The Alan Turing Institute

UNIVERSITY OF CAMBRIDGE

WARWICK
THE UNIVERSITY OF WARWICK

A. Tsakalidis, M. Bazzi, M. Cucuringu, P. Basile, B. McGillivray

Mining the UK Web Archive for Semantic Change Detection

RANLP
03 Sept '19
Varna, Bulgaria

INTRODUCTION

- **Semantic Change:** Identify words whose lexical semantics have changed over time (tweet, follow, blackberry, etc.)
- **Importance**: historical/social studies, NLP tasks

atsakalidis@turing.ac.uk

INTRODUCTION

- **Semantic Change:** Identify words whose lexical semantics have changed over time (tweet, follow, blackberry, etc.)

- **Importance**: historical/social studies, NLP tasks

- **Challenges**: (a) lack of labelled datasets
  (b) short-term Semantic Change Detection
  (c) evaluation

CONTRIBUTIONS

- **Data**: Labelled dataset of word vectors during 2000–2013
- **Approach**: Variant of Procrustes Alignment trained on an extremely small number of "stable" words
- **Evaluation**: Rank-based approach

atsakalidis@turing.ac.uk

BACKGROUND

- **Early work:**
  - comparison of frequency & co-occurrence patterns between words across time [Sagi et al., 2009; Cook & Stevenson, 2010; Gulordava & Baroni, 2011]
- **Current approaches:**
  - Learn word representations over different time intervals & compute shift [Kim et al., 2014; Hamilton et al., 2016; Del Tredici et al., 2018]
  - Diachronic word embeddings [Bamler and Mandt, 2017; Rosenfeld and Erk, 2018; Yao et al., 2018; Rudolph and Blei, 2018]

BACKGROUND

- **Early work:**
  - comparison of frequency & co-occurrence patterns between words across time [Sagi et al., 2009; Cook & Stevenson, 2010; Gulordava & Baroni, 2011]
- **Current approaches:**
  - Learn word representations over different time intervals & compute shift [Kim et al., 2014; Hamilton et al., 2016; Del Tredici et al., 2018]
  - Diachronic word embeddings [Bamler and Mandt, 2017; Rosenfeld and Erk, 2018; Yao et al., 2018; Rudolph and Blei, 2018]

MAIN ISSUE

- **Current approaches:** learn word representations over different time intervals & compute shift [Kim et al., 2014; Hamilton et al., 2016; Del Tredici et al., 2018]

Issue: comparison of word vectors is impossible

MAIN ISSUE

- **Current approaches:** learn word representations over different time intervals & compute shift [Kim et al., 2014; Hamilton et al., 2016; Del Tredici et al., 2018]

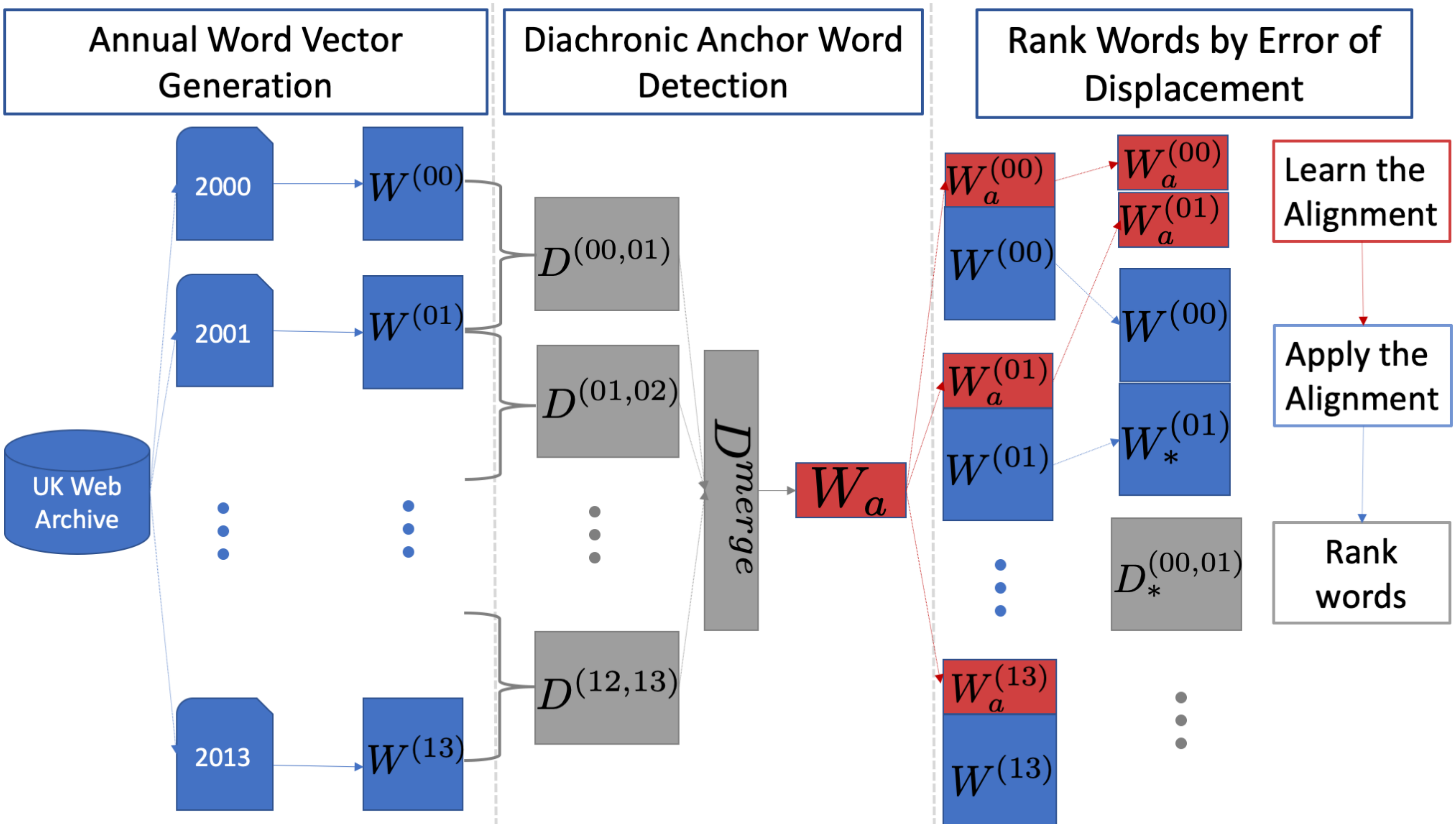
Issue: comparison of word vectors is impossible

