

Section Draft

Max Feng

Introduction

Linguistic systems are not always static. Although they are often well-balanced, language still evolves sometimes due to newly developed objects or ideas (Blank, 1999). In the digital era, certain words including bug, web, mouse, and cloud come to develop new meanings to accommodate the invention of new technological devices. This process, known as semantic change, is one of the central topics of historical linguistics.

Semantic change has been investigated historically through qualitative analyses of texts over time (Blank, 1999). However, this tradition method is time-consuming and can be subjective. In the modern era, the vast storage capability of computers allows us to possess an unprecedented amount of data presented in media, which makes the quantification of semantic change possible.

We employ diachronic word embeddings—a technique that represents words as high-dimensional vectors—to create a quantitative “map” of meaning over time (Hamilton et al., 2016). By training separate word embedding models on technology news articles from each year between 2013 and 2023, we create a temporal series of semantic snapshots. The core of our analysis involves statistically aligning these snapshots and measuring the **cosine distance** between a word’s vector representation across different years. This distance serves as a numerical index of semantic change, allowing us to identify which technology-related words have shifted most significantly over the past decade.