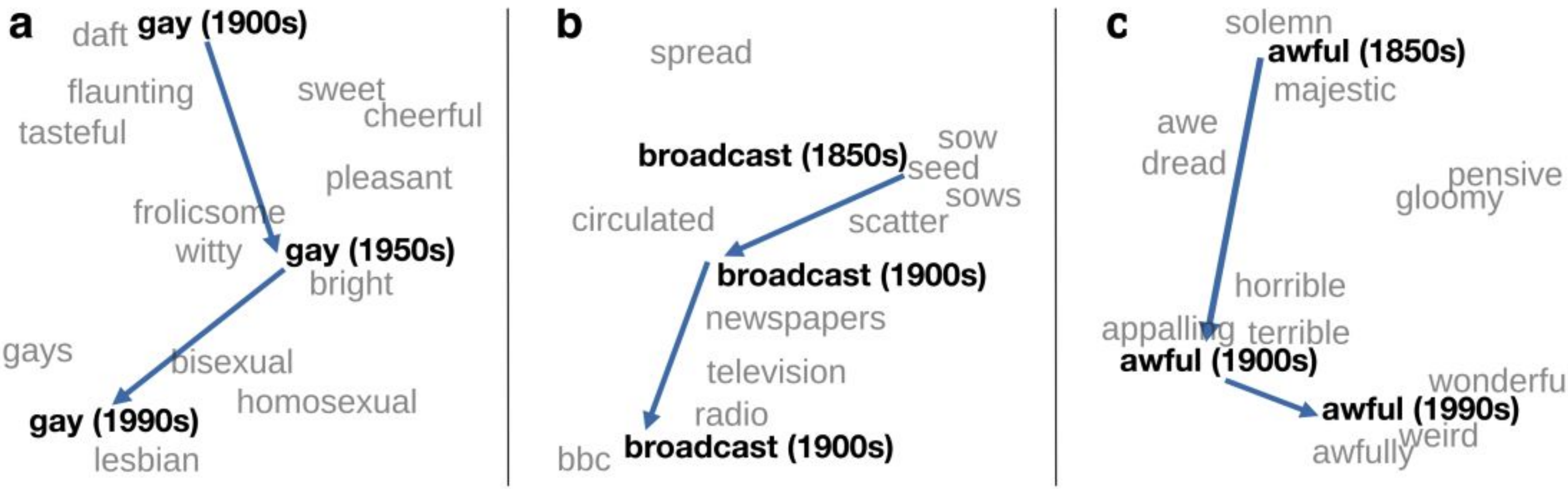
A Semantic Distance Metric Learning approach for Lexical Semantic Change Detection

Taichi Aida[1] and Danushka Bollegala[2]
[1]Tokyo Metropolitan University
[2]University of Liverpoolz

Background: Semantic Change Detection (SCD)

- Words can have different meanings overtime
- **Manual detection is laborious**

  → **Automatic detection** (e.g. vector space)

