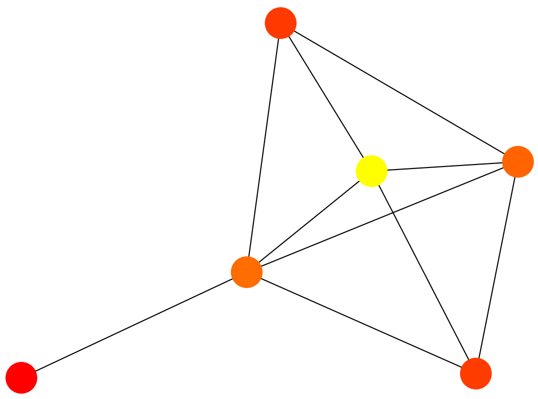
EPFL



Graph-based text representations

Student

Wentao Feng

Thesis supervisors

Maxime Peyrard

Akhil Arora

Thesis advisor

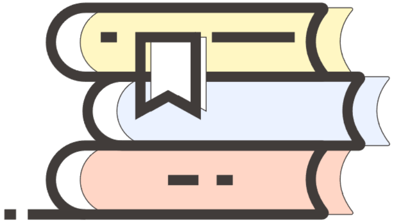
Robert West

External expert

Andreas Spitz

dlab

12 July 2021



Introduction

Method

Experiment

Conclusion

Introduction - Problem

- **Text representation**

  A process converts text into a mathematically computable form

- **Existing models**

  Word2Vec, GloVe, Fasttext, etc.

- **Drawback**

  Learned representations are not understandable to the human.

EX writings (0.189453, 0.210938, 0.205078, 0.289062, 0.21875, …)
yellow (-0.073242, 0.026367, 0.076171, 0.189453, -0.0471, …)

Introduction - Motivation

Why do we need an interpretable text representation?

- Debug pipeline
- Dimension reduction
- Improve downstream tasks
- ⋮

Introduction - Challenge

- **Interpretable embedding space**

  Sparse-vector-based methods: one-hot encoding, occurrence matrix

- **Efficiency**

  Dense-vector-based methods: Word2Vec, GloVe, etc.

Introduction - Previous work

SPOWV[1], SPINE[2]:

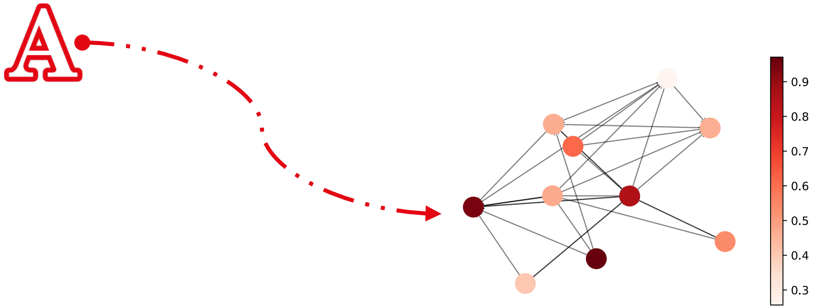
Transform dense-vector-based models with sparsity and non-negativity constraint.

- ⊕ Axis has a topic or concept
- ⊖ Small capacity of interpretability
- ⊖ Sub-optimality
- ⊖ Pre-trained models matter

SPOWV[1]: Sparse Overcomplete Word Vector
SPINE[2]: Sparse Interpretable Neural Embeddings

Introduction - Solution

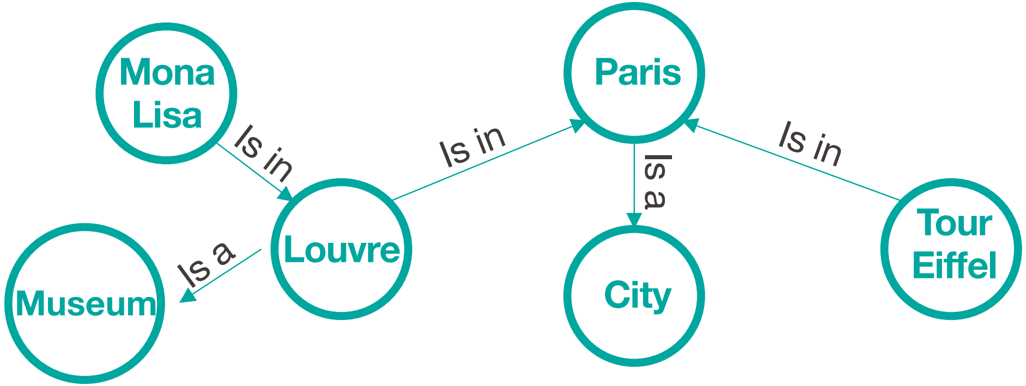
Representing word as a distribution on the knowledge graph



