORD-19

The scale of global COVID-19-related research has led to an unprecedented rate of new scientific findings, including developing understanding of the complex relationships between drugs, symptoms, comorbidities, and health outcomes for COVID-19 patients. We used TextEssence to study how the contexts of medical concepts in COVID-19-related scientific literature have changed over time. Table 1 shows the number of new articles indexed in the COVID-19 Open Research Dataset (CORD-19; Wang et al. (2020a)) from its beginning in March 2020 to the end of October 2020; while additions of new sources over time led to occasional jumps in corpus volumes, all are sufficiently large for embedding training. We created disjoint sub-corpora containing the new articles indexed in CORD-19 each month for our case study.

CORD-19 monthly corpora were tokenized using ScispaCy (Neumann et al., 2019), and concept embeddings were trained using JET (Newman-Griffis et al., 2018), a weakly-supervised concept embedding method that does not require explicit corpus annotations. We used SNOMED Clinical Terms (SNOMED CT), a widely-used reference representing concepts used in clinical reporting, as our terminology for concept embedding training, using the March 2020 interim release of SNOMED CT International Edition, which included COVID-19 concepts. We trained JET embeddings using a vector dimensionality $d = 100$ and 10 iterations,

to reflect the relatively small size of each corpus. We used 10 replicates per monthly corpus, and a high-confidence threshold of 0.5 for EC@5.

5.1 Findings

TextEssence captures a number of shifts in COVID-19 that reflect how COVID-19 science has developed over the course of the pandemic. Table 2 highlights key findings from our preliminary investigation into concepts known *a priori* to be relevant. Please note that while full nearest neighbor tables are omitted due to space limitations, they can be accessed by downloading our code and following the included guide to inspect COVID-19 results.

44169009 Anosmia While associations of anosmia (loss of sense of smell) were observed early in the pandemic (e.g., Hornuss et al. (2020), posted in May 2020), it took time to begin to be utilized as a diagnostic variable (Talavera et al., 2020; Wells et al., 2020). *Anosmia*’s nearest neighbors reflect this, staying stably in the region of other otolaryngological concepts until October (when Talavera et al. (2020); Wells et al. (2020), *inter alia* were included in COVID-19), where we observe a marked shift in utilization to play a similar role to other common symptoms of COVID-19.

116568000 Dexamethasone The corticosteroid dexamethasone was recognized early as valuable for treating severe COVID-19 symptoms (Lester et al. (2020), indexed July 2020), and its role has remained stable since (Ahmed and Hassan (2020), indexed October 2020). This is reflected in the shift of its nearest neighbors from prior contexts of traumatic brain injury (Moll et al., 2020) to a stable neighborhood of other drugs used for COVID-19 symptoms. However, in September 2020, 702806008 *Ruxolitinib* emerges as *Dexamethasone*’s nearest neighbor. This reflects a spike in literature investigating the use of ruxolitinib for severe COVID-19 symptom management (Gozzetti et al., 2020; Spadea et al., 2020; Li and Liu, 2020). As the similarity graph in Figure 3 shows, the contextual similarity between dexamethasone and ruxolitinib steadily increases over time, reflecting the growing recognition of ruxolitinib’s new utility (Caocci and La Nasa (2020), indexed May 2020).

83490000 Hydroxychloroquine Hydroxychloroquine, an anti-malarial drug, was misleadingly promoted as a potential treatment for COVID-19 by President Trump in March, May, and July 2020, leading to widespread misuse of the drug (Englund

Concept	Month(s)	Representative neighbors
44169009 Anosmia	Mar-Sep	2553606007 Gustatory 51388003 Pharyngeal pain 60707004 Taste
	Oct	15387003 Vomiting 73879007 Nausea 49727002 Cough
116568000 Dexamethasone	Mar	19130008 Injury 417746004 Traumatic injury
	May-Jul	116602009 Prednisone 108675009 Infliximab
	Sep	702806008 Ruxolitinib
83490000 Hydroxychloroquine	All	80229008 Antimalarial agent 96034006 Azithromycin
	Aug	198051006 Nosocomial infection 233765002 Respiratory failure without hypercapnia

Table 2: Representative nearest neighbors (manually selected from top 10) for three concepts in COVID-19, grouped by period of observation. Complete nearest neighbor tables are omitted for brevity, but may be viewed using our released code and data.

et al., 2020). As a result, a number of studies re-investigated the efficacy of hydroxychloroquine as a treatment for COVID-19 in hospitalized patients (Ip et al. (2020); Albani et al. (2020); Rahmani et al. (2020), all indexed August 2020). This shift is reflected in the neighbors of *Hydroxychloroquine*, adding investigative outcomes such as nosocomial (hospital-acquired) infections and respiratory failure to the expected anti-malarial neighbors.