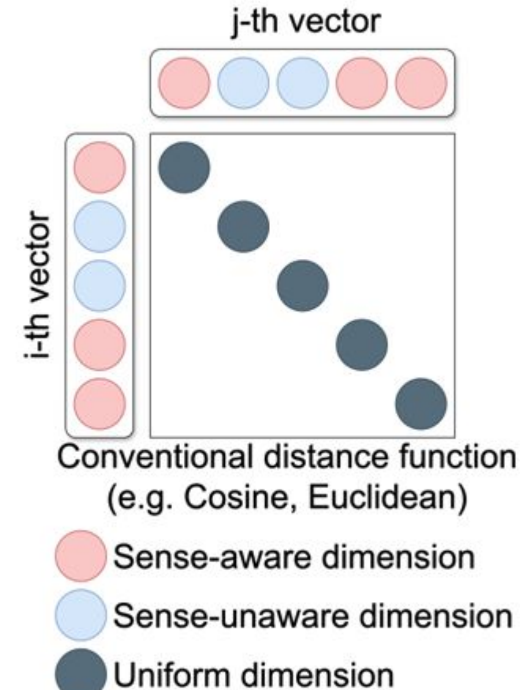


[Hamilton+16] Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

Background: Two challenges in SCD

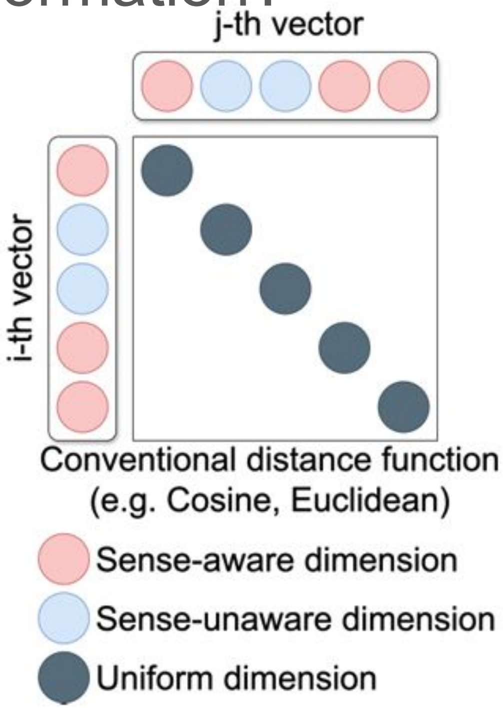
- **Representational challenge**

- **Measurement challenge**



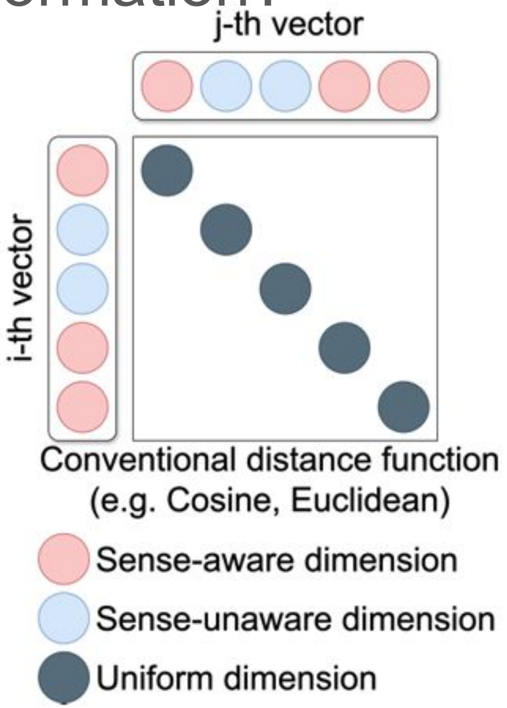
Background: Two challenges in SCD

- **Representational challenge**
  - word can take **different meanings** in different contexts within the **same corpus**
  - should we consider **time** or **sense** information?
    →**time << sense** [Cassotti+23]
- **Measurement challenge**



Background: Two challenges in SCD

- **Representational challenge**
  - word can take **different meanings** in different contexts within the **same corpus**
  - should we consider **time** or **sense** information?
    →**time << sense** [Cassotti+23]
- **Measurement challenge**
  - SCD is an **unsupervised task**
    →comparing vector sets using **parameter-free distance functions** (Cosine, Euclidean)
  - considers **sense-aware**/**unaware** dimensions **uniformly**

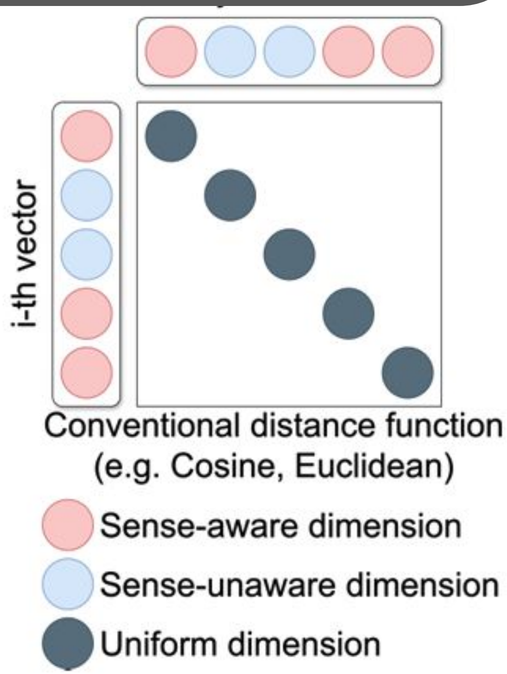


Background: Two challenges in SCD

How to overcome the measurement challenge?

