

Detection of Change in the Senses of AI in Popular Discourse

Ahmet Suerdem, Tugba Yıldız, Savaş Yıldırım

Istanbul Bilgi University
Santral Istanbul, Eyup, Istanbul
{ahmet.suerdem,
tdalyan,
savasy}@bilgi.edu.tr
<http://www.bilgi.edu.tr>

Abstract. As the chatbots, driverless cars and other robot like applications become a part of everyday life, we are witnessing an increase in the popularization of Artificial Intelligence (AI) by the mass media. While this has some potential in terms of informing the public about a technological development, it also makes the term a buzzword not pointing to any actual object with no agreed upon meaning. AI is usually deployed as an umbrella term for sealing a variety of analytical tools such as intelligent decision support systems, deep learning and computational linguistics disregarding their actual denotations. As the popular discourse and media represent its mundane features to connote miracles or apocalypses, AI gains a mythical status which can have different significations according to different cultural contexts. Our aim in this paper is to study the semantic shifts in the meaning of AI in different contexts by examining the mapping of the words to different semantic vector spaces over time.

Keywords: Word embeddings, semantic shifts, artificial intelligence

1 Introduction

Recently, the detection of semantic shifts in the meaning of words has gained considerable attention in the fields of information retrieval and computational linguistics. Semantic similarity can be measured according to a distributional model postulating that terms sharing similar contexts are semantically similar. First order representations represent a word in a one-hot long vector in a word-by-word matrix in terms of co-occurrence statistics measuring the semantic similarity. On the other hand, recent advances in the literature suggest that second-order representations overtake the former [1–3].

In that respect, Neural Network Language Models (NNLM) have demonstrated promising performance by reducing time complexity especially for word representation learning. The most important characteristics of NNLM is their capacity to generate dense and short word embeddings that are highly effective for finding semantic and syntactic regularities [4, 5].

In this paper, our primary goal is to detect the change in the meaning of “Artificial Intelligence” (AI) over time by using word embeddings. The corpus is collected from media articles taken from major UK newspapers by scraping the Lexis/Nexis¹ queries for the keyword “artificial intelligence”. Once learned word representation, semantic similarity could be easily measured by simple metrics in a static model. Detection of semantic shift requires dynamic analysis allowing us to track and detect the changes in the meaning of AI across time. Hence we measured semantic shifts in the sense of AI across last five years by means of word embedding vectors. Applying a word2vec model to build word vectors, we tracked the changes according to both global and local word embeddings models. Finally, to show the shifts in the semantic of AI, we present our results as time series patterns.

1.1 Related Works

In recent years, there have been variety of computational studies on the language changes over time [6–9]. [6] proposed distributional similarity approach to a relative-frequency-based method using the Google Books Ngram data from the 1960s and 1990s. In [7], they proposed three methods based on frequency to extract sudden change in word usage, syntactic times series over part of speech tag distribution and distributional times series over embedding space by using three different datasets, The Google Books Ngram Corpus, Amazon Movie Reviews and Twitter data. [8] proposed a method to monitor of vocabulary shifts over time proceeds as follows : using distributional semantic models to infer semantic spaces over time from a time-stamped textual documents, constructing semantic networks by applying graph-based measures to calculate saliency of terms, and shifting the vocabularies over time.

[9] showed that using a linear transformation is effective to find semantic shifts over time and how distributional methods can reveal the two statistical laws (law of conformity and law of innovation) of semantic change. [10] monitored semantic fluctuations over more than 400 years using time-stamped word representations per decade. They also proposed a visual analytics framework for visualizing lexical change at three different levels - individual words, word pairs, and sentiment orientation. Similar approach is proposed to compute the semantic shifts using word embeddings trained on corpora that represent specific viewpoints and evaluated on political speeches and media reports [11].

