Word embedding for social sciences: an interdisciplinary survey

Akira Matsui\(^1\) and Emilio Ferrara\(^2\)

\(^1\) College of Business Administration, Yokohama National University, Yokohama, Kanagawa, Japan

\(^2\) Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, California, United States

ABSTRACT

Machine learning models learn low-dimensional representations from complex high-dimensional data. Not only computer science but also social science has benefited from the advancement of these powerful tools. Within such tools, word embedding is one of the most popular methods in the literature. However, we have no particular documentation of this emerging trend because this trend overlaps different social science fields. To well compile this fragmented knowledge, we survey recent studies that apply word embedding models to human behavior mining. Our taxonomy built on the surveyed article provides a concise but comprehensive overview of this emerging trend of intersection between computer science and social science and guides scholars who are going to navigate the use of word embedding algorithms in their voyage of social science research.

INTRODUCTION

The advancement of machine learning technologies has revolutionized traditional research methodologies in the social sciences, enabling the extraction of insights from unconventional data sources\(^1\). Recently, we have been witnessing the expansion of social science research that explores textual data, captivated by the concept of “Text as Data” (Gentzkow, Kelly & Taddy, 2019; Benoit, 2020; Grimmer & Stewart, 2013). This trend is evident not only in fields that traditionally rely on quantitative analysis, such as economics (Cong, Liang & Zhang, 2018) and finance (Rahimikia, Zohren & Poon, 2021), but also in disciplines with a penchant for qualitative analysis, like history (Wevers & Koolen, 2020). The flourishing body of literature in this domain enables social scientists to meticulously select or develop machine learning models for text mining that meet the demands of their research, providing a vast array of options. While rule-based models, such as dictionary-based models, remain popular (Tausczik & Pennebaker, 2010; Hutto & Gilbert, 2014), neural network-based embedding models are gaining increasing traction in social science research.

Embedding models learn low-dimensional representations of entities from data by predicting relationships between these entities (Boykis, 2024). The advantage of embedding...
models in social science research lies in their ability to acquire the “embedded” structure in an unsupervised manner. This inherent flexibility allows social scientists to apply these models to their data with minimal constraints, facilitating their analyses. These entities could be nodes in a graph (Goyal & Ferrara, 2018; Barros et al., 2021; Cai, Zheng & Chang, 2018), pixels in an image (Liu et al., 2020; Baltrušaitis, Ahuja & Morency, 2018), or vertices in a network (Cui et al., 2018; Zhou et al., 2022a). Despite the multitude of embedding algorithms available for diverse purposes, word embedding models have been recognized as some of the most widely adopted methods (Chaubard et al., 2019; Mikolov et al., 2013a, 2013b; Goldberg & Levy, 2014; Mikolov, Yih & Zweig, 2013; Mikolov et al., 2015). Among these, the most accessible word embedding model for social science research is the word2vec model (Mikolov et al., 2013b; Mikolov, Yih & Zweig, 2013; Mikolov et al., 2013a), used in diverse fields ranging from management science (Lee & Kim, 2021; Toubia, Berger & Elashberg, 2021; Chanen, 2016; Li et al., 2021; Bhatia et al., 2021; Chen & Zhang, 2023; Zheng, 2020; Khaouja et al., 2019; Tshitoyan et al., 2019) to poetry (Nelson, 2021).

The rise in the use of word2vec and other related models in social science research has not been accompanied by adequate documentation of how these models are utilized. However, because this trend spans multiple disciplines, we are witnessing a lack of sufficient cross-communication. First, this absence of an integrated knowledge framework hampers discussions about the broader impact and relevance of word embedding in social science research. Second, the absence of comprehensive surveys spanning multiple disciplines has led social scientists to reference only literature within their own fields, inadvertently leading to the reinvention of methodologies already developed in other domains. Consequently, this solid approach often results in deserving scholarly works not receiving the recognition and citation they merit, thus undermining their rightful academic evaluation. Therefore, we need to review the broad spectrum of studies that should be referenced, many of which fall outside their immediate expertise.

To address the knowledge gap resulting from a lack of connections between disciplines, this article aims to identify emerging interdisciplinary trends in machine learning models arising across diverse fields. In this literature review, we seek to bridge the knowledge gaps between different fields and ask the research question “How similarly are word embedding models, such as word2vec, applied across different methodologies in social science research?” This literature review builds a taxonomy of these application methodologies to illustrate how social scientists harness word embedding models, particularly focusing on word2vec. The taxonomy emphasizes the applications of word embedding published in social science journals, outlining the contemporary landscape of social science research utilizing word embedding models, rather than delving into the technical intricacies of these neural network models. This literature review targets social scientists who are currently using or considering the use of word embedding models like word2vec in their research. Therefore, the presented article does not comprehensively review the state of the art in the technical aspects of embedding models. Instead, we offer a review that covers various social science disciplines, especially for those who may not specialize in machine learning or natural language processing. To provide an overview of such different fields, we develop a
taxonomy to guide potential readers in situating their studies within this interdisciplinary literature.

This interdisciplinary survey covers literature published across a wide range of research fields, helping researchers or practitioners access studies and insights outside their own areas of expertise. In addition, the taxonomy developed in this survey helps researchers in positioning their work within the context of the literature, which is too broad to be referred in a article in a single field. In this way, the interdisciplinary literature review offered here contributes to bridge barriers among research fields. This survey also helps us understand and enhance the application of word embedding models in social science contexts.

The structure of this survey is as follows. After the discussion of search methodology, we begin this survey with a brief overview of the technical background of word2vec in “Word Embedding Models”. Then, we introduce the nine labels used in our taxonomy (Table 1) and review the articles detailing the taxonomy throughout “Pre-trained Models” to “Non-text”. We discuss a pitfall of similarity measurements with a simple experiment. In addition, we employ the proposed taxonomy to understand the literature using the word2vec model for social science.

**SURVEY METHODOLOGY**

This survey primarily focuses on articles published in social science journals, aiming to conduct a concentrated literature survey. Although our main attention was not directed toward research in word embedding published in computer science conference proceedings, we selectively included some of these articles to ensure a comprehensive understanding of the field. This approach was essential to paint a complete picture of the literature landscape.

We primarily focused on articles that had a clear application or discussion of word embedding models in a social science context. We gave preference to articles published in peer-reviewed journals and conference proceedings to guarantee the credibility and academic rigor of our sources. Conversely, we excluded articles that did not fall within our specified subject areas. Non-English publications were also excluded due to language.

| Labels       | Definition                                                                 |
|--------------|-----------------------------------------------------------------------------|
| Pre-trained  | Pre-trained model                                                            |
| Working variable | Variables constructed from a word embedding model for analysis             |
| W/theory    | Validating a theory proposed in a research field for analysis, or quoting it to justify an analysis. |
| W/reference | Using references such as the words with specific semantics to define        |
| Non-text    | Applying word embedding model to non-text data (e.g., human behavior data) |
| Same words  | Comparing the same words from different models                             |
| Human subj  | Employing human subjects for analysis                                       |
| Clustering  | Clustering analysis using word embedding vectors                            |
| Prediction  | Prediction using word embedding vectors                                      |
constraints. This exclusion was necessary to ensure that our sources were reproducible and verifiable by our readers.

For this survey, we predominantly used the Scopus Database, implementing a specific query that included terms word2vec and combinations such as word AND embedding. To ensure that our survey was as exhaustive as possible, we expanded our search to include Google Scholar and the Social Science Research Network (SSRN) to include working articles or preprints, adapting our query to suit their unique search algorithms. This process is not limited to research articles but also includes other types of publications, such as articles, book chapters, and review articles. Additionally, we compile articles that cite (Mikolov et al., 2013b; Mikolov, Yih & Zweig, 2013) to capture the literature on the application of word2vec. From the collected articles, we select those published in fields related to social sciences. The specific fields we cover are listed in the research area column in Tables 2 and 3.