Solutions:

- initialise word vectors at *t+1* based on the word vectors at *t* [Kim et al., 2014]
- Orthogonal Procrustes [Schönemann,1966]: map the word vectors of *the whole vocabulary* at *t* to their corresponding ones at *t+1* [Hamilton et al., 2016]

Problem: trying to align *all* words can be noisy (those that have changed meaning?)

APPROACH: OVERVIEW



APPROACH: DETAILS

- We want to find:

$$R = \underset{\Omega;\Omega^T\Omega=I}{\operatorname{argmin}} \left\| W^{(t)}\Omega - W^{(t+1)} \right\|_F$$

- Solved as:

$$R = UV^T$$, where

$$U\Sigma V^T = SVD(W^{(t+1)}W^{(t)T})$$

atsakalidis@turing.ac.uk

APPROACH: DETAILS

- We want to find: $$R = \underset{\Omega;\Omega^T\Omega=I}{\operatorname{argmin}} \left\| W^{(t)}\Omega - W^{(t+1)} \right\|_F$$
- Solved as: $$R = UV^T$$, where $$U\Sigma V^T = SVD(W^{(t+1)}W^{(t)T})$$
- Align vectors as: $$W_*^{(t+1)} = W^{(t+1)}R^T s$$
- Displacement Error: $D^{(t,t+1)}$: *dist* $\{W^{(t)}, W_*^{(t+1)}\}$

APPROACH: DETAILS

- We want to find: $$R = \underset{\Omega;\Omega^T\Omega=I}{\operatorname{argmin}} \left\| W^{(t)}\Omega - W^{(t+1)} \right\|_F$$
- Solved as: $$R = UV^T$$, where $$U\Sigma V^T = SVD(W^{(t+1)}W^{(t)T})$$
- Align vectors as: $$W_*^{(t+1)} = W^{(t+1)}R^Ts$$
- Displacement Error: $D^{(t,t+1)}$: *dist* $\{W^{(t)}, W_*^{(t+1)}\}$
- **Anchors:** *k* words with lowest displacement error between *[t, t+1]*

APPROACH: DETAILS

- We want to find: $$R = \underset{\Omega;\Omega^T\Omega=I}{\operatorname{argmin}} \left\| W^{(t)}\Omega - W^{(t+1)} \right\|_F$$
- Solved as: $$R = UV^T$$, where $$U\Sigma V^T = SVD(W^{(t+1)}W^{(t)T})$$
- Align vectors as: $$W_*^{(t+1)} = W^{(t+1)}R^T s$$
- Displacement Error: $D^{(t,t+1)}$: *dist* $\{W^{(t)}, W_*^{(t+1)}\}$
- **Anchors:** *k* words with lowest displacement error between *[t, t+1]*
- **Diachronic Anchors:** *k* words with lowest avg displacement error across time