6 Discussion

Our case study on scientific literature related to COVID-19 demonstrates that TextEssence can be used to study diachronic shifts in usage of domain concepts. We highlight three directions for future work using TextEssence: mining for new shifts and associations in changing literature (§6.1); applications between comparative analysis of corpora (§6.2); and further investigation of embedding confidence as a tool for analysis (§6.3).

6.1 Mining shifts in the literature

While our primary focus in developing TextEssence was on its use as a qualitative tool for targeted

inquiry, diachronic embeddings have significant potential for knowledge discovery through quantitative measurement of semantic differences. For example, new embeddings could be generated for subsequent months of COVID-19 (or other corpora), and analyzed to determine what concepts shifted the most—indicating current trends—or what concepts are just starting to shift—suggesting potential future developments.

However, quantitative, vector-based comparison of embedding spaces faces significant conceptual challenges, such as a lack of appropriate alignment objectives and empirical instability (Gonen et al., 2020). While nearest neighbor-based change measurement has been proposed (Newman-Griffis and Fosler-Lussier, 2019; Gonen et al., 2020), its efficacy for small corpora with limited vocabularies remains to be determined. Our novel embedding confidence measure offers a step in this direction (see §6.3 for further discussion), but further research is needed.

6.2 Other applications of TextEssence

A previous study on medical records (Newman-Griffis and Fosler-Lussier, 2019) showed that the technologies behind TextEssence can be used for categorical comparison as well as analysis of temporal shifts. More broadly, the use of TextEssence is not limited to comparison of text corpora alone. In settings where multiple embedding strategies are available, such as learning representations of domain concepts from text sources (Beam et al., 2020; Chen et al., 2020), knowledge graphs (Grover and Leskovec, 2016), or both (Yamada et al., 2020; Wang et al., 2020b), TextEssence can be used to study the different regularities captured by competing algorithms, providing insight into the utility of different approaches. TextEssence also can function as a tool for studying the properties of different terminologies for domain concepts, something not previously explored in the computational literature.

In addition, the TextEssence interface can provide utility for other types of analyses as well. For example, the Inspect and Compare portions of the interface could be used to interact with topic models learned from different corpora. These components are largely agnostic to the nature of the underlying data, and could be extended for studying a variety of different types of NLP models.

6.3 Confidence estimation in embedding analysis

The relatively constrained size of corpora in our analysis motivated our novel embedding confidence measure, to help separate differences due to random effects in embedding training from differences in concept usage patterns. We used a fixed confidence threshold for our analyses; however, increasing or decreasing the threshold for high-confidence embeddings will affect both the set of reported neighbors and the visualization of the embedding space, and can inform the user of TextEssence which observations are more or less stable. We highlight varying this threshold as an important area for future investigation with TextEssence.

More broadly, prior work by Wendlandt et al. (2018), Antoniak and Mimno (2018), and Gonen et al. (2020), among others, has also shown embedding stability to be a concern in models trained on larger corpora than those used in this work. However, the role of random embedding effects on previous qualitative studies using word embeddings (e.g., Kulkarni et al. (2015), Hamilton et al. (2016)) has not been evaluated. A broader investigation of embedding confidence measures in qualitative studies will be invaluable in the continued development of embedding technologies as a tool for linguistics research.

7 Conclusion

TextEssence is an interactive tool for comparative analysis of word and concept embeddings. Our implementation and experimental code is available at <https://github.com/drgriffis/text-essence>, and the database derived from our COVID-19 analysis is available at <https://doi.org/10.5281/zenodo.4432958>. A screencast of TextEssence in action is available at <https://youtu.be/1xEEfsMwL0k>. All associated resources for TextEssence may be found at <https://textessence.github.io>.

Acknowledgments

This work made use of computational resources generously provided by the Ohio Supercomputer Center (Ohio Supercomputer Center, 1987) in support of COVID-19 research. The research reported in this publication was supported in part by the National Library of Medicine of the National Institutes of Health under award number T15 LM007059.