2 Word Representation

Distributional approaches represent words in vector space models (VSM) for the NLP problems. For example, [12] represented the sense of a word as a real-valued vector by using co-occurrence statistics to measure the semantic similarity. It is based on the idea that if two words share similar neighboring words, they are

¹ <https://www.lexisnexis.com/hottopics/scholastic/>

likely to be similar. The similarity between the vectors of the words are simply computed by cosine similarity and other metrics. The main disadvantages of this method is the size and sparsity of the matrix that is equal to the size of the vocabulary. As the dimension of the vector exceedingly increases, so does computational complexity of the designed system.

The widely applied solution is the feature elimination in the preparation step. It discards non-informative terms based on some metrics using corpus statistics. The study [1] pointed that the term frequency could be informative. Some feature selection criteria such as chi-square (χ^2) are found very effective to find informative terms from corpus, [1–3]. Another technique is to reduce the dimensionality such as in Latent Semantic Indexing [13] (or Latent Semantic Analysis). This technique is applied to produce informative and short latent dimensions. It uses Singular Value Decomposition (SVD) as a method for building significant dimensions derived from a document-term matrix. It is a member of a method family that can approximate an N-dimensional matrix using fewer dimensions such as Principle Components Analysis (PCA), Factor Analysis, etc [14–16].

Besides these dimension reduction techniques, NNLM have recently become widely used and demonstrated promising performance by reducing time complexity. The most important characteristics of NNLM is its capacity of generating dense and short embeddings, namely word embeddings [4, 5]. In the neural networks architecture, each word is initially associated with a random vector. As a two-layer neural network processes textual corpus, the vectors are iteratively updated by applying stochastic gradient descent (SGD) where the gradient is measured by back-propagation. The objective is to guess the last word from a given word sequence. Thus, the prediction task is typically similar to multi-class classification where soft-max function is used to compute class probability estimation. The network finally learns the embeddings for all words appeared in the corpus by convergence.

As one of the most popular word embeddings models, word2vec model [4] showed how word embeddings were efficiently trained within two different architectures, namely Continuous Bag of Words (CBow) and the Skip-Gram (SG). The architecture achieved both minimizing computational time complexity and maximizing model accuracy. Second model, GloVe, [5] proposed another word embedding model. It is based on matrix factorization and a new global log-bilinear regression model. These two popular word embedding models also proved that embeddings are very good at capturing syntactic and semantic regularities, using the vector offsets between word pairs.

3 Methodology

We propose a model that measures the change in the sense of AI through time. For building time series we have produced semantic similarities between AI and other words over time.

Our procedure are as follows:

Algorithm

```

Building Time Series(C, years):
    let C be a global corpus
    pre-processing(C)
    applyNPChunker(C)
    let C_year be a local corpus for each year
    globalModel= WordEmbeddings(C)
    V= buildVoc(C, globalModel)

    # Build Local Word Embedding for each year
    for each year in(2013,2017):
        localModel[year]= WordEmbeddings(C_year)

    # semantic similarities of terms w with AI over time
    TimeSeries=[]
    for each year in (2013, 2019):
        model=localModel[year]
        for w in V:
            TimeSeries.append((w,year), model.SemSim("AI",w))

    # Clustering Time Series
    cluster=HierCluster(Normalize(TimeSeries))

```

In order to measure semantic shift of a word, a representative time slot corpora is needed [7]. We have decided to start from 2013, a date when the term AI has started to creep into popular press because of Siri, Google now and Cortana and especially their application to smartphones using natural language to answer questions, make recommendations and perform actions. We have collected the corpus by means of automatically scraping Lexis/Nexis queries for retrieving the news articles in major UK newspapers containing the keyword “artificial intelligence” between 2013 and 2017. We balanced the corpus by randomly selecting an equal number of articles.

After collecting the corpus, some pre-processing steps such as: cleaning noisy terms, tokenization, sentence boundary detection, stop words removal have been applied. For entity detection, we segmented and annotated multi-word sequences by means of noun-phrase chunking. Collocation sets can be created according to some metrics depending on corpus overall statistics and n-gram statistics which in turn be used as a chunker where we used bigrams and tri-grams. With the ngram chunker, we captured the noun phrases such as Artificial Intelligence, Big Data and lemmatized them. Besides using the entire (i.e. global) corpus to measure such statistics, we have also used the local (i.e. annual subset) corpora to compute local statistics. First, n-gram chunkers were trained through global corpus statistics and then they were applied to each local corpus.