Through this meticulous and methodical approach, we aimed to ensure a comprehensive and unbiased coverage of the literature. This strategy was crucial in facilitating a robust analysis of the current state of word embedding applications in social science research, thereby contributing significantly to our understanding of the field. This survey focuses on the foundational role of the word2vec model introduced by Mikolov et al. (2013a) and we review the literature from 2013 to the present (2024). In addition to the word2vec application, we also acknowledge earlier research, including the seminal work of Rubenstein & Goodenough (1965) from 1965.

WORD EMBEDDING MODELS

We briefly review the technical aspects of word embedding models before diving into the main part of our survey. Since our focus is the literature that uses word2vec models, this section discusses the word2vec algorithm. This section ensures that the presented article provides sufficient technical prerequisites of word2vec, while in-depth explanations are relegated to relevant literature sources.

Word embedding models learn the relationships between words in sentences in the data and return embedding vectors based on their learned model. The embedding vectors are considered to represent the semantics of the words in the sentence, and the distributional hypothesis Harris (1954) supports this assumption. The theory asserts that a word is determined by the surrounding words, and experiments with human subjects have validated that word embedding models can capture semantics similar to human cognition (Caliskan, Bryson & Narayanan, 2017). This prompts social scientists to assume that studying word embedding models learned from the text of interest can reveal how humans process information embedded in the text.

Evaluations of word embedding models roughly take two forms: intrinsic evaluation and extrinsic evaluation. Typical tasks for intrinsic evaluations are analogy and similarity. Many datasets for intrinsic evaluation are publicly available and accessible (Analogy (Miller & Charles, 1991; Charles, 2000; Agirre et al., 2009; Rubenstein & Goodenough, 1965; Bruni, Tran & Baroni, 2014; Radinsky et al., 2011; Huang et al., 2012; Luong, Socher & Manning, 2013; Hill, Reichart & Korhonen, 2015), Similarity (Mikolov et al., 2013a;
For extrinsic evaluation, they often use word embedding models for translation or sentiment analysis. See Wang et al. (2019) for recent evaluation methods.

Although there are various word embedding models, social scientists have favored word2vec as the most popular model due to its simplicity, high user-friendliness, and longstanding usage (Chaubard et al., 2019; Rong, 2014). Thus, this survey article primarily focuses on the application of word2vec in social science research while also covering other language models that describe the new trend of this interdisciplinary research.
word2vec: a popular word embedding model

We begin the explanation of the technical background of word2vec by clarifying the terminology of “word2vec,” because the term does not designate a specific neural network model or algorithm but rather refers to a software program and tool. Pedagogically speaking, word2vec denotes software that combines a learning algorithm and a training method (Chaubard et al., 2019). For such packages, the continuous bag of words model (CBOW) and the skip-gram model are common implementations within word2vec, and the negative sampling algorithm is a widely-used technique (Mikolov et al., 2013a, 2013b; Goldberg & Levy, 2014; Mikolov, Yih & Zweig, 2013; Mikolov et al., 2015). While computer science articles usually describe the details of their models, social science articles often use word2vec as if it refers to a specific neural network model, and discussions of learning algorithms and training methods are not very common in social science articles except in those discussing research methods such as econometrics or statistical models. Therefore,
when a article mentions only “word2vec” for their methods, it may use the CBOW or the skip-gram model\(^3\).

**CBO\(\text{W}, \text{skip-gram model and SGNS model}\)**

The CBOW and skip-gram solve a “fake” problem that predicts what words appear in a given text. The main goal of the model is not to make predictions but to acquire the parameters obtained by solving the prediction problem as a low-dimensional representation of words. In other words, the prediction problem solved by word embedding is only to learn the parameters, and the prediction problem itself is a problem that is not of central interest. Given this background, we refer to prediction problems that word embedding solves to learn the low-dimensional representation as “fake” problems. To explain this fake problem in the word embedding model, we often use the terms “word” and “context” \(^4\). Target is the element in the text and is often a word. Context is the set of elements around the target and is often the set of words.

Consider the sentences, The stick to keep the bad away. The rope used to bring the good toward us. When the target is rope, its context could be [bad, away, the, used, to, bring] with a window size of three. The window size determines the context size and the number of elements to capture from the target. This example considers the three words to the left and right of the target word. While both the CBOW and skip-gram models solve the prediction model using the context and target, they differ in which part predicts the other. The CBOW model predicts the target \(w\) using the context \(c\), \(P(w \mid c)\). Conversely, the skip-gram model predicts the context using the target, \(P(c \mid w)\). The most used word2vec is the skip-gram with negative sampling (SGNS) model that combines the skip-gram model and negative sampling algorithm because this combination is provided in the popular ready-to-use library gensim (Rehurek & Sojka, 2010a, 2010b).

The SGNS model considers the distribution, \(p(d \mid w, c)\), where \(d\) takes 1 when a pair of target \(w\) and context \(c\) is observed in the data, 0 otherwise. The SGNS model maximizes the following conditional log-likelihood (Dyer, 2014),

\[
\sum_{(w,c) \in D} \{ \log p(d = 1 \mid c, w) + k \mathbb{E}_{\tilde{w} \sim q} \log p(d = 0 \mid c, \tilde{w}) \} \tag{1}
\]

where \(q\) is the noise distribution in negative sampling, and \(k\) is the sample size from the noise\(^5\); \(D\) is the set of pairs of target \(w\), context \(c\), \(\mathbb{E}_{\tilde{w} \sim q}\) represents the expectation under the probability distribution of \(q(\tilde{w})\), and \(k\) refers to the number of negative samples\(^6\).

The SGNS model especially calculates the conditional probability \(p(d = 1 \mid c, w)\) by \(\sigma(v_c \cdot v_w)\), where \(\sigma(x)\) is sigmoid function and \(v_w, v_c \in \mathbb{R}^d\). In other words, it seeks the parameter \(v_w, v_c\) that maximizes the above conditional probability, which are the embedding vectors of interest.

The word2vec\(\text{f}\) model can construct an arbitrary context for the word2vec model. While a typical word2vec model constructs a set of words around the target word in its learning process, the word2vec model does not necessarily use the surrounding words of the target word for a set; it can be arbitrary. Since context determines the relationship between targets and words in the data, we can define the context based on external information.

---

3 In this survey, we review the articles (Ash, Chen & Ornaghi, 2020; Lee & Kim, 2021; Caliskan, Bryson & Narayanan, 2017) that uses GloV model (Pennington, Socher & Manning, 2014; Kawaguchi, Kuroda & Sato, 2021) that uses fastText model (Bojanowski et al., 2017).

4 While “target” is also popularly used instead of “word,” we stick to using “word” in this survey to avoid potential confusions.

5 In the literature, it is often mentioned that this log-likelihood maximization with the noise distribution is considered as Noise Contrastive Estimation (NCE) (Gutmann & Hyvärinen, 2010), and the SGNS model is one of the variations of NCE.

6 Gutmann & Hyvärinen (2010) describes the details, and this article also follows the notation and equations of that article.
Levy & Goldberg (2014), for example, propose the word2vecf model, a modified version of the word2vec model, which obtains word embedding vectors by constructing the context based on the dependency between words. Moreover, the relationships to be learned can be in non-text data as we discuss in “Non-text”.

TAXONOMY OF APPLIED METHODS AND LABELING LITERATURE

Now, we turn to the main focus of this survey article, which provides an overview of a new emerging trend that applies word embedding models to social science research. To give a clear overview of the applications of word embeddings in interdisciplinary research, we first develop a taxonomy of the analytical methods used in the surveyed social science articles. We then label the literature based on the developed taxonomy. This approach allows us to review the studies from a concrete perspective on the trend of word embedding applications, which spans a wide range of disciplines.