References

- Mukhtar H Ahmed and Arez Hassan. 2020. [Dexamethasone for the Treatment of Coronavirus Disease \(COVID-19\): a Review](#). *SN Comprehensive Clinical Medicine*, 2(12):2637–2646.
- Filippo Albani, Federica Fusina, Alessia Giovannini, Pierluigi Ferretti, Anna Granato, Chiara Prezioso, Danilo Divizia, Alessandra Sabaini, Marco Marri, Elena Malpetti, and Giuseppe Natalini. 2020. [Impact of Azithromycin and/or Hydroxychloroquine on Hospital Mortality in COVID-19](#). *Journal of clinical medicine*, 9(9).
- Maria Antoniak and David Mimno. 2018. [Evaluating the Stability of Embedding-based Word Similarities](#). *Transactions of the Association for Computational Linguistics*, 6:107–119.
- Andrew L Beam, Benjamin Kompa, Allen Schmalz, Inbar Fried, Griffin Weber, Nathan Palmer, Xu Shi, Tianxi Cai, and Isaac S Kohane. 2020. [Clinical Concept Embeddings Learned from Massive Sources of Multimodal Medical Data](#). *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 25:295–306.
- Olivier Bodenreider. 2004. [The Unified Medical Language System \(UMLS\): integrating biomedical terminology](#). *Nucleic Acids Research*, 32(90001):D267–D270.
- Gemma Boleda. 2020. [Distributional Semantics and Linguistic Theory](#). *Annual Review of Linguistics*, 6(1):213–234.
- Giovanni Caocci and Giorgio La Nasa. 2020. [Could ruxolitinib be effective in patients with COVID-19 infection at risk of acute respiratory distress syndrome \(ARDS\)?](#)
- Juntian Chen, Yubo Tao, and Hai Lin. 2018. [Visual exploration and comparison of word embeddings](#). *Journal of Visual Languages & Computing*, 48:178–186.
- Qingyu Chen, Kyubum Lee, Shankai Yan, Sun Kim, Chih-Hsuan Wei, and Zhiyong Lu. 2020. [BioConceptVec: Creating and evaluating literature-based biomedical concept embeddings on a large scale](#). *PLOS Computational Biology*, 16(4):e1007617.
- Rene Cutura, Michaël Aupetit, Jean-Daniel Fekete, and Michael Sedlmair. 2020. [Comparing and exploring high-dimensional data with dimensionality reduction algorithms and matrix visualizations](#). In *Proceedings of the International Conference on Advanced Visual Interfaces, AVI '20*, New York, NY, USA. Association for Computing Machinery.
- 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 Information and Knowledge Management - CIKM '14*, CIKM '14, pages 1819–1822, Shanghai, China. ACM.
- Tessa R Englund, Alan C Kinlaw, and Saira Z Sheikh. 2020. [Rise and Fall: Hydroxychloroquine and COVID-19 Global Trends: Interest, Political Influence, and Potential Implications](#). *ACR Open Rheumatology*, 2(12):760–766.
- Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. [Word embeddings quantify 100 years of gender and ethnic stereotypes](#). *Proceedings of the National Academy of Sciences*, 115(16):E3635—E3644.
- Hila Gonen, Ganesh Jawahar, Djamel Seddah, and Yoav Goldberg. 2020. [Simple, Interpretable and Stable Method for Detecting Words with Usage Change across Corpora](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 538–555, Online. Association for Computational Linguistics.
- Alessandro Gozzetti, Enrico Capochiani, and Monica Bocchia. 2020. [The Janus kinase 1/2 inhibitor ruxolitinib in COVID-19](#).
- Aditya Grover and Jure Leskovec. 2016. [Node2Vec: Scalable Feature Learning for Networks](#). In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pages 855–864, New York, NY, USA. ACM.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. [Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.
- Benjamin Hoover, Hendrik Strobelt, and Sebastian Gehrmann. 2020. [exBERT: A Visual Analysis Tool to Explore Learned Representations in Transformer Models](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 187–196, Online. Association for Computational Linguistics.
- Daniel Hornuss, Berit Lange, Nils Schröter, Siegbert Rieg, Winfried V Kern, and Dirk Wagner. 2020. [Anosmia in COVID-19 patients](#). *medRxiv*, page 2020.04.28.20083311.
- Andrew Ip, Donald A Berry, Eric Hansen, Andre H Goy, Andrew L Pecora, Brittany A Sinclair, Urszula Bednarz, Michael Marafelias, Scott M Berry, Nicholas S Berry, Shivam Mathura, Ihor S Sawczuk, Noa Biran, Ronaldo C Go, Steven Sperber, Julia A Piwoz, Bindu Balani, Cristina Cicogna, Rani Sebti, Jerry Zuckerman, Keith M Rose, Lisa Tank, Laurie G Jacobs, Jason Korcak, Sarah L Timmapuri, Joseph P Underwood, Gregory Sugalski, Carol Barsky, Daniel W Varga, Arif Asif, Joseph C

