Evaluating the timing and magnitude of semantic change in diachronic word embedding models

Anonymous ACL submission

Abstract

Recent studies have suggested that diachronic word embedding models are able to track the direction of changes in public perception. Building on these works, we evaluate the ability of diachronic word embedding models to accurately capture such changes both qualitatively and quantitatively, such as their timing and magnitudes. Using a longitudinal dataset on public perception of brands, we found that evolution of word meaning as captured by diachronic word embedding models, trained on New York Times articles, reflected the timing and magnitudes of general consumer awareness of companies. In contrast, this was not the case for other readily available characteristics, such as stock market prices. This comparison is enabled by a new feature extraction method which summarizes the semantic changes encoded in diachronic word embeddings.

1 Introduction

Recently there has been a surge of interest in using diachronic word embedding models to understand historical evolution of human perception. Garg et al., 2018 provides an early example of this effort in quantifying the change of gender and ethnic stereotype over the past century. Other works include inferences about the evolution of a range of psychological and cultural constructs of a subjective nature, including moral sentiment (Xie et al., 2019), antisemitism (Tripodi et al., 2019), and perception of social class (Kozlowski et al., 2019).

These pioneering studies demonstrated the value of diachronic word embedding models primarily in capturing the direction of changes. For instance, Garg et al., 2018 has shown that text-based gender bias measure has been decreasing, consistent with the real-world trend.

A natural next question to ask is whether diachronic word embedding models can accurately capture timing and magnitude of temporal changes, above and beyond direction of changes. For example, can diachronic word embedding models accurately pinpoint salient time series markers, such as when a term comes closest to a particular meaning, or periods of particularly rapid change? Ability to answer these questions can greatly increase the practical value of these models.

In particular, we aim to overcome two challenges associated with understanding the timing and magnitudes of semantic changes. Firstly, it remains difficult to conduct temporal analysis on semantic changes in diachronic word embedding models. For example, it is easy to know that the word "trump" has changed in association from "real estate" to "president". We only need to look at its most similar words in two time points. However, to know the timing and magnitudes of the change, we need a feature extraction method to compute the semantic vector of the change.

Secondly, the lack of high-resolution longitudinal datasets to evaluate findings about subjective perception has been well recognized (Kutuzov et al., 2018). This prevents researchers from characterizing the relationship between the text-based measures and public perception. Are texts leading, lagging, or synchronized with the construct of interest? How sensitive are diachronic word embedding models to changes in public perception? The application value of diachronic word embedding models critically depends on answers to these questions.

In this study, we seek to address the two challenges by (1) proposing a feature extraction method for diachronic word embedding models which summarizes the semantic changes of words, and (2) introducing a longitudinal dataset measuring public perception of consumer brands. With the dataset, we are able to elucidate the relationship between diachronic word embedding models and public perception of named entities. To the best of our knowledge, this is the first paper which directly compares diachronic word embedding results to time series of human cognitive metric.

2 Related Work

A growing volume of work in natural language processing has developed various approaches to track lexical semantic changes across time (Bamler and Mandt, 2017; Kutuzov et al., 2018). A majority of these approaches build on distributional semantic models (Hamilton et al., 2016; Yao et al., 2018), and have been shown to deliver excellent performance in tasks such as detecting changes in word meanings (Hamilton et al., 2016), identifying temporal analogies (Szymanski, 2017), and characterizing the trajectories of semantic changes (Yao et al., 2018; Li et al., 2019).

Because of the deep theoretical connections between word meanings and semantic memory, i.e., people's shared knowledge about concepts and facts of the world (Kumar, 2021), computational linguists and cognitive scientists have long held the belief that distributional semantic models are more than a method for word meaning analysis, but also a lens into how people perceive the world (McRae and Jones, 2013; Lenci, 2018; Günther et al., 2019).

The majority of recent studies that examine this general hypothesis measure keywords along a predetermined dimension such as gender, ethnicity, morality, etc (Garg et al., 2018; Xie et al., 2019). While such an approach has proven successful in measuring semantic changes on a specific, predefined demension, it is less suitable for characterizing semantic changes in general, since it requires researchers to handpick the representative words for the change. For instance, to summarize the semantic change of the word "facebook" during the past 20 years, we need to assume that "facebook" was mainly moving in the direction of social media (not technology or communication), and pick the correct representative words for the direction ("whatsapp" and "instagram" are good choices, whereas "social platform" is less suitable since it is not used as frequently in the news). Therefore, we need an automatic feature extraction method to compute the direction of major semantic changes.