We first review the literature on applications of word embedding models for social science. Our taxonomy introduces the labels of popular methods found in the surveyed literature. This method-oriented classification benefits us when surveying interdisciplinary topics. The nature of these topics prevents us from discussing each article according to its research topics or questions because showcasing such highly diverse outcomes dilutes the visibility of the survey. Therefore, instead of following the literature of each research field, we study the way the surveyed articles use word embedding models for their analyses.

We discuss our taxonomy and labels with the literature in “Taxonomy”, and we summarize the labels in Table 1. We define eight labels and summarize the descriptive definitions of the analytical methods. We also list the surveyed articles in Tables 2 and 3 with the corresponding labels. This process identifies several important research orientations in the literature. For example, we identify a popular method that constructs variables for study using reference words (W/reference label). In this survey, we also observe a new line of literature that applies word2vec to non-textual data, such as relationships between users on web platforms.

Labeling literature

The central aim of this survey is to label the literature on applications of word embedding models for social science. To this aim, we first build the taxonomy on the literature in “Taxonomy” and label them based on them.

In addition, we group the articles by their research topics based on the journals or conferences in which they were published. For working articles, we infer the topic based on the author’s faculty affiliation and past publication history to determine the research area, resulting in 14 research areas. Note that we understand that defining or classifying research fields is not a trivial task.

Taxonomy

This section describes the taxonomy built with the labels introduced in “Labeling Literature”. We discuss several representative articles for each methodological label.
**Pre-trained models**

Word embedding models extract information from text data for research questions. Obtaining low-dimensional representations from data allows us to extract the semantics or hidden patterns in the data. When social scientists are interested in a specific subject, they train a word embedding model on their own data, such as descriptions of central banks (Matsui, Ren & Ferrara, 2021; Baumgartner & Zahner, 2021), judges (Ash, Chen & Ornaghi, 2020), organizations (Chen & Zhang, 2023), job openings (Khaouja et al., 2019), or smartphone applications (Kawaguchi, Kuroda & Sato, 2021). To study these topics, researchers build their own datasets and train their word embedding models on the data.

While most studies construct their own word embedding models, pre-trained word embedding models are also popular, with articles using pre-trained models published in prestigious journals (Bhatia, 2019; Kozlowski, Taddy & Evans, 2019; Garg et al., 2018). Pre-trained word embedding models are ready-to-use models that have already been trained on large corpora. Most pre-trained models are trained on general corpora and are open to the public, such as the Google News pre-trained model (Google, 2013) and pre-trained models on historical corpora in several languages (Hamilton, Leskovec & Jurafsky, 2016b). Models trained on general corpora are supposed to represent general semantics, allowing studies with these pre-trained models to answer broad research questions, such as evaluating biases in publications (Garg et al., 2018; Caliskan, Bryson & Narayanan, 2017) and culture (Kozlowski, Taddy & Evans, 2019).

Studies have demonstrated that employing human subjects to validate the findings of pre-trained models can aid our understanding of crucial human perceptions (Caliskan, Bryson & Narayanan, 2017; Bhatia et al., 2021; Zheng, 2020). Such studies often use models pre-trained on general corpora to understand common human biases. For example, management scientists combine surveys from human subjects and pre-trained models to investigate essential concepts in their discipline, such as perceptions of risk (Bhatia, 2019) and leadership (Bhatia et al., 2021). Additionally, because of their generality, pre-trained models help researchers validate theories proposed in their disciplines. For example, researchers in neuroscience propose features for interpreting word embeddings based on a method called neurosemantic decoding (Chersoni et al., 2021).

**Overfitting models**

Rather than focusing on general social phenomena, social science research often targets specific subjects or classes. In such studies, researchers often “overfit” word embedding models on corpora that represent specific subjects. Overfitting is typically avoided by computer scientists to ensure the generality of their trained models since overfitted models do not perform well on computational tasks such as analogical reasoning or generating language models, as discussed in “Word Embedding Models”.

However, social science researchers are also interested in the results from overfitted models because these models can capture the specificity of the data. There is a line of research, for example, that investigates human biases, such as stereotypes, “embedded” in the text of interest (Bhatia & Bhatia, 2021; Ash, Chen & Ornaghi, 2020; Kroon, Trilling & Raats, 2021; Garg et al., 2018; Charlesworth, Caliskan & Banaji, 2022;
Boutyline, Arseniev-Koehler & Cornell, 2023; Durrheim et al., 2023). Word embedding models learned from the corpora can reveal biases where certain words are overfitted with specific human characteristics, such as gender, and these computed biases are similar to the biases humans hold (Caliskan, Bryson & Narayanan, 2017). In the context of fairness in machine learning, significant computer science literature proposes methods for debiasing such biases (Bolukbasi et al., 2016; Zhao et al., 2018; Brunet et al., 2019; Font & Costa-Jussa, 2019) and evaluates their validity (Gonen & Goldberg, 2019).

**Working variable**

With trained word embedding models, most surveyed articles construct variables that embody the concepts or research questions to be examined in their study, while some articles use word embedding vectors for clustering (Hu et al., 2020; Nanni & Fallin, 2021; Zheng, 2020; Khaouja et al., 2019; Kim et al., 2021) or prediction (Bhatia et al., 2021; Richie et al., 2019; Lee & Kim, 2021; Wang, 2019) tasks. To clarify this trend, we call this procedure the working variable” using an analogy from working hypothesis,” which is a well-used term in social science research (Taylor, 2021). A working hypothesis is testable using empirical analysis and is often transformed from a conceptual hypothesis or theoretical hypothesis. While a theoretical hypothesis is a conceptual description of the hypothesis to be proved in the research, a working hypothesis describes an executable procedure or test that can prove the theoretical hypothesis. This method allows social scientists to transform a theoretical hypothesis into a working hypothesis that can be tested by experimental or observational research.

In this survey, we define “working variable” as a variable that is a proxy variable of theoretical interest. With working variables, researchers investigate theoretical hypotheses in various fields and external objects such as human perception. For example, to analyze stereotypes in the documents of interest, researchers can use word embedding to calculate some working variables that proxy stereotypes (Garg et al., 2018). Not only human perception, but also qualitative concepts can be defined as working variables such as the speed of semantics in documents like research articles (Soni, Lerman & Eisenstein, 2021; Toubia, Berger & Eliashberg, 2021), law documents (Soni, Lerman & Eisenstein, 2021), and the plots of movies and TV shows (Soni, Lerman & Eisenstein, 2021; Toubia, Berger & Eliashberg, 2021).

Defining working variables allows us to analyze social phenomena or concepts of interest through concepts or theories in social science. For example, working variables can be proxies of important concepts, such as opacity (Boyce-Jacino & DeDeo, 2020) or grammar (Čech et al., 2019). With embedding models, authors can construct their proxies for their research question (Johnson et al., 2023). Such variables are often defined as a scalar value obtained by the embedding model learned in the data of interest. Calculating these working variables usually employs the reference words described in “Reference Words”. Defining not only a single working variable but also multiple working variables clarifies the research questions. Toubia, Berger & Eliashberg (2021), for instance, quantified the shape of stories by defining several variables such as speed, volume, and circuitousness in the text.
There are studies that directly calculate vectors obtained from word embedding models and some researchers use word embedding models to estimate their own statistical model for causal inference (Wang, 2019) and equilibrium estimation (Kawaguchi, Kuroda & Sato, 2021), rather than analyzing the subject represented using the embedding vectors.

**Reference words**

To construct working variables, the literature often introduces “reference words” that serve as a point from which researchers calculate the distance of word embedding vectors. In this method, “reference” words represent a concept to be analyzed.