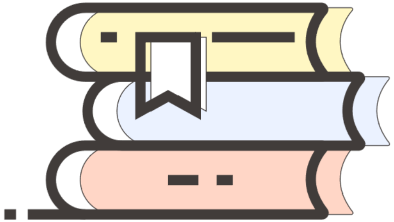
Method in a nutshell:

1. Build Skip-gram dataset.
2. Embed word as Gaussian mixture distribution
3. Measure the statistical distance between two distributions
4. Maximize the objective of negative samplings

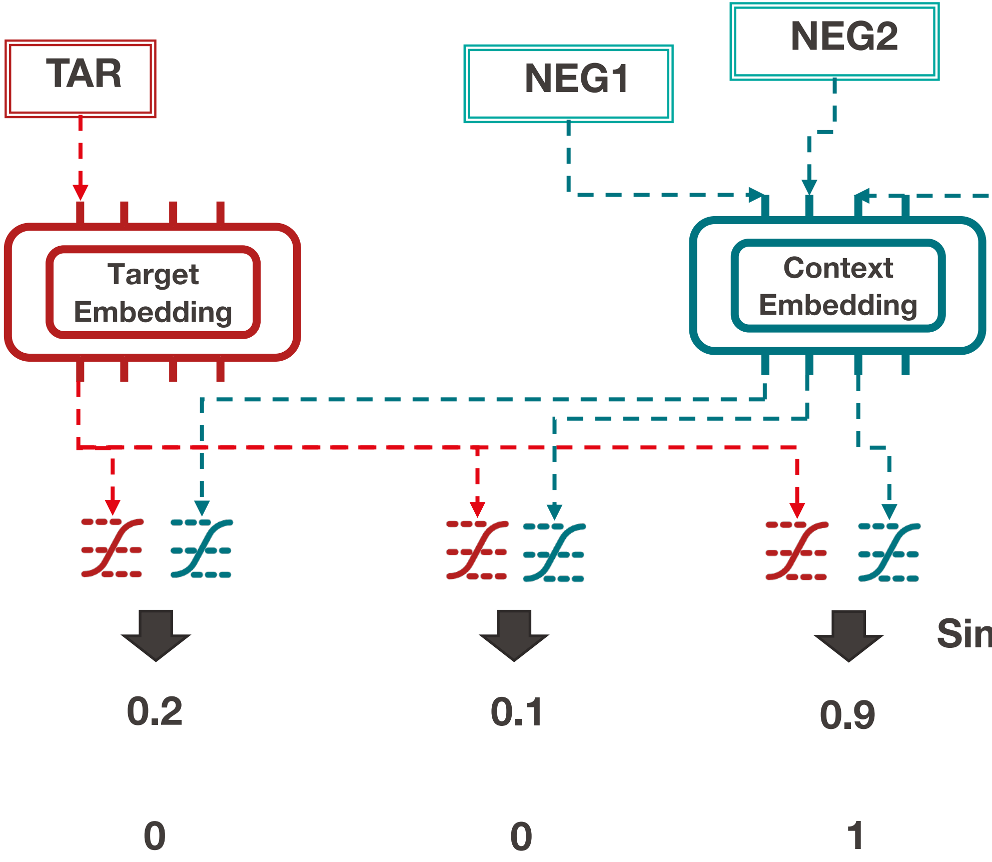
Introduction - Solution



- **Naturally interpretable**
  - EX academia (efficiency: 0.5061 traveling: 0.4707 rookie: 0.4327 upset: 0.4250 ricardo: 0.3571 penalty: 0.3485…)
- **High efficiency and scalability after relaxations**



Method - Skip-gram



TAR: target word

CTX: context word

NEG: negative sampling

Similarity function & Sigmoid

Label

Method - Objective function

$$\log \sigma(\mathcal{S}(w_j, w_i|\mathcal{G})) + \sum_{k=1}^{K} \mathbb{E}_{w_k \sim p_n(w)}[\log \sigma(-\mathcal{S}(w_k, w_i|\mathcal{G}))]$$

w_i: *target word*

w_j: *context word*

w_k: *Negative sampling*

$p_n(w)$: *Negative sampling distribution*

σ: *sigmoid function*

S: *similarity function*

G: *knowledge graph*

Method - Issue

$$\log \sigma(\mathcal{S}(w_j, w_i|\mathcal{G})) + \sum_{k=1}^{K} \mathbb{E}_{w_k \sim p_n(w)} [\log \sigma(-\mathcal{S}(w_k, w_i|\mathcal{G}))]$$