- Landolfi, and Stuart L Goldberg. 2020. [Hydroxychloroquine and tocilizumab therapy in COVID-19 patients-An observational study](#). *PloS one*, 15(8):e0237693.
- Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. [What Does BERT Learn about the Structure of Language?](#) In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3651–3657, Florence, Italy. Association for Computational Linguistics.
- Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. [Statistically Significant Detection of Linguistic Change](#). In *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, pages 625–635, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.
- Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. [Diachronic word embeddings and semantic shifts: a survey](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Mohammed Lester, Ali Sahin, and Ali Pasyar. 2020. [The use of dexamethasone in the treatment of COVID-19](#). *Annals of Medicine and Surgery*, 56:218–219.
- Hailan Li and Huaping Liu. 2020. [Whether the timing of patient randomization interferes with the assessment of the efficacy of ruxolitinib for severe COVID-19](#). *Journal of Allergy and Clinical Immunology*, 146(6):1453.
- S Liu, X Wang, C Collins, W Dou, F Ouyang, M El-Assady, L Jiang, and D A Keim. 2019. [Bridging Text Visualization and Mining: A Task-Driven Survey](#). *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2482–2504.
- Shixia Liu, Michelle X Zhou, Shimei Pan, Yangqiu Song, Weihong Qian, Weijia Cai, and Xiaoxiao Lian. 2012. [Tiara: Interactive, topic-based visual text summarization and analysis](#). *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(2):1–28.
- X. Liu, Z. Zhang, R. Leontie, A. Stylianou, and R. Pless. 2020. [2-map: Aligned visualizations for comparison of high-dimensional point sets](#). In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2539–2547.
- A T McCray, A Burgun, and O Bodenreider. 2001. [Aggregating UMLS semantic types for reducing conceptual complexity](#). *Studies in health technology and informatics*, 84(Pt 1):216–220.
- Apolonia Moll, Mónica Lara, Jaume Pomar, Mónica Orozco, Guiem Frontera, Juan A Llompарт-Pou, Lesmes Moratinos, Víctor González, Javier Ibáñez, and Jon Pérez-Bárcena. 2020. [Effects of dexamethasone in traumatic brain injury patients with pericontusional vasogenic edema: A prospective-observational DTI-MRI study](#). *Medicine*, 99(43).
- Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. 2019. [ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing](#). In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 319–327, Florence, Italy. Association for Computational Linguistics.
- Denis Newman-Griffis and Eric Fosler-Lussier. 2019. [Writing habits and telltale neighbors: analyzing clinical concept usage patterns with sublanguage embeddings](#). In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 146–156, Hong Kong. Association for Computational Linguistics.
- Denis Newman-Griffis, Albert M Lai, and Eric Fosler-Lussier. 2018. [Jointly Embedding Entities and Text with Distant Supervision](#). In *Proceedings of The Third Workshop on Representation Learning for NLP*, pages 195–206. Association for Computational Linguistics.
- Dong Nguyen, Maria Liakata, Simon DeDeo, Jacob Eisenstein, David Mimno, Rebekah Tromble, and Jane Winters. 2020. [How We Do Things With Words: Analyzing Text as Social and Cultural Data](#). *Frontiers in Artificial Intelligence*, 3:62.
- Ohio Supercomputer Center. 1987. [Ohio Supercomputer Center](#).
- Hamid Rahmani, Effat Davoudi-Monfared, Anahid Nourian, Morteza Nabiee, Setayesh Sadeghi, Hossein Khalili, Ladan Abbasian, Fereshteh Ghiasvand, Arash Seifi, Malihe Hasannezhad, Sara Ghaderkhani, Mostafa Mohammadi, and Mir Saeed Yekaninejad. 2020. [Comparing outcomes of hospitalized patients with moderate and severe COVID-19 following treatment with hydroxychloroquine plus atazanavir/ritonavir](#). *Daru : journal of Faculty of Pharmacy, Tehran University of Medical Sciences*, 28(2):625–634.
- Andrew Runge and Eduard Hovy. 2020. [Exploring Neural Entity Representations for Semantic Information](#). In *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 204–216, Online. Association for Computational Linguistics.
- Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. [SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23, Barcelona (online). International Committee for Computational Linguistics.
- Daniel Smilkov, Nikhil Thorat, Charles Nicholson, Emily Reif, Fernanda B. Viégas, and Martin Wattenberg. 2016. [Embedding projector: Interactive visualization and interpretation of embeddings](#).