APPROACH: DETAILS

- We want to find: $$R = \underset{\Omega;\Omega^T\Omega=I}{\operatorname{argmin}} \left\| W^{(t)}\Omega - W^{(t+1)} \right\|_F$$
- Solved as: $$R = UV^T$$, where $$U\Sigma V^T = SVD(W^{(t+1)}W^{(t)T})$$
- Align vectors as: $$W_*^{(t+1)} = W^{(t+1)}R^T s$$
- Displacement Error: $$D^{(t,t+1)}: \quad \textit{dist}\,\{W^{(t)}, W_*^{(t+1)}\}$$
- **Anchors:** *k* words with lowest displacement error between *[t, t+1]*
- **Diachronic Anchors:** *k* words with lowest avg displacement error across time
- **Core idea:** learn a new alignment based *strictly* on (diachronic) anchors

atsakalidis@turing.ac.uk

DATASET

- **Word Vectors**:
  - *UK Web Domain Dataset 1996-2013* (>20TB) [Basile & McGillivray, 2018]
  - Annual word2vec representations [Mikolov et al., 2013] during 2000–2013

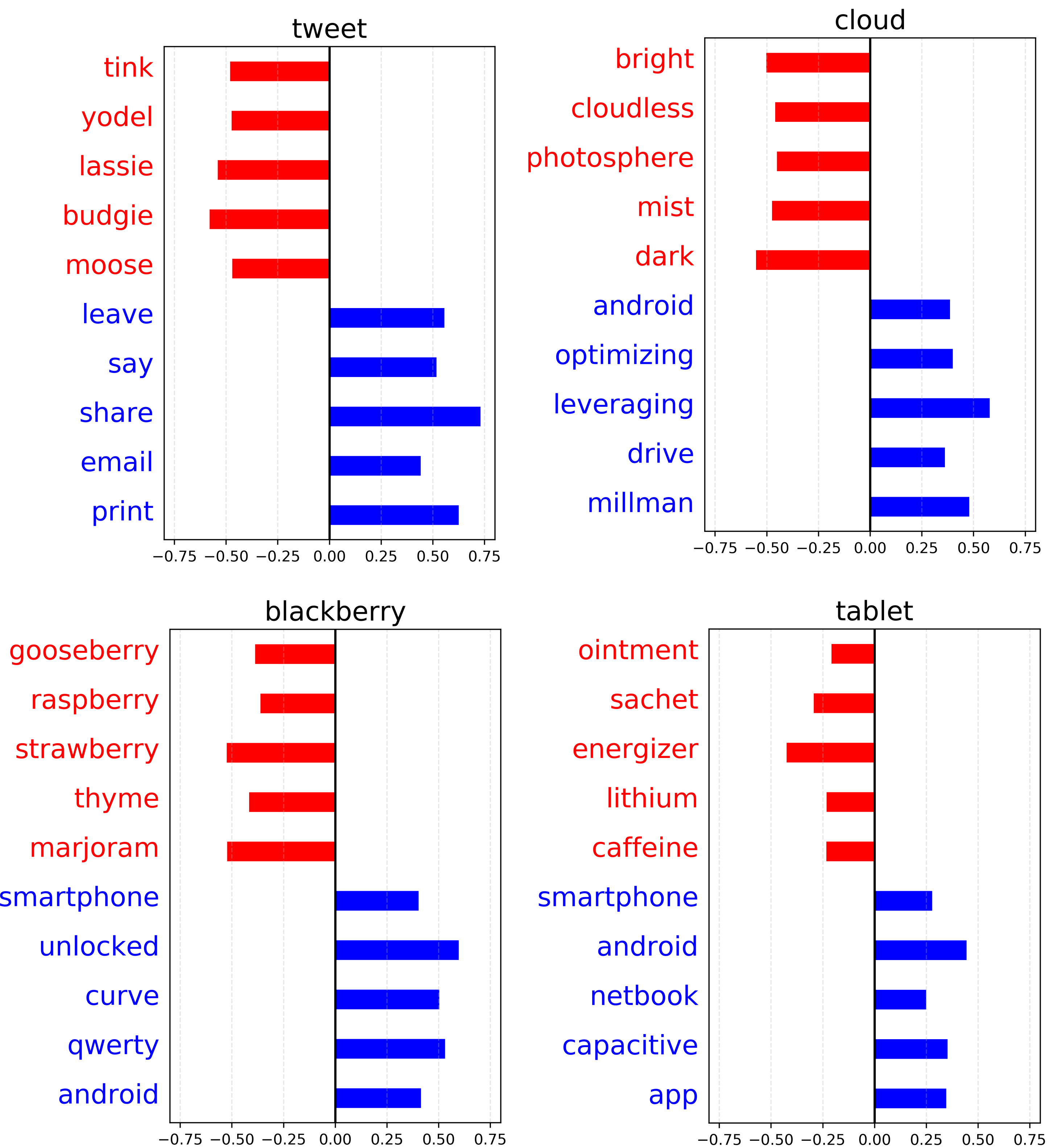
DATASET

- **Word Vectors**:
  - *UK Web Domain Dataset 1996-2013* (>20TB) [Basile & McGillivray, 2018]
  - Annual word2vec representations [Mikolov et al., 2013] during 2000–2013
- **Ground Truth**:
  - *Oxford English Dictionary* (>250K lemmata)
  - Gold standard: 208 words with a new meaning after the year 2000