For example, Garg et al. (2018) analyze the gender stereotypes in text over one century by calculating the relative similarity between the words related to gender (men or women) and the words related to specific occupations. To calculate the relative similarity, they first calculated the distance between the word embedding vectors of words related to men and the words describing specific occupations. They also calculated the same distance for women-related words. Then, they computed the relative differences between these two distances. When the relative distance with an occupation is large, it means that the stereotype in the text is considerable. For example, they determined that “engineer” is close to the men-related words in the historical corpus and argued that this suggests the existence of stereotypes in the corpus. The article showed that the measured gender stereotypes are correlated with occupational employment rates by gender. They also conducted the same analysis for ethnic stereotypes.

There are two advantages to using reference words for analysis. The first is that introducing references improves the interpretability of the results; if we use the Subjectivity Lexicon as the reference word, we can measure the degree to which the word of interest is subjective (Matsui, Ren & Ferrara, 2021). Secondly, this method is free from the problems associated with coordinate systems when comparing different models detailed in this point in “Comparing the same words”.

**Comparing the same words**

Some prefer simple methods of analyzing the similarity between words. For example, this category contains the analysis of enumerating words that are similar to a certain word, which is often done in word embedding tutorials because the similarity of each word is calculated to list similar words. Some articles have used this simple analysis as seminal work and have yielded useful findings (Chanen, 2016).

Calculating the similarity between the same words from different models could also answer important research questions for social science. The same word often has different meanings depending on the speaker and the time period in which the word was used. Examining such diachronic changes in semantics or meaning by the speaker can address profound research questions in social science. Nyarko & Sanga (2020), for example, examined the difference in perception between experts and laypeople by investigating whether these two groups use the same word to describe different semantics according to their profession.
When comparing the semantic differences of the same word from two different documents, we need to train two separate word embedding models on each document. In such a situation, we should note that we cannot directly compare the two models, even if they were trained with the same algorithm because algorithms typically use an arbitrary coordinate system with each training. To address this issue, **Hamilton, Leskovec & Jurafsky (2016a)** uses orthogonal Procrustes to align different models to reveal historical changes in the semantics of words, and some other studies follow this method (**Nyarko & Sanga, 2020**). Orthogonal Procrustes, however, may not be stable (**Gillani & Levy, 2019**). Therefore, an improved method using the dynamic word embedding model, which explicitly incorporates semantic changes, has been proposed (**Gillani & Levy, 2019; Yao et al., 2018**), and some articles use the dynamic word embedding model to analyze semantic changes (**Matsui, Ren & Ferrara, 2021**). Thanks to recent studies on the statistical properties of word embedding models (**Arora et al., 2016**), researchers have proposed a way to compare different models (**Zhou, Huang & Zheng, 2020**). Without a model alignment, which will advance the use of word embedding models from a more social science perspective. Additionally, this problem can be avoided if the comparison is done within the same model and is projected onto a scalar value, **Garg et al. (2018)** calculate stereotypes as scalar values and investigate the trajectory of the values over time. As discussed in “Reference Words”, this is the second advantage of introducing reference words that allow us to avoid the problems stemming from the issue of word embedding models using an arbitrary coordinate system in their learning.

**Non-text**

This survey observes a recent trend that applies word2vec to non-textual data. As discussed in “CBOW, Skip-gram Model and SGNS Model”, word embedding can learn relationships between not only words but also any entities such as metadata (**Rheault & Cochrane, 2020**), and symbols (**Tshitoyan et al., 2019**), in addition to text data. This subsection also surveys works that apply word2vec algorithms (skip-gram or CBOW) to digital platforms and geographic data (**Mikolov et al., 2013a, 2013b; Goldberg & Levy, 2014; Mikolov, Yih & Zweig, 2013; Mikolov et al., 2015**).

The relationships that word embedding models learn are not limited to words in text. Word embedding models can obtain low-dimensional representations from relationships composed of non-text data. For example, **Hu et al. (2020)** and **Niu & Silva (2021)** obtain embedding vectors of Points of Interest (POI) from geological data describing the relationships between POIs. With their clustering algorithm, they suggested it can discover the usage pattern of urban space with reasonable accuracy. **Murray et al. (2020)** noted an important theoretical insight for analyzing geographic data with word2vec. They revealed that word2vec is mathematically equivalent to the gravity model **Zipf (1946)**, which is often used to analyze human mobility, and conducted experiments with real datasets to validate their findings.

Word embedding models can also learn relationships between chemical equations and the text around them in a research article. **Tshitoyan et al. (2019)** apply word2vec to academic articles in chemistry to learn the relationship between text and symbols.
(chemical formulae). They demonstrate that potential knowledge about future discoveries can be obtained from past publications by showing that the obtained embedding model can recommend discoveries several years in advance of the actual discovery. Not only symbols but also other information can be embedded. Rheault & Cochrane (2020) analyze politicians’ speeches using the word2vec algorithm to embed information about politicians’ ideologies along with the text data of their speeches.

Non-text data may yield useful insights even when text data is available in the dataset. Kim, Askin & Evans (2024) learn the representation of music track by applying the word2vec algorithm to the sequence of music listening history. Waller & Anderson (2021) adopt a modified version of the word2vec algorithm, using algorithm called word2vecf, which is a modified version of word2vec (Levy & Goldberg, 2014; Goldberg, 2017), to analyze the posting behavior of Reddit users. They obtained embedding vectors by learning the relationships between Reddit users and communities (subreddits). It is worth noting that they do not analyze the textual data of Reddit posts to investigate users’ behavior. Studying Reddit text data might unveil the attributes of Reddit users, such as demographic information, preferences, and age groups. However, user comments do not always reflect such meta-information. Some supervised machine learning models may predict such meta-information; there are successful unsupervised methods that predict the characteristics of text, such as sentiment analysis (Pennebaker, Francis & Booth, 2001; Bird, Klein & Loper, 2009; Hutto & Gilbert, 2014) or moral foundations dictionary (Graham et al., 2011; Haidt & Graham, 2007). Moreover, the validity of training data is difficult to prove due to issues regarding the annotator’s subjectivity. To overcome this problem, the authors used a data-driven and linguistically independent method that characterizes subreddits only from the perspective of “what kind of users are posting in the community.” By setting reference “words” (subreddits), they created indicators such as age, gender, and partisanship. For example, to calculate the working variable of “partisanship,” they chose opposing communities in Reddit: “democrats” and “conservatives.” Then, they calculated the relative distance for each community. Applying our taxonomy to Waller & Anderson (2021) reveals that, while they do not use text, they share characteristics with other studies in their methodologies. They define a working variable (partisanship) using reference words (subreddits). In addition, they are similar to the analysis by Toubia, Berger & Eliashberg (2021) in terms of setting multiple working variables.

Several studies apply the word embedding algorithm to learn relationships between text and non-text data or between non-text entities. As discussed in the introduction, this survey primarily focuses on the applications of word embedding models and algorithms. Therefore, the non-text label does not cover embedding models for non-text data if they are not word embedding models, such as graph embedding (Goyal & Ferrara, 2018; Barros et al., 2021), image embedding (Liu et al., 2020; Baltrušaitis, Ahuja & Morency, 2018), or network embedding (Cui et al., 2018; Zhou et al., 2022a).

**COSINE SIMILARITY OR EUCLIDEAN DISTANCE?**

We find that many articles examine the similarity between embedding vectors when they employ the methods discussed so far such as reference words (“Reference Words”) or non-
For similarity measurement, most use a cosine similarity but Euclidean distance ($L_2$ norm) is also popular. Because the similarity between vectors can be interpreted as the distance between vectors, we have various options to define the distance between vectors⁷. However, it is not trivial whether the results can change depending on the measure we use for analysis. Some articles analyze the robustness of the results using multiple options, such as Euclidean distance and cosine similarity Garg et al. (2018).⁸

Toubia, Berger & Eliashberg (2021), in their supplementary information, argues that Euclidean distance is richer than cosine similarity that only measures the angle. Indeed, cosine similarity is not a perfect alternative to Euclidean distance. Cosine similarity can disregard the information on the distance between two given vectors. Figure 1 depicts five different vectors. Vectors 1 and 2 have the same angle $\theta_1$, and therefore, the cosine similarity of these is one. In contrast, the Euclidean distance between the two is $d$, which means they are different in terms of Euclidean distance.