- Manuela Spadea, Francesca Carraro, Francesco Saglio, Elena Vassallo, Rosanna Pessolano, Massimo Berger, Carlo Scolfaro, Sergio Grassitelli, and Franca Fagioli. 2020. [Successfully treated severe COVID-19 and invasive aspergillosis in early hematopoietic cell transplantation setting.](#)
- Blanca Talavera, David García-Azorín, Enrique Martínez-Pías, Javier Trigo, Isabel Hernández-Pérez, Gonzalo Valle-Peñacoba, Paula Simón-Campo, Mercedes de Lera, Alba Chavarría-Miranda, Cristina López-Sanz, María Gutiérrez-Sánchez, Elena Martínez-Velasco, María Pedraza, Álvaro Sierra, Beatriz Gómez-Vicente, Ángel Guerrero, and Juan Francisco Arenillas. 2020. [Anosmia is associated with lower in-hospital mortality in COVID-19.](#) *Journal of the Neurological Sciences*, 419.
- Laurens van der Maaten and Geoffrey Hinton. 2008. [Visualizing data using t-sne.](#) *Journal of Machine Learning Research*, 9(86):2579–2605.
- Ekaterina Vylomova, Sean Murphy, and Nicholas Haslam. 2019. [Evaluation of Semantic Change of Harm-Related Concepts in Psychology.](#) In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 29–34, Florence, Italy. Association for Computational Linguistics.
- Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, and Others. 2020a. [CORD-19: The Covid-19 Open Research Dataset.](#) *arXiv preprint arXiv:2004.10706*.
- Qingyun Wang, Manling Li, Xuan Wang, Nikolaus Parulian, Guangxing Han, Jiawei Ma, Jingxuan Tu, Ying Lin, Haoran Zhang, Weili Liu, and Others. 2020b. [Covid-19 literature knowledge graph construction and drug repurposing report generation.](#) *arXiv preprint arXiv:2007.00576*.
- Vincent Warmerdam, Thomas Kober, and Rachael Tatman. 2020. [Going Beyond T-SNE: Exposing what lies in Text Embeddings.](#) In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 52–60, Online. Association for Computational Linguistics.
- Rebecca Weiss. 2014. [MUCK: A toolkit for extracting and visualizing semantic dimensions of large text collections.](#) In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 53–58, Baltimore, Maryland, USA. Association for Computational Linguistics.
- Philippa M Wells, Katie J Doores, Simon Couvreur, Rocío Martínez Nunez, Jeffrey Seow, Carl Graham, Sam Acors, Neophytos Kouphou, Stuart J D Neil, Richard S Tedder, Pedro M Matos, Kate Poulton, Maria Jose Lista, Ruth E Dickenson, Helin Sertkaya, Thomas J A Maguire, Edward J Scourfield, Ruth C E Bowyer, Deborah Hart, Aoife O’Byrne, Kathryn J A Steel, Oliver Hemmings, Carolina Rosadas, Myra O McClure, Joan Capedevilla-pujol, Jonathan Wolf, Sebastien Ourselin, Matthew A Brown, Michael H Malim, Tim Spector, and Claire J Steves. 2020. [Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England.](#) *Journal of Infection*, 81(6):931–936.
- Laura Wendlandt, Jonathan K Kummerfeld, and Rada Mihalcea. 2018. [Factors Influencing the Surprising Instability of Word Embeddings.](#) In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2092–2102. Association for Computational Linguistics.
- Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, and Yuji Matsumoto. 2020. [Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia.](#) In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 23–30, Online. Association for Computational Linguistics.
- Yiyun Zhao and Steven Bethard. 2020. [How does BERT’s attention change when you fine-tune? An analysis methodology and a case study in negation scope.](#) In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4729–4747, Online. Association for Computational Linguistics.
- Tamara Zhordaniya, Andrey Kutuzov, and Elizaveta Kuzmenko. [Vec2graph: A python library for visualizing word embeddings as graphs.](#) Springer.