DATASET

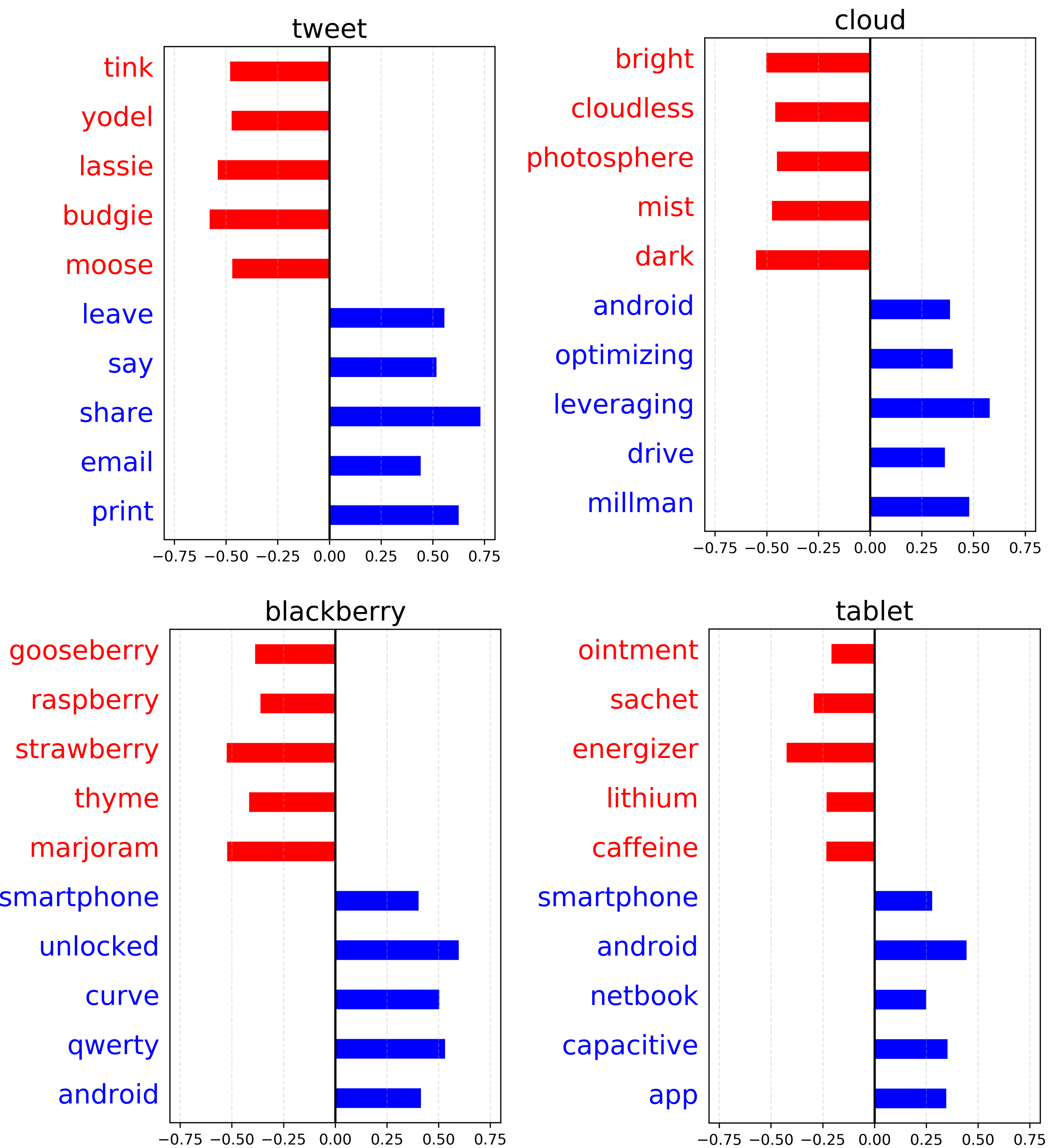
- **Word Vectors**:
  - *UK Web Domain Dataset 1996-2013* (>20TB) [Basile & McGillivray, 2018]
  - Annual word2vec representations [Mikolov et al., 2013] during 2000–2013
- **Ground Truth**:
  - *Oxford English Dictionary* (>250K lemmata)
  - Gold standard: 208 words with a new meaning after the year 2000
- **Intersected Dataset**:
  - 47,886 word vectors per year (2000-13)
  - 65 words with known semantic shift