It is important to recognize that equal Euclidean distances do not always correspond to equivalent cosine similarity values and that Euclidean distance cannot always be interchanged with cosine similarity, nor vice versa. The Euclidean distance between Vectors 2 and 3 is the same as the Euclidean distance between Vectors 1 and 2, but these two pairs have different angles ($\theta_1$ and $\theta_2$). Figure 1 also depicts that using the normalized Euclidean distance eliminates the distance between Vectors 1 and 2. The normalized Vector 1 will be on the same point as Vector 2, which is on the unit circle. In this case, these two vectors are the same from the perspective of both cosine similarity and Euclidean distance. Also, we should note that small distances do not always mean small cosine distances. For example, Fig. 1 depicts Vectors 4 and 5 is small in Euclidean distance compared to the other vectors, but their angle is $\theta_1$ (approximately $\pi/4$); therefore, the cosine similarity between them is not small.

---

⁷ We also should note that cosine similarity is not a distance metric.

⁸ Regarding the robustness check, Ash, Chen & Ornaghi (2020) also studied the correlations between their stereotype measurements of 100 and 300 dimension embedding vectors. In addition, they tested three sets of window sizes in their robustness check.
Specifically, the relationship between two measurements becomes dismissed when $\theta$ is small or around $\pi$. Let us consider the distance between two points in a unit circle: $A (0, 1)$ and $B (\cos \theta, \sin \theta)$. We can consider two points in a unit circle by $(x, y)$ and $(x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$, where the angle between the two is $\theta$. Since we only study the distance between the two, we set $x = 1$ and $y = 0$ without loss of generality. The distance between two is $(2(1 - \cos \theta))^{1/2}$, and the cosine similarity is $\cos \theta$. Therefore, the error between the Euclidean distance of normalized vectors and cosine similarity is a function of $\theta$, and that is

$$e(\theta) = \cos \theta - (2(1 - \cos \theta))^{1/2}. \quad (2)$$

Since the first derivative of $e(\theta)$ with respect to $\theta$ is

$$e'(\theta) = -\sin(\theta) \left( 1 + (2(1 - \cos(\theta))^{1/2}) \right), \quad (3)$$

the relationship between the two metrics become zero around ($\theta = \pi$ or 0) where angle $\theta \in R$. This fact implies that the two metrics are incompatible when their angle is small or large. We also note that the fact that the angle is not always relevant to the distance.

To further discuss the comparison between the cosine similarity and Euclidean distance, we study the empirical relationship between them using the Google News pre-trained word2vec model that contains the embedding vectors of 3,000,000 words [Google (2013)]. We construct 1,500,000 random word pairs and plot the relationships in Figs. 2A and 2B. While Fig. 2A demonstrates that the two metrics are correlated, their correlation is not strong (the Pearson correlation coefficient $\rho = -0.357$), which implies that using cosine similarity would lose some information and vice versa. Figure 2A also demonstrates that the cosine similarity distributes uniformly where the Euclidean distance is small, meaning that the two different pairs of vectors with the same Euclidean distance can assume different values in their cosine similarities. This deviation becomes large where the Euclidean distance is small. This finding is consistent even when we normalize the metrics. Figure 2B demonstrates the empirical relationship between the Euclidean distance of normalized vectors and cosine similarity. While they capture the difference of two word embedding vectors in a similar way (the Pearson correlation coefficient $\rho = -0.991$), they are not the same. The Euclidean distance of normalized vectors is more sensitive than cosine similarity in which the distance between two points is close (i.e., $\theta$ is small) or large (i.e., $\theta$ is around $\pi$).

The above discussion suggests that the cosine similarity and Euclidean distance can qualitatively return in a similar manner, but they are not always compatible. Notably, the cosine similarity can capture the difference between two vectors when their Euclidean distance is small. However, when we study the normalized vectors, the Euclidean distance captures the differences, but the cosine similarity does not when the Euclidean distance is small. Given this, we should know what characteristics we intend to capture when selecting a metric.
These findings also imply that the Euclidean distance and cosine similarity (angle) capture different characteristics. Recently, some studies in computer science literature have proposed word embedding models that explicitly model this relation, such as hierarchical relationships (Tifrea, Bécigneul & Ganea, 2018; Nickel & Kiela, 2017; Vendrov et al., 2016; Vilnis & McCallum, 2015). For example, (Iwamoto, Kohita & Wachi, 2021) proposed the embedding models that use the polar coordinate system and illustrate that their model captures the hierarchy of words in which the radius (distance) represents generality, and the angles represent similarity.

**Pitfalls with cosine similarity for similarity measurement with word embedding vectors**

Among the surveyed articles, cosine similarity is the most popular similarity measurement but the recent literature points out several critical pitfalls of analysis with this popular measurement. Zhou et al. (2022b) demonstrate that the cosine similarity measurement can underestimate the similarity between word embedding vectors compared to human judgments. They attribute this underestimation to differences in word frequency, suggesting that low-frequency and high-frequency words have different geometric representations in the embedding space. Steck, Ekanadham & Kallus (2024) also detects issues with cosine similarity measurement, showing that it can return arbitrary similarity values in certain settings using regularized linear models. Wannasuphoprasit, Zhou & Bollegala (2023) identify a problem where cosine similarity underestimates the actual similarity between high-frequency words as Zhou et al. (2022b). They propose a method to mitigate this problem that discounts the $\ell_2$ norm of a word embedding by its frequency to
correct this underestimation, demonstrating that the proposed method effectively resolves the issue.

**CONCLUSION**

This survey comprehensively overviewed the emerging interdisciplinary research trend. We survey the literature on the application of word embedding models to social science. Since this research trend covers a wide range of research areas, we first construct the taxonomy of the method used in the surveyed articles, constructing the labels of popular methods. We then discuss the literature with the constructed labels and this helps us with documenting and bridging the literature in different research areas from neuroscience to humanity. In addition, we discuss several important issues regarding the similarity measurement, which researcher should note in their application.

We have made several contributions to the literature and the science community. First, this survey detects the emerging trend of an application in social science that uses word embedding algorithms, and we document this trend not only by focusing on a single research area or discipline. This survey would serve “a hub” for researchers from various academic fields to refer to findings from other disciplines when using word embedding models for social science analysis. The second contribution is the introduction of the taxonomy of the methods used in the surveyed article. Based on this, we provide the labels that would help researchers classify their work based on their methods. Before this work, the literature has not particularly contextualized their work based on the way they use word embedding because of the lack of notions that bridge the literature. The present research also provides researchers with useful guidance and a well-organized framework for how to explore or classify these applications of word embedding models in social sciences. Lastly, we contribute to the literature by summarizing several important issues when applying word embedding to social science research. More specifically, the survey discusses some pitfalls of using popular similarity measurements. We would like to expect that this survey article harnesses the researcher working on this interdisciplinary area to explore and contextualize their studies, connecting their work across different research areas.

While the present article contributes to the science community in the aforementioned points, it has limitations as the other works do. First, we mainly focused on the application of the word2vec algorithm and did not deeply explore ones utilizing other embedding algorithms or neural-based approaches. For example, this survey did not cover the sentiment analysis or topic modeling by machine learning including the embedding-based method such as BART (Lewis et al., 2020). In addition, the presented work did not deeply discuss the methodological discussion about word embedding. There are a substantial amount of topics about the theoretical aspect of word embedding algorithms as the famous quote goes “Good question. We don’t really know.” in Goldberg & Levy (2014). The further direction for this literature could establish the theoretical foundation and empirical evidence that provides a framework or procedure for researchers who apply word embedding algorithms to social science research.
Funding
This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (No. JP22K20159); Research Institute of Science and Technology for Society, Japan, Grant Number JPMJRS23L4. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures
The following grant information was disclosed by the authors:
JSPS KAKENHI Grant-in-Aid for Scientific Research: JP22K20159.
Research Institute of Science and Technology for Society, Japan: JPMJRS23L4.