In terms of validation, the focus on what were in the minds of human beings living in the past poses significant challenges for validating that metrics derived from diachronic word embedding models indeed reflect people's subjective thoughts across time. Here, an evaluation dataset collected from a contemporary sample of human subjects is not sufficient, and a longitudinal dataset with the suitable time span and resolution is required, which in most cases does not exist. Existing studies typically sidestep this difficulty by using one of two strategies: (1) performing validation analysis only for selected time periods (often the contemporary period) for which data can be collected or are available (Xie et al., 2019; Kozlowski et al., 2019), or (2) relying on other longitudinal real-world data that are related to, but do not directly measure, the construct of interest (Garg et al., 2018).

Therefore, it remains unclear how well diachronic word embedding models may represent subjective thoughts and perception shared by the public in a more precise manner, beyond showing general consistency in the direction of change. Aiming to fill this gap in existing research, our work proposes a feature extraction method for summarizing semantic changes, and leverages a rare opportunity from a long-running market research dataset that measures the changing public perception of a large number of consumer brands, allowing for direct, formal validation of inferences from diachronic word embedding.

3 Data Description

3.1 Brand Asset Valuator (BAV)

Due to their importance to the economy and firms, there is rich availability of data on the public perception of consumer brands from market research firms, academic researchers, or governments. In this paper, we chose to focus on the Brand Asset Valuator (BAV) dataset which covers a period from 1993 to the present. This dataset, published by the BAV Group, a firm that specializes in measuring public perception of consumer brands, is particularly useful for our purposes as it has been continuously collected since 1993, and the collection methodology has been stable except for a change in 1996. For these reasons, the BAV dataset has been widely used by both practitioners and academic marketing researchers to track changes in various aspects of brand perception (Mizik and Jacobson, 2008, Stahl et al., 2012, Datta et al., 2017). Specifically, the BAV dataset measures brand awareness on a scale of 0 to 100, which can be further decomposed into four components. We use the brand awareness data from 1996-2019 in this study.

3.2 Other Metrics

In addition to the BAV, we investigated other potential evaluation datasets reflecting tangible real-world outcomes. This includes historical stock prices of Blackberry (GoogleFinance, 2022) and the number of monthly active users of Facebook from 2008 to 2018 (Statista, 2021).

4 Method

4.1 Diachronic Word Embedding Models

We evaluated two diachronic word embedding models in this study: the Orthogonal Procrustes Alignment model proposed by Hamilton et al., 2016, and the Dynamic Word Embedding (DWE) model proposed by Yao et al., 2018. The former is a two-step procedure which aligns the single-year word vectors, whereas the latter jointly optimizes word vectors and alignment during model training. All of our results hold for both models.

4.1.1 Orthogonal Procrustes Alignment

We first trained a single-year GloVe model (Pennington et al., 2014) on New York Times articles for each year from 1996 to 2019 [1]. We then aligned the word vectors in each year sequentially to the vectors in the previous year by finding the orthogonal rotation matrix which minimizes the differences between the vectors.

4.1.2 Dynamic Word Embedding (DWE)

We trained a 50-dimensional DWE model on all New York Times articles from 1996 to 2019 using the codes provided by the authors. All parameters are set using the default values in the codes.

4.2 Semantic Trajectory

To plot the semantic trajectory for a given target word, we first take the union of the two most similar words to the target word in each year. These words serve as landmarks to help interpret the trajectory. Then we project the word vectors of the target word in all years along with the word vectors of the landmark words in 2019 onto the two-dimensional plane using Multidimensional Scaling (MDS). MDS is chosen since it respects the cosine similarities between pairs of word vectors.

4.3 Semantic Change Feature Extraction

For a given brand we applied Principal Component Analysis (PCA) to all of their word vectors from 1996 to 2019. The first principal component is then the direction of the largest semantic movement of that brand during the 24 years. We then computed the cosine similarity of the brand word vectors in each year with the first principal component. This entire procedure summarizes the semantic change of the brand across time by projecting its high-dimensional semantic trajectory along the direction of the largest semantic movement.