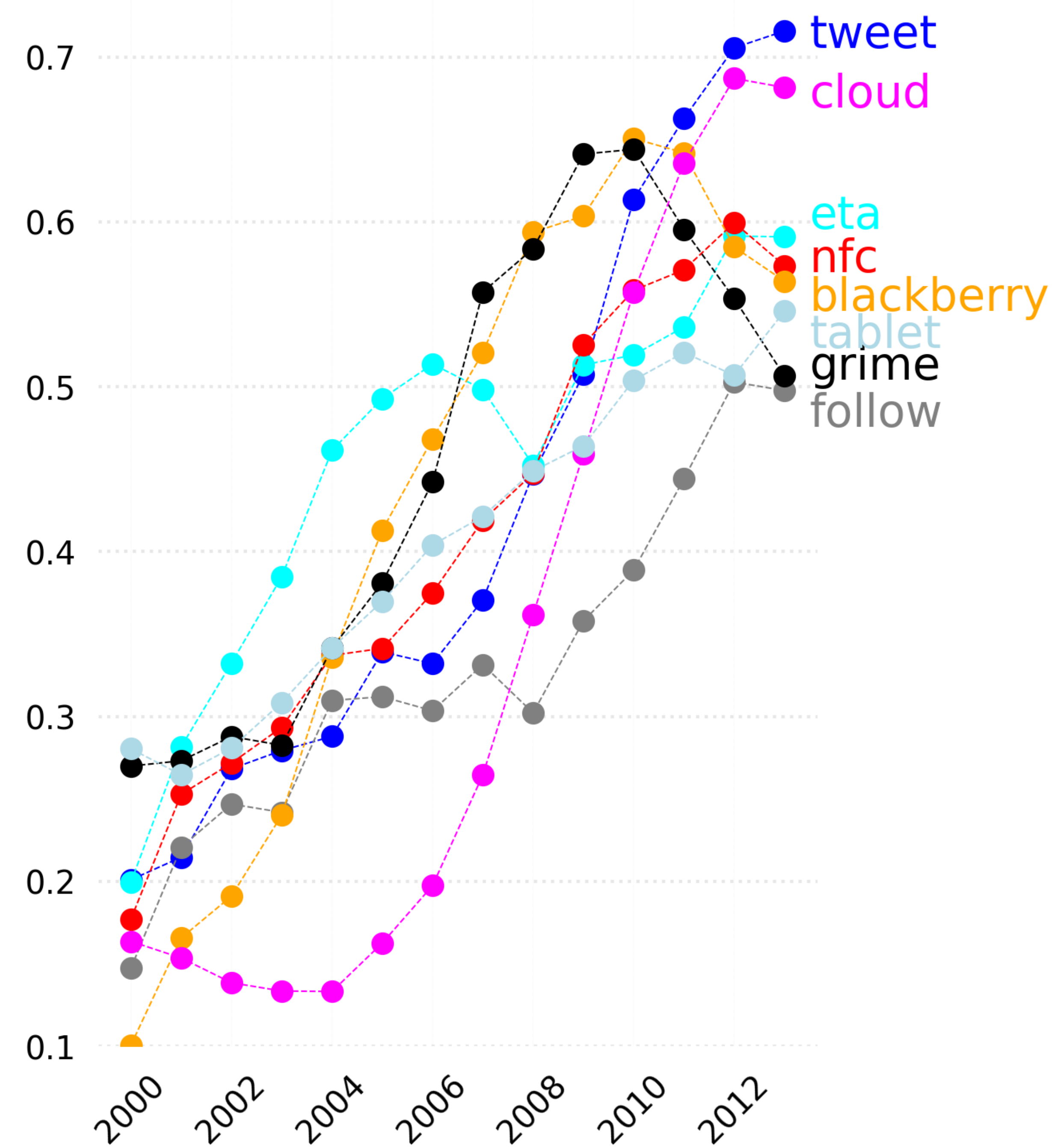
EMPIRICAL INSIGHTS



atsakalidis@turing.ac.uk

EMPIRICAL INSIGHTS





atsakalidis@turing.ac.uk

EXPERIMENTS

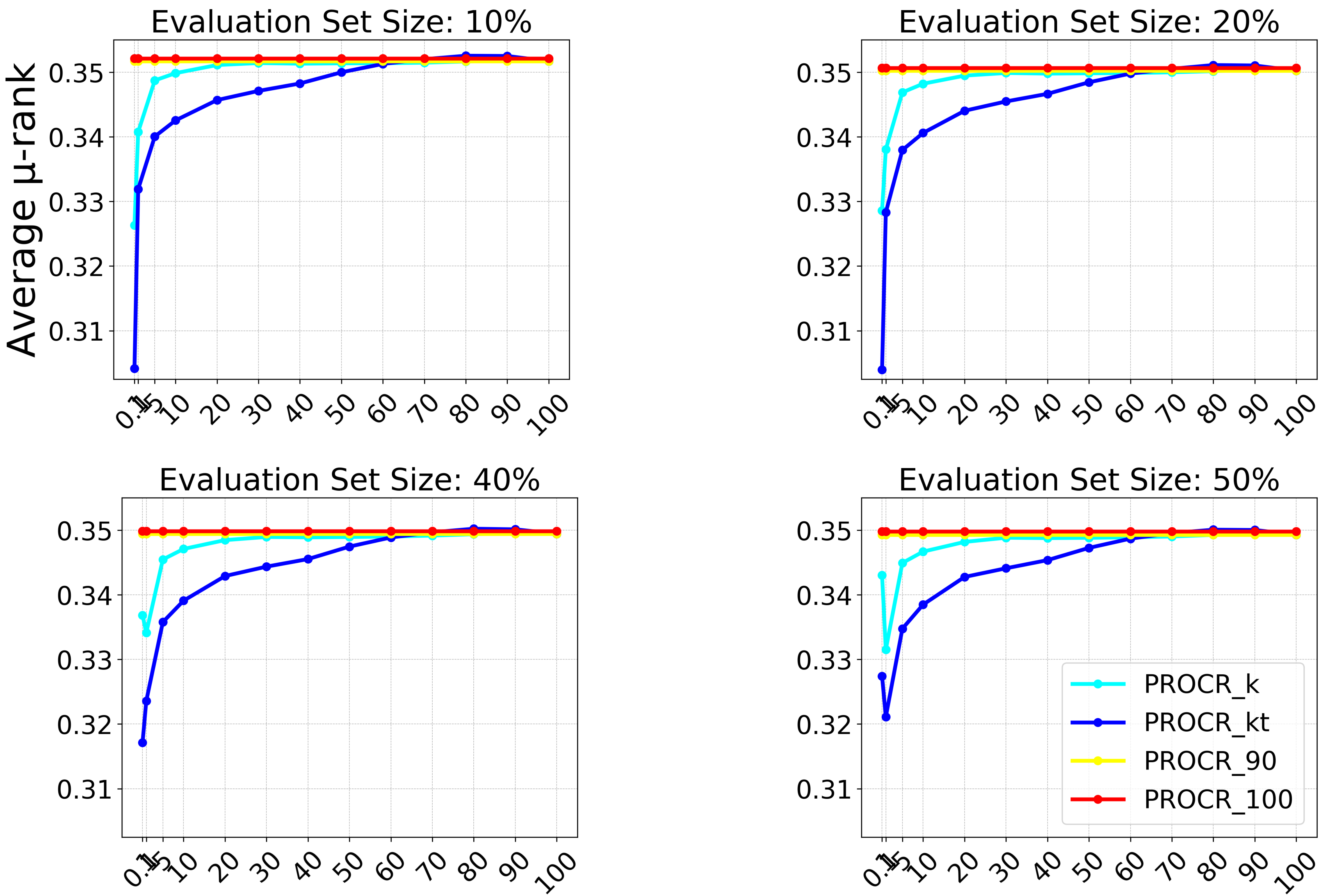
- **Experiment 1:** train/eval split | eval set size: {10%, …, 50%} | 40 runs
- **Experiment 2:** single run (whole dataset)

- **Models:**
  - PROCR_90: alignment based on train set (Experiment 1)
  - PROCR_100: alignment using all words
  - PROCR_k: alignment based on anchor words
  - PROCR_kt: alignment based on diachronic anchor words