→**time << sense** [Cassotti+23]

- **Measurement challenge**
  - SCD is an **unsupervised task** →comparing vector sets using **parameter-free distance functions** (Cosine, Euclidean)
  - considers **sense-aware**/**unaware** dimensions **uniformly**



Proposal: Semantic Distance Metric Learning (SDML)

- SDML = **sense-aware supervision** + **metric learning**

Proposal: Semantic Distance Metric Learning (SDML)

- SDML = **sense-aware supervision**
  \+ **metric learning**
  - We use **Mahalanobis distance**

$$h(\boldsymbol{w}_i, \boldsymbol{w}_j; \mathbf{A}) = (\boldsymbol{w}_i - \boldsymbol{w}_j)^\top \mathbf{A}(\boldsymbol{w}_i - \boldsymbol{w}_j)$$

  - Mahalanobis matrix $\mathbf{A}$ is optimised for **sense-aware supervision** using **metric learning**
    - **sense-aware supervision**: Word-in-Context (WiC)
    - **metric learning**: Information Theoretic Metric Learning (ITML)

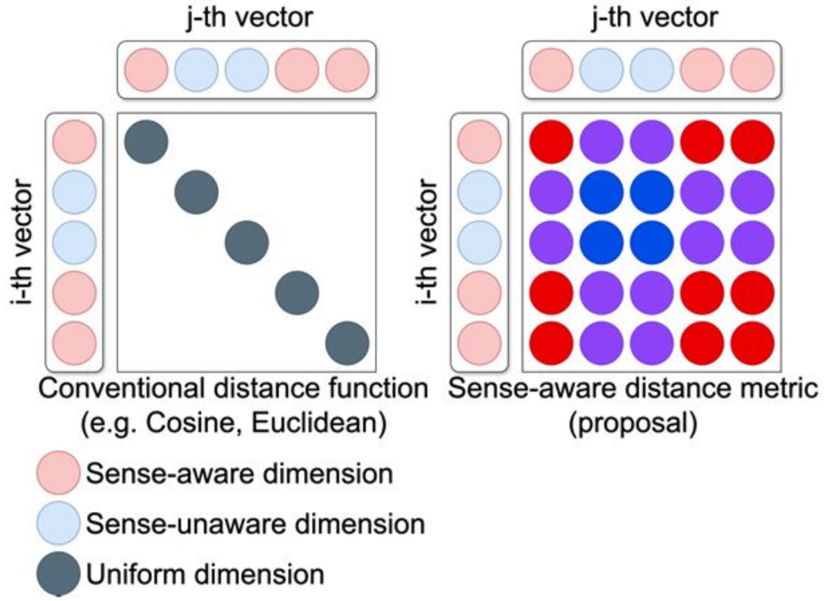
$$\begin{aligned} \min_{\mathbf{A}} \quad & \mathrm{KL}(p(\boldsymbol{w}; \mathbf{A}_0) || p(\boldsymbol{w}; \mathbf{A})) \\ \text{subject to} \quad & h(\boldsymbol{w}_1, \boldsymbol{w}_2) \leq u \quad y = 1, \\ & h(\boldsymbol{w}_1, \boldsymbol{w}_2) \geq \ell \quad y = 0. \end{aligned}$$

Proposal: Semantic Distance Metric Learning (SDML)

- SDML vs measurement challenge

✅ considers **sense-aware**/**unaware** dimensions
(diagonal elements of Mahalanobis matrix $\mathbf{A}$)

✅ accounts for **cross-dimensional** information
(off-diagonal elements of Mahalanobis matrix $\mathbf{A}$)