After preprocessing the corpus, we have trained word embeddings model using both global corpus and other five local corpora. For each subset corpus, a separate word embeddings model was built to measure the semantic and other

differences between the terms across time. To train word embedding model, we used word2vec model. The preprocessed and annotated textual data were consumed by word2vec model with mostly default configuration where dimension size is 300, minimum word frequent threshold is 5 and the context window is 10. We divided our article corpus into five temporal subsets S year. We create a vocabulary V by selecting those terms that appear in each S year, local corpus, and are similar to the term AI. The terms whose global corpus frequency is less than 50 are eliminated. Finally, we constructed a time series data frame for word semantics and usage. We applied time series clustering technique to a dynamic data for tracking the change where the terms are grouped in term of their characteristics such losing similarities or gaining similarities with AI. We present our findings as plots visualizing the time series trends in similarities.

4 Experiments

4.1 Preliminary Analysis

The pre-processed corpus is summarized in Fig.1. The terms are sorted according to their frequencies and plotted in the histogram. A quick examination of the ten most frequent words shows that all these terms (i.e. machine learning, big data and virtual reality) denote to “artificial intelligence” as a generic concept. These two figures verify the Zipfs law indicating that the frequency of any word should be inversely proportional to its rank in the table of word frequency. Thus, the most frequent word in corpus occurs roughly twice as often as the second most frequent word, and so forth. The corpus does not show any idiosyncrasies.

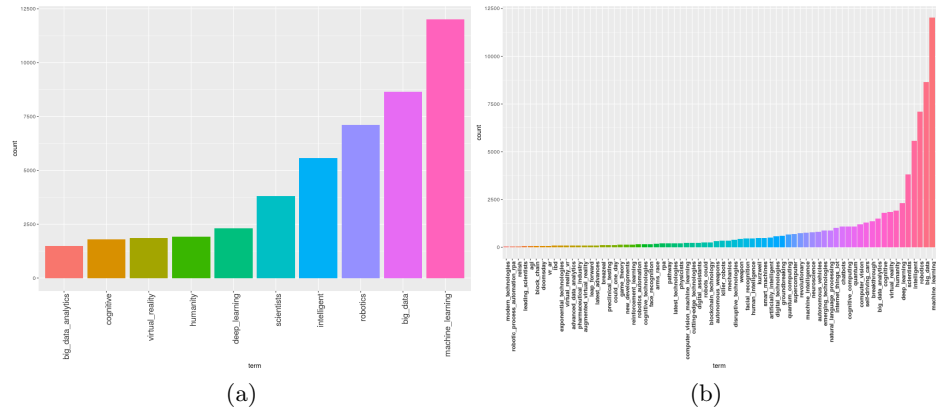


Fig. 1: A set of two subfigures describes: (a) Histogram of first ten terms; (b) Histogram of all terms

4.2 Word Usage Analysis

After examining the corpus, we evaluated the change in word usage by creating a year by word count matrix where each cell represents how many times a word appears in the corresponding year. To simplify the word space and represent it on a two-dimensional space, we used Correspondence Analysis. Correspondence Analysis (CA) is a multivariate dimensionality reduction technique designed to explore relationships among categorical variables and jointly represent the patterns in their categories [18]. This technique is especially effective for cross-tabulated data and is widely used across many disciplines such as social sciences, history and psychology because of its ease of understanding for the non-technical audience. The following Fig.2 shows the Correspondence Plot where columns (years) and the rows (words) are jointly mapped into a 2D space. The positioning of words and years along the coordinates of this space represents Euclidean distances and nicely summarizes the groupings of years and words in terms of their semantic proximities. The circles in the figure show the marginal frequency of the word concerned. Greater circles represent more frequent words and since we have balanced the distribution of news articles by selecting an equal number of articles per year, the sizes of the circles are comparable.