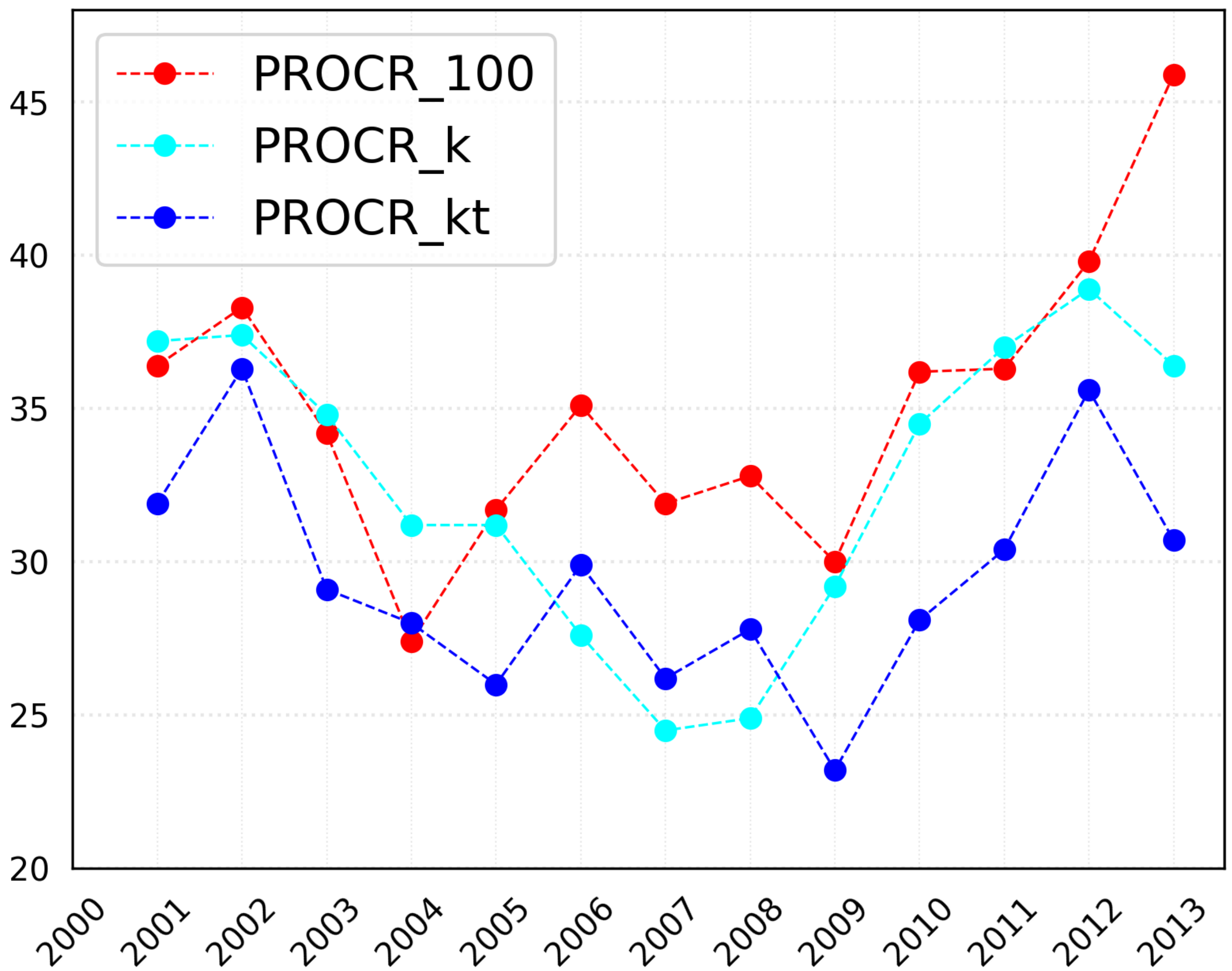
- **Evaluation:** rank (%) of a semantically shifted word based on displacement error (*μ-rank*)