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Akira Matsui conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Emilio Ferrara conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability
The following information was supplied regarding data availability:
This is a literature review.

REFERENCES
Aceves P, Evans JA. 2024. Human languages with greater information density have higher communication speed but lower conversation breadth. Nature Human Behaviour 8(4):644–656 DOI 10.1038/s41562-024-01815-w.
Agirre E, Alfonseca E, Hall K, Kravalova J, Pašca M, Soroa A. 2009. A study on similarity and relatedness using distributional and WordNet-based approaches. In: Proceedings of Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics. 19–27.
Arora S, Li Y, Liang Y, Ma T, Risteski A. 2016. A latent variable model approach to PMI-based word embeddings. Transactions of the Association for Computational Linguistics 4:385–399 DOI 10.1162/tacl_a_00106.
Ash E, Chen DL, Ornaghi A. 2020. Stereotypes in high-stakes decisions: evidence from us circuit courts. SSRN Electronic Journal DOI 10.2139/ssrn.3749842.
Baltrušaitis T, Ahuja C, Morency LP. 2018. Multimodal machine learning: a survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence 41(2):423–443 DOI 10.1109/TPAMI.2018.2798607.
Barros CD, Mendonça MR, Vieira AB, Ziviani A. 2021. A survey on embedding dynamic graphs. ACM Computing Surveys 55(1):1–37 DOI 10.1145/3483595.
Baumgärtner M, Zahner J. 2021. Whatever it takes to understand a central banker: Embedding their words using neural networks. Technical report, MAGKS Joint Discussion Paper Series in Economics.

Benoit K. 2020. Text as data: an overview. In: The SAGE Handbook of Research Methods in Political Science and International Relations. London: SAGE Publishing.

Bhatia S. 2019. Predicting risk perception: new insights from data science. Management Science 65(8):3800–3823 DOI 10.1287/mnsc.2018.3121.

Bhatia N, Bhatia S. 2021. Changes in gender stereotypes over time: a computational analysis. Psychology of Women Quarterly 45(1):106–125 DOI 10.1177/0361684320977178.

Bhatia S, Olivola CY, Bhatia N, Ameen A. 2021. Predicting leadership perception with large-scale natural language data. The Leadership Quarterly 33:101535 DOI 10.1016/j.leaqua.2021.101535.

Bird S, Klein E, Loper E. 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5:135–146 DOI 10.1162/tacl_a_00051.

Bojanowski P, Grave E, Joulin A, Mikolov T. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in Neural Information Processing Systems 29:4349–4357.

Boutyline A, Arseniev-Koehler A, Cornell DJ. 2023. School, studying, and smarts: gender stereotypes and education across 80 years of American print media, 1930–2009. Social Forces 102(1):263–286 DOI 10.1093/sf/soac148.

Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in Neural Information Processing Systems 29:4349–4357.

Boycko-Jacino C, DeDeo S. 2020. Opacity, obscurity, and the geometry of question-asking. Cognition 196(2):104071 DOI 10.1016/j.cognition.2019.104071.

Cai H, Zheng VW, Chang KCC. 2018. A comprehensive survey of graph embedding: problems, techniques, and applications. IEEE Transactions on Knowledge and Data Engineering 30(9):1616–1637 DOI 10.1109/TKDE.2018.2807452.

Čech R, Hůla J, Kubát M, Chen X, Milička J. 2019. The development of context specificity of lemma. A word embeddings approach. Journal of Quantitative Linguistics 26(3):187–204 DOI 10.1080/09296174.2018.1491748.

Dahbura NN, Boyd-Graber JH, Quinlan ER. 2014. Deep learning for extracting word-level meaning from safety report narratives. In: 2014 International Conference on Machine Learning. PMLR, 803–811.

Deng J, Li J, Li F, Mei AM. 2017. Multimodal distributional semantics. Journal of Artificial Intelligence Research 49:7–102 DOI 10.1613/jair.4135.

Economides E, Stroup JF, Schriber AO, Chawla NV. 2014. Automatic sense disambiguation in text. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 1273–1283.

Eisenstein J, Kollnig T, Epstein E. 2014. Language is the key to hard news. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Prague: Association for Computational Linguistics, 1122–1132.

Eisenstein J, Stoyanov J, Kollnig T. 2015. Embedding models of language: an overview. In: Proceedings of the 14th Conference on Computational Natural Language Learning. Brown University: Association for Computational Linguistics, 1–10.

Eisenstein J, Shitrit S, Smith D, Stoyanov J. 2017. Future directions in the theory and practice of deep learning for NLP. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 1918–1927.

Eisenstein J, Shitrit S, Smith D, Stoyanov J. 2017. Future directions in the theory and practice of deep learning for NLP. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 1918–1927.

Ellis H, 2016. How can deep learning be used in natural language processing? In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 1606–1610.

Eisenstein J, Stoyanov J, Kollnig T. 2015. Embedding models of language: an overview. In: Proceedings of the 14th Conference on Computational Natural Language Learning. Brown University: Association for Computational Linguistics, 1–10.

Eisenstein J, Kollnig T, Stoyanov J, Epstein E. 2014. Case studies in the theory and practice of deep learning for NLP. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 1122–1132.

Eisenstein J, Stoyanov J, Kollnig T. 2015. Embedding models of language: an overview. In: Proceedings of the 14th Conference on Computational Natural Language Learning. Brown University: Association for Computational Linguistics, 1–10.

Eisenstein J, Stoyanov J, Kollnig T. 2015. Embedding models of language: an overview. In: Proceedings of the 14th Conference on Computational Natural Language Learning. Brown University: Association for Computational Linguistics, 1–10.

Eisenstein J, Stoyanov J, Kollnig T. 2015. Embedding models of language: an overview. In: Proceedings of the 14th Conference on Computational Natural Language Learning. Brown University: Association for Computational Linguistics, 1–10.

Eisenstein J, Stoyanov J, Kollnig T. 2015. Embedding models of language: an overview. In: Proceedings of the 14th Conference on Computational Natural Language Learning. Brown University: Association for Computational Linguistics, 1–10.
Chaubard F, Fang M, Genthial G, Mundra R, Socher R. 2019. CS224n: natural language processing with deep learning, lecture notes: part I. Available at https://web.stanford.edu/class/cs224n/readings/cs224n-2019-notes01-wordvecs1.pdf (accessed 1 March 2022).

Chen H, Zhang R. 2023. Identifying nonprofits by scaling mission and activity with word embedding. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations 34:39–51 DOI 10.1007/s11266-021-00399-7.

Chersoni E, Santus E, Huang CR, Lenci A. 2021. Decoding word embeddings with brain-based semantic features. Computational Linguistics 47(3):663–698 DOI 10.1162/COLI_a_00412.

Cong L, Liang T, Zhang X. 2018. Textual factors: a scalable, interpretable, and data-driven approach to analyzing unstructured information. SSRN Electronic Journal DOI 10.2139/ssrn.3307057.

Cui P, Wang X, Pei J, Zhu W. 2018. A survey on network embedding. IEEE Transactions on Knowledge and Data Engineering 31(5):833–852 DOI 10.1109/TKDE.2018.2849727.

Durrheim K, Schulz M, Mafunda M, Mazibuko S. 2023. Using word embeddings to investigate cultural biases. British Journal of Social Psychology 62:617–629 DOI 10.1111/bjso.12560.

Dyer C. 2014. Notes on noise contrastive estimation and negative sampling. ArXiv preprint DOI 10.48550/arXiv.1410.8251.