[1] The data is available at this link.

5 Evaluation

5.1 Timing and Magnitudes of Changes

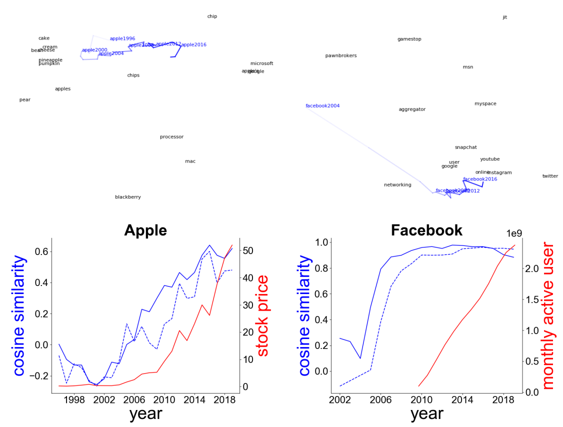

Figure 1: Top panels: Semantic trajectories of "apple" and "facebook" from 1996 to 2019 using DWE. Words in black are in their 2019 meanings. Bottom panels: Projected semantic trajectories vs. stock prices (Apple) and monthly active users (Facebook). Solid blue time series are DWE trajectories. Dashed blue time series are Orthogonal Procrustes Alignment trajectories.

Evaluation Component	Previous Works
Direction	Xie et al., 2019 Kozlowski et al., 2019 Garg et al., 2018 Hamilton et al., 2016
Timing and magnitudes	?

Table 1: Previous works which evaluate different components of semantic changes in diachronic word embedding models.

In this section, we demonstrate the importance of examining the timing and magnitudes components of semantic changes. The top panels in Figure 1 shows the semantic trajectories of the word "facebook" and "apple". Consistent with previous findings (Yao et al., 2018), the trajectories correctly illustrate the direction of semantic changes for both words: "apple" moved from fruit to technology, and "facebook" moved steadily toward the social media direction.

Upon seeing the promising results of semantic direction, one might ask further questions about the timing and magnitudes of the semantic trends, which has been absent from previous studies (Table 1). For instance, can the semantic trajectories serve as indicators of objective metrics such as monthly active users of Facebook or stock prices of Blackberry? We examined the possibility by comparing the projected semantic trajectories with the time series of both metrics. However, as shown in the bottom panels of Figure 1, this is not the case. The linear growth of Facebook is not captured by its projected semantic trajectory, and there is a significant dip in Apple's semantic trajectory which is not reflected in its stock price. Therefore, examination of timing and magnitudes of changes reveals information about the model's range of application. Note that our point is not that diachronic word embedding models cannot be applied to predict objective metrics, but that this possibility has to be empirically tested by examining the timing and magnitudes of semantic changes.

5.2 Evaluation Using BAV Time Series

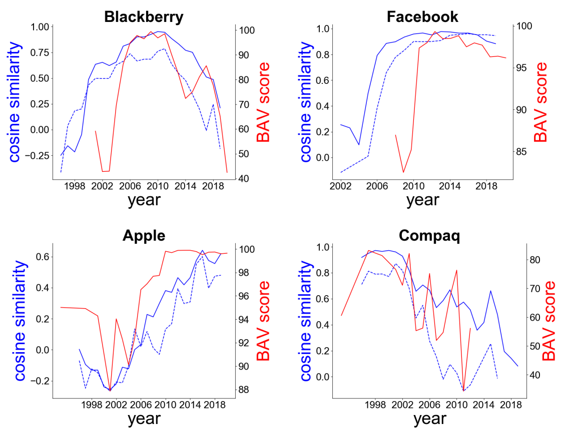

Figure 2: BAV brand awareness scores and projected semantic trajectories of Blackberry, Facebook, Apple, and Compaq. Solid blue time series are DWE trajectories. Dashed blue time series are Orthogonal Procrustes Alignment trajectories.