When we examine the plot, the years 2013 and 2014 are fairly close to each other and words signifying more generic AI such as “big data” and “computers” are grouped around them. This suggests that the sense of AI is not yet much differentiated from the generic computer science and the term is perceived as a sub-discipline of information sciences. Yet, other years significantly diverge to signify a different agenda. For example, words like “machine intelligence”, “humanity”, “cognitive computing” group together around 2015 and some words from 2014 such as “doomsday”, “smart machines”, “human intelligence”, “weapons” and “doomsday” are close to this group. This suggests that like every new technology AI invokes the public imagination for utopian and dystopian fantasies. When we have a closer look into the articles in this group, we can see that they widely discuss the idea that humans will no longer be the dominant species on earth and will be replaced by intelligent machines. Reference to doomsday scenarios like in the movie *Terminator* where the artificial intelligence Skynet becomes self-aware and starts a nuclear strike on humanity are prevalent in these articles.

Some of the articles are more optimistic, suggesting a more symbiotic relationship between the intelligent machines and humans and perceive AI as a great collaborator to reduce the labour burden on humanity. All in all, these articles commonly discuss more philosophical and existential issues about the effects of AI on the future of humanity rather than concrete applications. 2016 and 2017 largely diverge from the previous years and the words around them are mostly grouped at the right-hand quadrant of the map. The words like “disruptive technologies”, “driverless cars”, “emerging technologies”, “machine learning”, “facial recognition”, “advanced data analytics” all signify prospects about the practical applications of AI and the changes they could bring to everyday life. Although 2016 and 2017 are quite distant in the map, this distance is because of the reference to the type of the technology rather than a substantive signifi-

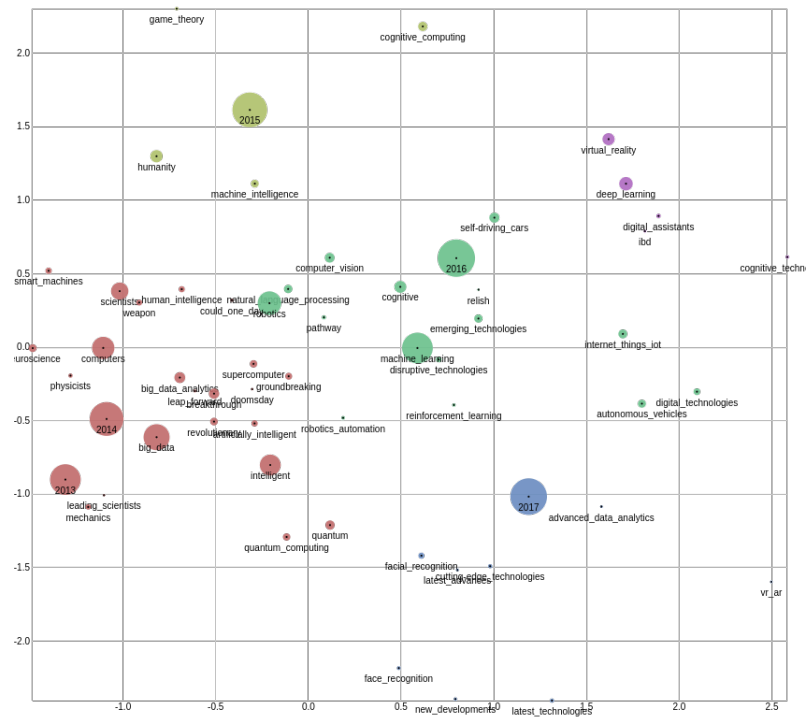


Fig. 2: Corresponding analysis of word usage

cation difference. While the hot topic for 2016 is “driverless cars”, “IOT” and “face recognition” underlines the year 2017. This suggests that each year a new technology is launched to excite the public imagination about its prospective applications. When we make a deeper reading of representative articles, the articles largely discuss about ground breaking products displacing an established technology and creating a completely new industry.