Font JE, Costa-Jussa MR. 2019. Equalizing gender biases in neural machine translation with word embeddings techniques. ArXiv preprint DOI 10.48550/arXiv.1901.03116.

Garg N, Schiebinger L, Jurafsky D, Zou J. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences of the United States of America 115(16):E3635–E3644 DOI 10.1073/pnas.1720347115.

Gennaro G, Ash E. 2022. Emotion and reason in political language. The Economic Journal 132:1037–1059 DOI 10.1093/jej/ueab104.

Gentzkow M, Levy O. 2014. Simple dynamic word embeddings for mapping perceptions in the public sphere. In: Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science. 94–99.

Gonen H, Goldberg Y. 2019. Lipstick on a pig: debiasing methods cover up systematic gender biases in word embeddings but do not remove them. ArXiv preprint DOI 10.48550/arXiv.1903.03862.

Google. 2013. Google news word embedding model. Available at https://code.google.com/archive/p/word2vec/.

Goyal P, Ferrara E. 2018. Graph embedding techniques, applications, and performance: a survey. Knowledge-Based Systems 151:78–94 DOI 10.1016/j.knosys.2018.03.022.

Graham J, Nosek BA, Haidt J, Iyer R, Koleva S, Ditto PH. 2011. Mapping the moral domain. Journal of Personality and Social Psychology 101(2):366 DOI 10.1037/a0021847.

Grand G, Blank IA, Pereira F, Fedorenko E. 2022. Semantic projection recovers rich human knowledge of multiple object features from word embeddings. Nature Human Behaviour 6(7):975–987 DOI 10.1038/s41562-022-01316-8.

Grimmer J, Stewart BM. 2013. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Political Analysis 21(3):267–297 DOI 10.1093/pan/mps028.
Gutmann M, Hyvärinen A. 2010. Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: Proceedings of the Conference on Artificial Intelligence and Statistics.

Haidt J, Graham J. 2007. When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. Social Justice Research 20(1):98–116 DOI 10.1007/s11211-007-0034-z.

Hamilton W, Leskovec J, Jurafsky D. 2016a. Diachronic word embeddings reveal statistical laws of semantic change. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics.

Hamilton W, Leskovec J, Juraňsky D. 2016b. Histwords: word embeddings for historical text.

Harris ZS. 1954. Distributional structure. WORD 10(2–3):146–162 DOI 10.1080/00437956.1954.11659520.

Hill F, Reichart R, Korhonen A. 2015. Simlex-99: evaluating semantic models with (genuine) similarity estimation. Computational Linguistics 41(4):665–695 DOI 10.1162/COLI_a_00237.

Hu S, He Z, Wu L, Yin L, Xu Y, Cui H. 2020. A framework for extracting urban functional regions based on multiprototype word embeddings using points-of-interest data. Computers, Environment and Urban Systems 80(4):101442 DOI 10.1016/j.compenvurbsys.2019.101442.

Hu K, Qi K, Yang S, Shen S, Cheng X, Wu H, Zheng J, McClure S, Yu T. 2018. Identifying the “Ghost City” of domain topics in a keyword semantic space combining citations. Scientometrics 114(3):1141–1157 DOI 10.1007/s11192-017-2604-7.

Huang E, Socher R, Manning C, Ng A. 2012. Improving word representations via global context and multiple word prototypes. In: Proceedings of Annual Meeting of the Association for Computational Linguistics, 873–882.

Hutto C, Gilbert E. 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media 8(1):216–225 DOI 10.1609/icwsm.v8i1.14550.

Iwamoto R, Kohita R, Wachi A. 2021. Polar embedding. In: Proceedings of the 25th Conference on Computational Natural Language Learning. Online: Association for Computational Linguistics, 470–480.

Jha M, Liu H, Manela A. 2021. Natural disaster effects on popular sentiment toward finance. Journal of Financial and Quantitative Analysis 56:1–35 DOI 10.1017/S0022109021000466.

Johnson DR, Kaufman JC, Baker BS, Patterson JD, Barbot B, Green AE, van Heli J, Kennedy E, Sullivan GF, Taylor CL, Ward T, Beaty RE. 2023. Divergent semantic integration (DSI): extracting creativity from narratives with distributional semantic modeling. Behavior Research Methods 55(7):3726–3759 DOI 10.3758/s13428-022-01986-2.

Jones J, Amin M, Kim J, Skiena S. 2020. Stereotypical gender associations in language have decreased over time. Sociological Science 7(3):1–35 DOI 10.15195/v7.a1.

Kawaguchi K, Kuroda T, Sato S. 2021. Merger analysis in the app economy: an empirical model of ad-sponsored media. In: Conference paper. TPRC48: The 48th Research Conference on Communication, Information and Internet Policy, Vol. 3746830.

Kelly M, Ghafoorian M, West RL, Reitter D. 2020. Indirect associations in learning semantic and syntactic lexical relationships. Journal of Memory and Language 115:104153 DOI 10.1016/j.jml.2020.104153.

Khaoja I, Mezzour G, Carley KM, Kassou I. 2019. Building a soft skill taxonomy from job openings. Social Network Analysis and Mining 9(1):43 DOI 10.1007/s13278-019-0583-9.
Kim K, Askin N, Evans JA. 2024. Disrupted routines anticipate musical exploration. *Proceedings of the National Academy of Sciences of the United States of America* **121**(6):e2306549121 DOI 10.1073/pnas.2306549121.

Kim Y, Shin H. 2017. Finding sentiment dimension in vector space of movie reviews: an unsupervised approach. *Journal of Cognitive Science* **18**(1):85–101 DOI 10.17791/jcs.2017.18.1.85.

Kim B, Yoo M, Park KC, Lee KR, Kim JH. 2021. A value of civic voices for smart city: a big data analysis of civic queries posed by Seoul citizens. *Cities* **108**(1):102941 DOI 10.1016/j.cities.2020.102941.

Kozłowski AC, Taddy M, Evans JA. 2019. The geometry of culture: analyzing the meanings of class through word embeddings. *American Sociological Review* **84**(5):905–949 DOI 10.1177/0003122419877135.

Kroon AC, Trilling D, Meer TGvd, Jonkman JG. 2020. Clouded reality: news representations of culturally close and distant ethnic outgroups. *Communications* **45**(s1):744–764 DOI 10.1515/commun-2019-2069.

Kroon AC, Trilling D, Raats T. 2021. Guilty by association: using word embeddings to measure ethnic stereotypes in news coverage. *Journalism & Mass Communication Quarterly* **98**(2):451–477 DOI 10.1177/1077699020932304.

Leavy S, Keane MT, Pine E. 2019. Patterns in language: text analysis of government reports on the Irish industrial school system with word embedding. *Digital Scholarship in the Humanities* **34**(Supplement 1):i110–i122 DOI 10.1093/llc/fqz012.

Lee R, Kim J. 2021. Developing a social index for measuring the public opinion regarding the attainment of sustainable development goals. *Social Indicators Research* **156**(1):201–221 DOI 10.1007/s11205-021-02643-5.

Levy O, Goldberg Y. 2014. Dependency-based word embeddings. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*.

Lewis M, Cahill A, Madnani N, Evans J. 2023. Local similarity and global variability characterize the semantic space of human languages. *Proceedings of the National Academy of Sciences of the United States of America* **120**(51):e230986120 DOI 10.1073/pnas.230986120.

Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky D, Chai J, Schluter N, Tetreault J, eds. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 7871–7880.

Li K, Mai F, Shen R, Yan X. 2021. Measuring corporate culture using machine learning. *The Review of Financial Studies* **34**(7):3265–3315 DOI 10.1093/rfs/hhaa079.

Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikäinen M. 2020. Deep learning for generic object detection: a survey. *International Journal of Computer Vision* **128**(2):261–318 DOI 10.1007/s11263-019-01247-4.