If diachronic word embedding does not capture changes in objective outcomes, what is it capturing? We answer this question using the BAV brand awareness data. As shown in Figure 2, brand awareness is well tracked by both diachronic word embedding models, with qualitatively similar tracking performance as Garg et al., 2018. The models correctly identify the timing of peaks (Blackberry) and valleys (Apple), and differentiate between different patterns of changes in brand perception (Facebook rose sharply, whereas Compaq declined steadily). The results suggest that diachronic word embedding models better capture subjective perception as opposed to objective metrics.

To the best of our knowledge, this is the first paper which directly compares diachronic word embedding results to time series of human cognitive metric. Garg et al., 2018 compared text-based bias metrics to historical occupation participation data of minorities. Xie et al., 2019 evaluated their model against moral judgment data at only one contemporary time point. Kozlowski et al., 2019 verified that social class associations can be inferred from semantic associations against one contemporary and one historical time points. Our analysis clearly illustrates the strength of diachronic word embedding models in tracking subjective perception.

6 Conclusion

Timing and magnitudes are important components to consider when evaluating the results of diachronic word embedding models. In this paper, we show that it is possible for diachronic word embedding models to capture such information above and beyond general directions of change. Intriguingly, we show that trajectories of companies captured in our diachronic word embedding models of the New York Times reflect general consumer awareness of these companies, rather than more "objective" characteristics of these companies, such as stock market prices.

One general limitation of using historical time series data to evaluate performances of diachronic word embedding models is the difficulty in providing quantitative metrics of accuracy. Naive application of linear regression does not work here since most real-world time series are highly autocorrelated, which could lead to the notorious "spurious correlation" issue (Granger and Newbold, 1974). This is the reason why many studies employ only qualitative evaluations when comparing time series data (Garg et al., 2018, Yao et al., 2018, Michel et al., 2011), including ours. It is therefore desirable to have quantitative methods for comparing time series data in the context of diachronic word embeddings and natural language processing in general.

References

Robert Bamler and Stephan Mandt. 2017. Dynamic word embeddings. In *International Conference on Machine learning*, pages 380–389. PMLR.

Hannes Datta, Kusum L Ailawadi, and Harald J Van Heerde. 2017. How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response? *Journal of Marketing*, 81(3):1–20.

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.

GoogleFinance. 2022. Blackberry ltd. https://www.google.com/finance/quote/BB:NYSE.

Clive WJ Granger and Paul Newbold. 1974. Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120.

Fritz Günther, Luca Rinaldi, and Marco Marelli. 2019. Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, 14(6):1006–1033.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Austin C Kozlowski, Matt Taddy, and James A Evans. 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5):905–949.

Abhilasha A Kumar. 2021. Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, 28(1):40–80.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. *arXiv preprint arXiv:1806.03537*.

Alessandro Lenci. 2018. Distributional models of word meaning. *Annual Review of Linguistics*, 4:151–171.

Ying Li, Tomas Engelthaler, Cynthia SQ Siew, and Thomas T Hills. 2019. The macroscope: A tool for examining the historical structure of language. *Behavior Research Methods*, 51(4):1864–1877.

Ken McRae and Michael Jones. 2013. Semantic memory. *The Oxford Handbook of Cognitive Psychology*, 206.

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Google Books Team, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

Natalie Mizik and Robert Jacobson. 2008. The financial value impact of perceptual brand attributes. *Journal of Marketing Research*, 45(1):15–32.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Florian Stahl, Mark Heitmann, Donald R Lehmann, and Scott A Neslin. 2012. The impact of brand equity on customer acquisition, retention, and profit margin. *Journal of Marketing*, 76(4):44–63.

Statista. 2021. Number of monthly active facebook users worldwide as of 2nd quarter 2021. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

Terrence Szymanski. 2017. Temporal word analogies: Identifying lexical replacement with diachronic word embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (volume 2: short papers)*, pages 448–453.

Rocco Tripodi, Massimo Warglien, Simon Levis Sullam, and Deborah Paci. 2019. Tracing antisemitic language through diachronic embedding projections: France 1789-1914. *arXiv preprint arXiv:1906.01440*.

Jing Yi Xie, Renato Ferreira Pinto Junior, Graeme Hirst, and Yang Xu. 2019. Text-based inference of moral sentiment change. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4654–4663.

Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. 2018. Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, pages 673–681.