4.3 Change in Meaning of AI across Time

To monitor the semantic shifts in the sense of AI across time, we have first extracted a list of approximately eighty terms having the highest semantic similarity of word embeddings to AI from the entire corpus. This was needed to identify a lexicon representing the language of AI and facilitate the interpretation. Then, we have used these as a trimmed feature vector and recalculated the similarities of these terms to AI in each particular annual subset corpora. Hence, we were able to detect which set of words converged or diverged to represent the sense of AI for that particular year. We have further reduced the word space into two dimensions by applying Principal Components Analysis (PCA) and mapped the distance of the term AI in each year to the words in the lexicon. Fig.3 shows that the second dimension shows more variability around opposite poles and hence more easy to interpret. The south-end of this dimension is populated by words such as “killer robots”, “humanity” and “Kurzweil” that represent more philosophical-speculative side of AI discourse. The north end, on the other hand, is populated by words such as “IOT”, “machine learning”, “big data analytics” which represent more concrete applications of AI. Hence, over the years, the sense of AI changes from more speculative to more concrete.

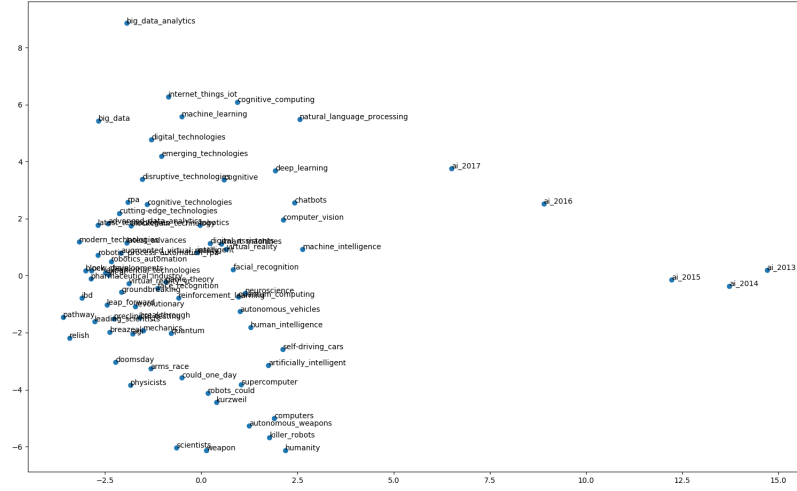


Fig. 3: PCA mapping of word embeddings for AI across years

Another interesting finding is that, while AI is quite distant to the overall lexicon in earlier years, it converges to the lexicon in the later years after 2015. While its sense stays more or less similar for the years before 2015 and distant to the global lexicon indicating a relative lexical poorness, we witness semantic shifts both in 2016 and 2017 consecutively towards the overall lexicon. We can interpret this as while AI did not have a steady language before 2016s, it has started to establish its own language with a specialized lexicon afterwards. To monitor the evolution of the semantic distance of each term to AI, we grouped the time series of words to five meaningful clusters by means of hierarchical cluster analysis. The clusters group terms in terms of their similarities in changes across time (for time-series clustering techniques see [17]). This was a necessary step for a tidier representation of the evolution of the more than eighty words as can be seen from the messy Fig.4. This helped us to focus on a deeper examination of the trend in semantic shifts.