atsakalidis@turing.ac.uk

RESULTS: EXPERIMENT 1



RESULTS: EXPERIMENT 2

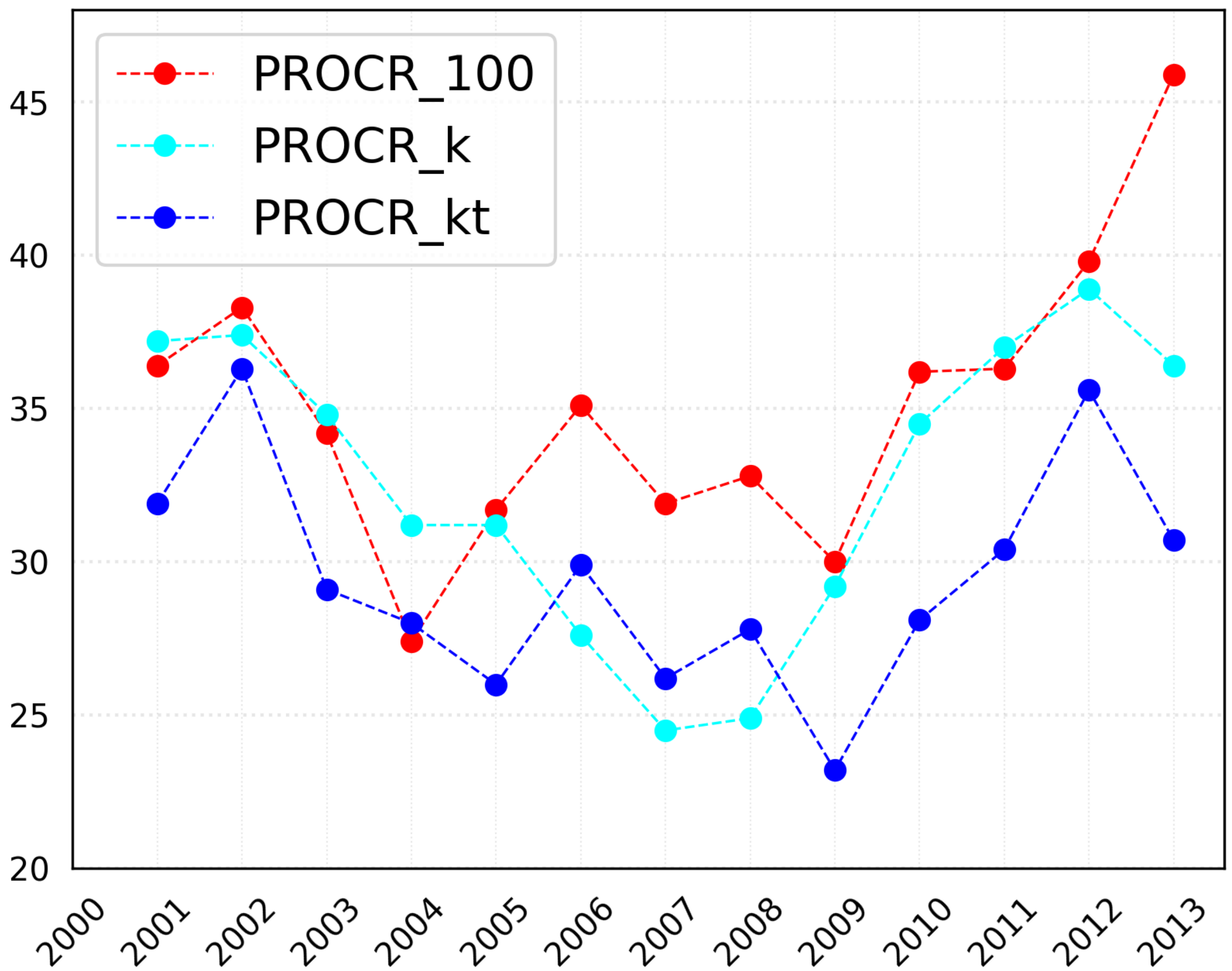


PROCR_100: 35.08±4.71

PROCR_k: 32.68±4.93

PROCR_kt: 29.48±3.67

RESULTS: EXPERIMENT 2



PROCR_100: 35.08±4.71

PROCR_k: 32.68±4.93

PROCR_kt: 29.48±3.67

***Anchor* vs *Diachronic Anchor* words:**

- #Diachronic Anchor words: 48 (i.e., top-0.1%)
- #Anchor words: 434 (from an overall possible of 631)
- Only 16% of anchor are diachronic anchor words (noise)

Diachronic Anchor words: more robust short-term alignments

atsakalidis@turing.ac.uk

CONCLUSION & FUTURE WORK

- **Summary**:
  - New labelled dataset for Semantic Change Detection
  - Rank-based task & evaluation
  - Procrustes Alignment based on a few diachronically stable words
- **Future Work**:
  - Generalised Procrustes Alignment [Gower, 1975]
  - Temporal Modelling (Temporal Clustering, Changepoint Detection)

atsakalidis@turing.ac.uk

REFERENCES

[1] Robert Bamler and Stephan Mandt. 2017. Dynamic word embeddings via skip-gram filtering. stat 1050:27.

[2] Pierpaolo Basile and Barbara McGillivray. 2018. Exploiting the Web for Semantic Change Detection. In International Conference on Discovery Science. Springer, pages 194–208.

[3] Paul Cook and Suzanne Stevenson. 2010. Automatically Identifying Changes in the Semantic Orientation of Words. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).

[4] Marco Del Tredici, Raquel Fernandez, and Gemma Boleda. 2018. Short-term meaning shift: an exploratory distributional analysis. arXiv preprint arXiv:1809.03169 .

[5] John C Gower. 1975. Generalized procrustes analysis. Psychometrika 40(1):33–51.

[6] Kristina Gulordava and Marco Baroni. 2011. 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. pages 67–71.

[7] William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). volume 1, pages 1489–1501.

[8] Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal Analysis of Language through Neural Language Models. In Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science. pages 61–65.

[9] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, pages 625–635.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems. pages 3111–3119.

[11] Alex Rosenfeld and Katrin Erk. 2018. Deep neural models of semantic shift. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). pages 474–484.

[12] Maja Rudolph and David Blei. 2018. Dynamic embeddings for language evolution. In Proceedings of the 2018 World Wide Web Conference on World Wide Web. International World Wide Web Conferences Steering Committee, pages 1003–1011.

[13] Eyal Sagi, Stefan Kaufmann, and Brady Clark. 2009. Semantic density analysis: Comparing word meaning across time and phonetic space. In Proceedings of the Workshop on Geometrical Models of Natural Language Semantics. Association for Computational Linguistics, pages 104–111.

[14] Peter H Schönemann. 1966. A generalized solution of the orthogonal procrustes problem. Psychometrika 31(1):1–10.

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

THANK YOU!

Any
Questions?

Dataset: https://github.com/adtsakal/Semantic_Change

@adtsakal
@BarbaraMcGilli
@basilepp

atsakalidis@turing.ac.uk

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1 and the seed funding grant SF099.