Luong MT, Socher R, Manning CD. 2013. Better word representations with recursive neural networks for morphology. In: *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*. 104–113.

Matsui A. 2022. Mining and Modeling Temporal Structures of Human Behavior in Digital Platforms. Phd thesis, University of Southern California, Los Angeles, CA.

Matsui A, Ferrara E. 2022. Word embedding for social sciences: an interdisciplinary survey. ArXiv preprint DOI 10.48550/arXiv.2207.03086.
Matsui A, Ren X, Ferrara E. 2021. Using word embedding to reveal monetary policy explanation changes. In: Proceedings of the Third Workshop on Economics and Natural Language Processing. 56–61.

Mikolov T, Chen K, Corrado G, Dean J. 2013a. Efficient estimation of word representations in vector space. In: International Conference on Learning Representations.

Mikolov T, Chen K, Corrado GS, Dean J. 2015. US Patent 9, 037, 464. Computing numeric representations of words in a high-dimensional space.

Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. 2013b. Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems.

Mikolov T, Yih Wt, Zweig G. 2013. Linguistic regularities in continuous space word representations. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 746–751.

Miller GA, Charles WG. 1991. Contextual correlates of semantic similarity. Language and Cognitive Processes 6(1):1–28 DOI 10.1080/01690969108406936.

Nanni A, Fallin M. 2021. Earth, wind, (water), and fire: measuring epistemic boundaries in climate change research. Poetics 88(4):101573 DOI 10.1016/j.poetic.2021.101573.

Nelson LK. 2021. Leveraging the alignment between machine learning and intersectionality: using word embeddings to measure intersectional experiences of the nineteenth century U.S. South. Poetics 88(2):101539 DOI 10.1016/j.poetic.2021.101539.

Nyarko J, Sanga S. 2020. A statistical test for legal interpretation: theory and applications. SSRN Electronic Journal 1(6):86 DOI 10.2139/ssrn.3737292.

Pennington J, Socher R, Manning CD. 2014. Glove: global vectors for word representation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 1532–1543.

Peterson JC, Chen D, Griffiths TL. 2020. Parallelograms revisited: exploring the limitations of vector space models for simple analogies. Cognition 205(1):104440 DOI 10.1016/j.cognition.2020.104440.

Radinsky K, Agichtein E, Gabrilovich E, Markovitch S. 2011. A word at a time: computing word relatedness using temporal semantic analysis. In: Proceedings of the International Conference on World Wide Web. 337–346.

Rahimikia E, Zohren S, Poon SH. 2021. Realised volatility forecasting: machine learning via financial word embedding. SSRN Electronic Journal 24(3):1313 DOI 10.2139/ssrn.3895272.

Rehurek R, Sojka P. 2010a. Gensim. Available at https://github.com/RaRe-Technologies/gensim.

Rehurek R, Sojka P. 2010b. Software framework for topic modelling with large corpora. In: Proceedings of the LREC, 2010 Workshop on New Challenges for NLP Frameworks, Citeseer.
Rheault L, Cochrane C. 2020. Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis* 28(1):112–133 DOI 10.1017/pan.2019.26.

Richie R, Zou W, Bhatia S, Vazire S. 2019. Predicting high-level human judgment across diverse behavioral domains. *Collabra: Psychology* 5(1) DOI 10.1525/collabra.282.

Rodman E. 2020. A timely intervention: tracking the changing meanings of political concepts with word vectors. *Political Analysis* 28(1):87–111 DOI 10.1017/pan.2019.23.

Rong X. 2014. word2vec parameter learning explained. ArXiv preprint DOI 10.48550/arXiv.1411.2738.

Rozado D, al Gharbi M. 2021. Using word embeddings to probe sentiment associations of politically loaded terms in news and opinion articles from news media outlets. *Journal of Computational Social Science* 5:427–448 DOI 10.1007/s42001-021-00130-y.

Rubenstein H, Goodenough JB. 1965. Contextual correlates of synonymy. *Communications of the ACM* 8(10):627–633 DOI 10.1145/365628.365657.

Soni S, Lerman K, Eisenstein J. 2021. Follow the leader: documents on the leading edge of semantic change get more citations. *Journal of the Association for Information Science and Technology* 72(4):478–492 DOI 10.1002/asi.24421.

Steck H, Ekanadham C, Kallus N. 2024. Is cosine-similarity of embeddings really about similarity? In: *Companion Proceedings of the ACM on Web Conference 2024, WWW ’24*. Stroudsburg: Association for Computing Machinery, 887–890.

Tausczik YR, Pennebaker JW. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1):24–54 DOI 10.1177/0261927X09351676.

Taylor SD. 2021. Concepts as a working hypothesis. *Philosophical Psychology* 35(1):569–594 DOI 10.1080/09515089.2021.2014439.

Tifrea A, Bécigneul G, Ganea OE. 2018. Poincar’e glove: hyperbolic word embeddings. In: *Proceeding of International Conference on Learning Representations*.

Toubia O, Berger J, Eliashberg J. 2021. How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences* 118(26):e2011695118 DOI 10.1073/pnas.2011695118.

Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, Persson KA, Ceder G, Jain A. 2019. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* 571(7763):95–98 DOI 10.1038/s41586-019-1335-8.

Vendrov I, Kiros R, Fidler S, Urtasun R. 2016. Order-embeddings of images and language. In: *Proceeding of International Conference on Learning Representations*.

Vilnis L, McCallum A. 2015. Word representations via gaussian embedding. In: *Proceeding of International Conference on Learning Representations*.

Waller I, Anderson A. 2021. Quantifying social organization and political polarization in online platforms. *Nature* 600(7888):264–268 DOI 10.1038/s41586-021-04167-x.

Wang W. 2019. Two-stage learning of a systemic risk factor. *SSRN Electronic Journal* DOI 10.2139/ssrn.3017446.

Wang B, Wang A, Chen F, Wang Y, Kuo CCJ. 2019. Evaluating word embedding models: methods and experimental results. *APSIPA Transactions on Signal and Information Processing* 8(1):313 DOI 10.1017/ATSI.P.2019.12.

Wannasuphoprasit S, Zhou Y, Bollegala D. 2023. Solving cosine similarity underestimation between high frequency words by $\ell_2$ norm discounting. In: Rogers A, Boyd-Graber J, Okazaki N,
Wevers M, Koolen M. 2020. Digital begriffsgeschichte: tracing semantic change using word embeddings. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53(4):226–243 DOI 10.1080/01615440.2020.1760157.

Yao Z, Sun Y, Ding W, Rao N, Xiong H. 2018. Dynamic word embeddings for evolving semantic discovery. In: *Proceedings of the Proceedings of the International Conference on Web Search and Data Mining*, 673–681.

Zhao J, Zhou Y, Li Z, Wang W, Chang KW. 2018. Learning gender-neutral word embeddings. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 4847–4853.

Zheng C. 2020. Comparisons of the city brand influence of global cities: word-embedding based semantic mining and clustering analysis on the big data of GDELT global news knowledge graph. *Sustainability* 12(16):6294 DOI 10.3390/su12166294.

Zhou K, Ethayarajh K, Card D, Jurafsky D. 2022b. Problems with cosine as a measure of embedding similarity for high frequency words. In: Muresan S, Nakov P, Villavicencio A, eds. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Stroudsburg: Association for Computational Linguistics, 401–423.

Zhou X, Huang S, Zheng Z. 2020. RPD: a distance function between word embeddings. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. Stroudsburg: Association for Computational Linguistics, 42–50.

Zhou J, Liu L, Wei W, Fan J. 2022a. Network representation learning: from preprocessing, feature extraction to node embedding. *ACM Computing Surveys* 55(2):1–35 DOI 10.1145/3491206.

Zipf GK. 1946. The P1 P2D hypothesis: on the intercity movement of persons. *American Sociological Review* 11(6):677–686 DOI 10.2307/2087063.