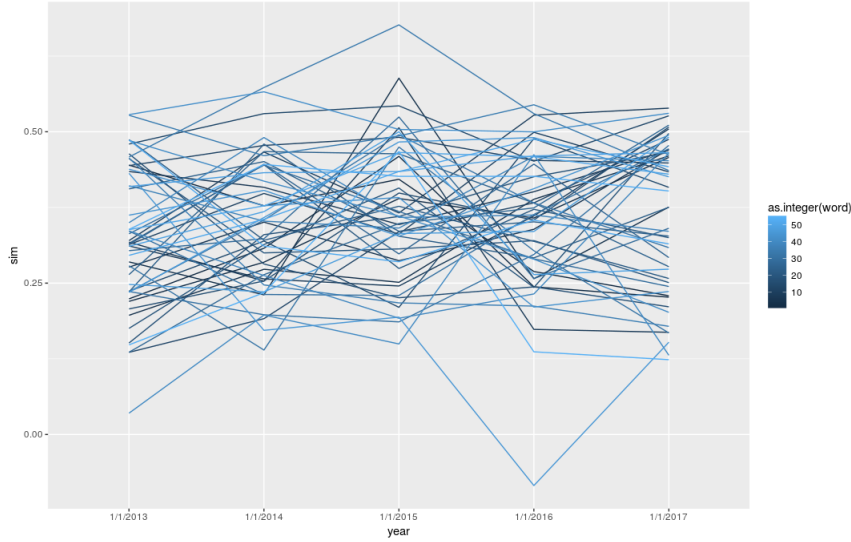


Fig. 4: Semantic similarities between all terms and AI over time

No matter how similar the words are to AI on average, those words showing same patterns across years are clustered by normalizing the matrix rows and applying the Euclidian distance. The first cluster contains the words denoting particular AI techniques such as “big data”, “cognitive computing”, “natural language processing”, “reinforcement learning” as shown in Fig.5-(a). They show an increasing trend in terms of converging to AI and their final similarity scores are about 0.4 and over in 2017. The second cluster as shown in Fig.5-(b), depicts a declining trend in an opposite manner to the first one. This cluster contains generic science and computer terms such as “game theory”, “mechanics”,

“quantum” and “super-computers”. This trend provides another evidence to our hypothesis that AI is distinguishing its language from that of generic science and establishing its own vocabulary. Third cluster is particular only to 2015 and is about the AI declaration by the experts we have mentioned earlier as we can see from the words it contains such as “arms race”, “autonomous weapons” and so on as shown in Fig.5-(c).

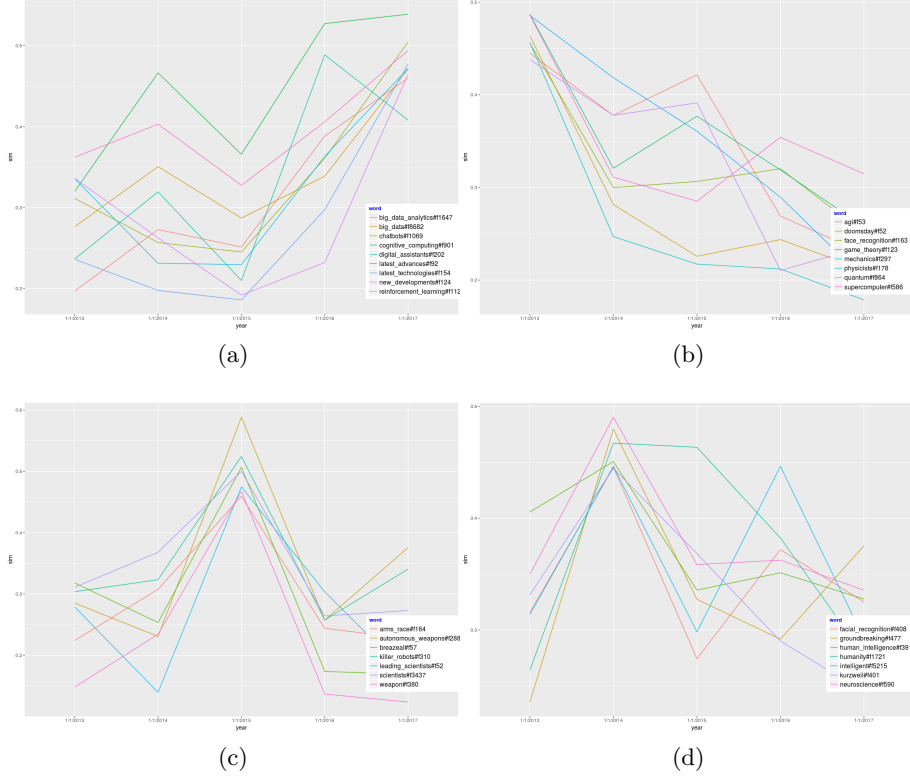


Fig. 5: A set of four subfigures describes: (a) Semantic similarities for first cluster; (b) Semantic similarities for second cluster; (c) Semantic similarities for third cluster; and, (d) Semantic similarities for fourth cluster