Experiment: Settings

- **Task 1: WiC (binary classification)**
  - "They stopped at an open *space*(🏠) in the jungle"
  - "The astronauts walked in outer *space*(🪐) without a tether"
  - word "*space*" takes the same meaning?: False
- **Task 2: SCD (ranking)**
  - "If a *plane*(🔲) be parallel to the horizontal…"
    "The sun is in the same *plane*(🔲) as the picture…"
  - "The President's *plane*(✈️) landed at Goose Bay…"
    "The *plane*(✈️) kept climbing and climbing…"
  - the meaning of "*plane*" is changed?: True (degree: 0.7)

Experiment: WiC (binary classification)

- **SDML significantly enhances performance** in multiple languages

Method	En	De	Fr	It	Ru
XLM-RoBERTa [Conneau+20]	86.6	84.0	76.2	72.3	80.9
+Sense-aware Fine-tuning [Cassotti+23]	78.0	78.3	73.2	67.1	78.2
+SDML [ours]	**90.3**	**84.9**	**78.7**	**75.3**	**87.6**

Experiment: SCD (ranking)

- **SDML improves performance by 2~5%**
- **A**(*diag*) < **A**(full): **cross-dimensional** information is also important

Method	En	De	Sv	La	Ru
Sense-aware Fine-tuning [Cassotti+23]	0.757	0.877	**0.754**	0.056	0.775
+SDML (*diag*) [ours]	0.750	0.902	0.642	0.083	0.804
+SDML (full) [ours]	**0.774**	**0.902**	0.656	**0.124**	**0.805**

Ablation: Correlation analysis

- Calculate correlations with
  - gold labels (Gold)
  - number of senses (#Synsets)
  - frequency (Freq. C1/C2)

Method	Gold	WordNet #Synsets	Freq. C1	Freq. C2
Sense-aware Fine-tuning [Cassotti+23]	0.757	**0.427**	-0.182	-0.062
+SDML (*diag*) [ours]	0.750	0.355	**-0.205**	**-0.121**
+SDML (full) [ours]	**0.774**	0.404	-0.122	-0.037

Ablation: Correlation analysis

- **Higher correlation with polysemy** contributes to the performance improvement
  - **Consistent with the law of innovation** [Hamilton+16] "polysemous words tend to change their meanings"

Method	Gold	WordNet #Synsets	Freq. C1	Freq. C2
Sense-aware Fine-tuning [Cassotti+23]	0.757	**0.427**	-0.182	-0.062
+SDML (*diag*) [ours]	0.750	0.355	**-0.205**	**-0.121**
+SDML (full) [ours]	**0.774**	0.404	-0.122	-0.037

Ablation: Correlation analysis

- **Lower correlation with frequency** contribute to the performance improvement
  - **Differ from the law of conformity** [Hamilton+16] "frequent words tend to be stable"

Method	Gold	WordNet #Synsets	Freq. C1	Freq. C2
Sense-aware Fine-tuning [Cassotti+23]	0.757	**0.427**	-0.182	-0.062
+SDML (*diag*) [ours]	0.750	0.355	**-0.205**	**-0.121**
+SDML (full) [ours]	**0.774**	0.404	-0.122	-0.037

Conclusion

- SCD task has two challenges
  - **Representational challgenge**
  - **Measurement challenge**
- To overcome the measurement challenge, we propose SDML
  - SDML = **sense-aware supervision** + **metric learning**
- Experimental results show that
  - **SDML enhances performance** in WiC and SCD tasks
  - Considering **cross-dimensional** information is also important

Future directions

- Considering **time** **+** **sense** information
    - **time**: Time Masking [Rosin+22a], Temporal Attention [Rosin+22b] …
        - **LLMs are not temporally grounded [Qiu+24]**
    - **sense**: Sense-aware fine-tuning [Cassotti+23] / distance metric [Ours] …
    - **time + sense = ? (e.g. TempoWiC [Loureiro+22])**
- Beyond the **degree** of semantic change
    - semantic change **types** [Cassotti+24, Periti+24]
    - find **new sense** [Mariia+24]
    - more …?