The fourth cluster in Fig.5-(d), shows a declining cyclic pattern and is also about the philosophical aspects of AI as we can observe from the words (Kurzweil, Humanity, human intelligence, neuro-science, ground-breaking) it contains. It represents more optimistic prospects about what AI would offer to humanity in the future. A close reading of the articles in this cluster presents interesting clues about prophetic claims about AI. Ray Kurzweil, a prominent futurist makes interesting claims about “transhumanism” which represents an intellec-

tual movement aiming to transform the human condition by means of sophisticated technologies to greatly enhance human intellect and physiology. This sense of AI fluctuates and comes to fore according to interesting events or declarations. Final cluster, in Fig.6-(a), shows a steadily increasing trend and represents the innovations and concrete applications of AI technology to everyday life (cutting-edge technologies, autonomous cars, IOT, robot automation etc.) providing another evidence to the hypothesis that AI is evolving from a more speculative sense towards a more down to earth sense and is establishing its own language.

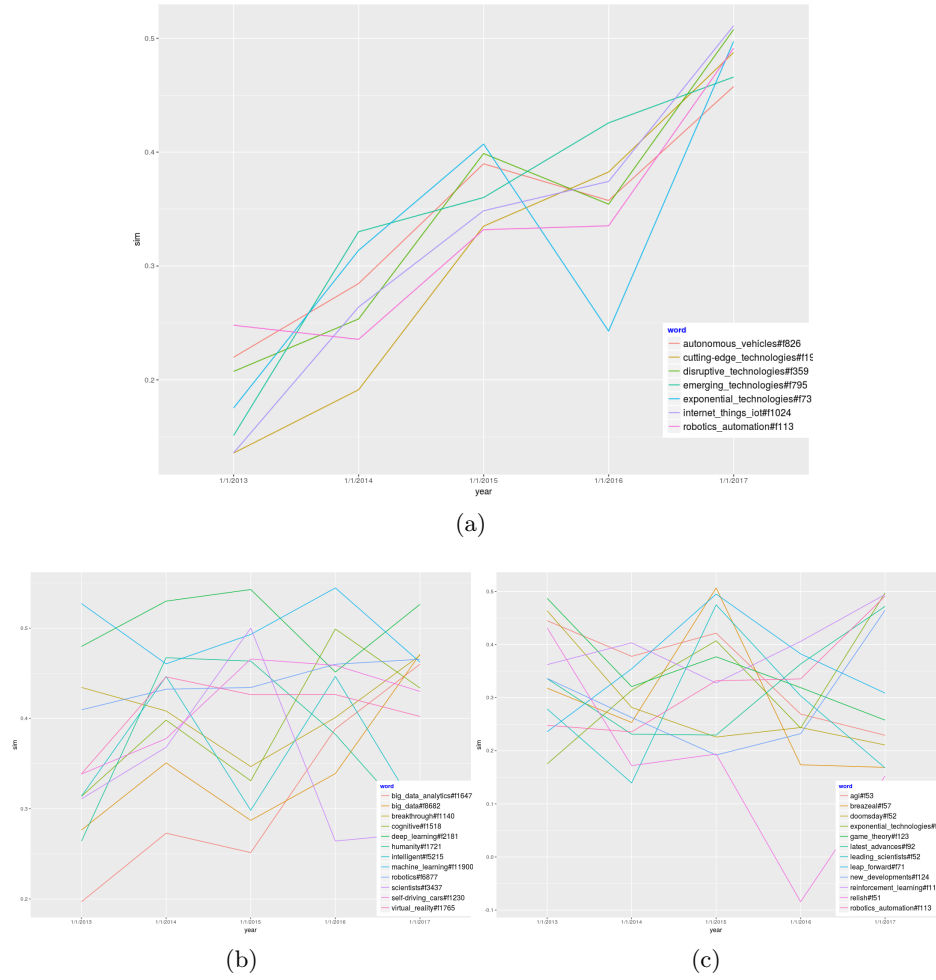


Fig. 6: A set of four subfigures: (a) Semantic similarities for fifth cluster; (b) Semantic shift of most frequent words; and, (c) Semantic shift of low frequent words

Finally, we checked the trend in most frequent and least frequent words. Time series is stationary and these terms do not show any trend or patterns in terms of their similarities to AI. This is understandable as these are either generic terms that might occur in every documents or specific terms occurring in a few documents as plotted in Fig.6-(b,c).

5 Conclusion

Our aim in this paper is to study the semantic shifts in the meaning of AI in popular discourse by examining the mapping of the words to different semantic vector spaces over time. The corpus is collected from media articles taken from major UK newspapers. We have applied a variety of techniques to understand the change in the meaning of AI. While Corresponding Analysis is applied to plot the word usage across time, the word embedding model has been utilized to represent the semantic shift of the words. To monitor the changes in word embeddings, PCA is applied to reduce dimensionality of embeddings and map the words on 2D space. We also use word embedding series to see the change of the meaning of AI over time where time series clustering is an important techniques to understand the series of terms. This led us to interpret the change and to monitor the semantic shifts in the sense of AI across time. All these experiments showed that, over the years, the sense of AI changes from more speculative to more concrete. They also provided some evidences to our hypothesis that AI is distinguishing its language from that of generic science and establishing its own vocabulary and AI is evolving from a more speculative sense towards a more down to earth sense.

References

1. Schutze, H., Hull, D.A., Pedersen, J.O. : A comparison of classifiers and document representations for the routing problem. In: Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '95), pp.229–237, (1995).
2. Lewis, D.D., Ringuette, M. : A comparison of two learning algorithms for text categorization. In: Symposium on Document Analysis and Information Retrieval, pp.81–83, (1994).
3. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, (2008).
4. Mikolov, T., Chen, K., Corrado, G., Dean, J. : Efficient estimation of word representations in vector space, arXiv:1301.3781. (2013).
5. Pennington, J., Socher, R., Manning, C. D. : Glove: Global Vectors for Word Representation. In : EMNLP, pp.1532–1543, (2014).
6. Gulordava, K., Baroni, M. : A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In : Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics (GEMS '11). pp.67–71, (2011).
7. Kulkarni, V., Al-Rfou', R., Perozzi, B., Skiena, S.: Statistically Significant Detection of Linguistic Change. CoRR abs/1411.3315, (2014).

8. Kenter, T., Wevers, M., Huijnen, P., de Rijke, M. : Ad Hoc Monitoring of Vocabulary Shifts over Time. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1191–1200 .
9. Hamilton, W.L., Leskovec, J., Jurafsky, D : Diachronic word embeddings reveal statistical laws of semantic change. In : ACL'16, pp 1489–1501, (2016).
10. Jatowt, A., Duh, K. : A framework for analyzing semantic change of words across time. In : JCDL, IEEE Computer Society, pp.229–238, (2014).
11. Azarbondy, H., Dehghani, M., Beelen, K., Arkut, A., Marx, M., Kamps, J. : Words are Malleable: Computing Semantic Shifts in Political and Media Discourse. In: CIKM17, pp.1509–1518, (2017).
12. Salton, G. : The SMART Retrieval System Experiments in Automatic Document Processing. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1971).
13. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R. : Indexing by latent semantic analysis. J. of the American Society for Information Science 41, pp.391–407, (1990).
14. Jurafsky, D., James H.M., Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. 2nd edition. Prentice-Hall, (2009).
15. Jolliffe, I.T. Principal Component Analysis, Springer, (2002).
16. Stephen R.B., Jonathan M.C., The role of factor analysis in the development and evaluation of personality scales, ProQuest, 2007.
17. Aghabozorgi, S., Shirkhorshidi, A.S. : Time-series clustering - A decade review. Information Systems 53, 16–38 (2015)
18. Greenacre, M., Blasius, J. : Correspondence Analysis and its Interpretation. In : Correspondence Analysis in the Social Sciences. Academic Press, London, (1994).