- **Discrete structure**
  - Back propagation
  - Measure similarity
- **Efficiency and scalability**
  - Work on large knowledge graph

Method - Relaxation

Discrete structure

- Back propagation → Node embedding
- Measure similarity → Node embedding

Efficiency and scalability

- Work on large knowledge graph → Gaussian mixture model (GMM), Anchor

Node embedding

Gaussian mixture model (GMM)

Anchor

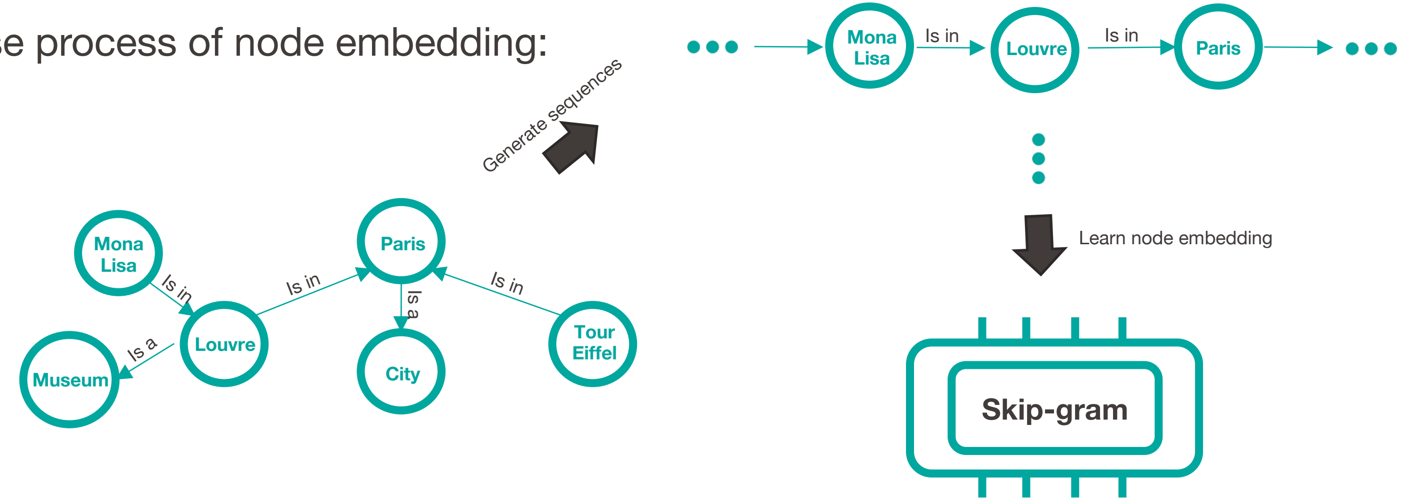
Method - Relaxation

Node embedding Gaussian mixture model (GMM) Anchor

Desiderata:

- ❑ Use dense vector represent node.
- ❑ Preserve structural information and node similarity.

Concise process of node embedding:



Method - Relaxation

Node embedding   **Gaussian mixture model (GMM)**   Anchor

$$P(x) = \sum_{m=1}^{M} a_m \mathcal{N}(x|\mu_m, \Sigma_m) \quad s.t. \sum_{m=1}^{M} a_m = 1, a_m \geq 0$$

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

Desiderata:

- Efficient: closed-form statistical distance
- Expressiveness: more components, more expressive.
- Sparse: small area has positive probability density

Method - Relaxation

Node embedding **Gaussian mixture model (GMM)** Anchor

Simplify Σ (Positive semidefinite matrix):

1. Diagonal matrix.
2. Treat as hyperparameter.
3. All Σ are the same.

Squared l_2 distance between two mixtures of Gaussian P, Q:

$$\ell_2^2(P,Q) = \sum_{m,m'} a_m a_{m'} \mathcal{N}(\mu_m | \mu_{m'}, \Sigma_m + \Sigma_{m'}) + \sum_{n,n'} b_n b_{n'} \mathcal{N}(\eta_n | \eta_{n'}, \Lambda_n + \Lambda_{n'}) - 2 \sum_{m,n} a_m b_n \mathcal{N}(\mu_m | \eta_n, \Sigma_m + \Lambda_n)$$

Method - Relaxation

⊕ Node embedding   Gaussian mixture model (GMM)   **Anchor**

Definition: An anchor is a carefully selected node (item) whose name (label) is existing in the dictionary, aiming to regulate the word's GMM.

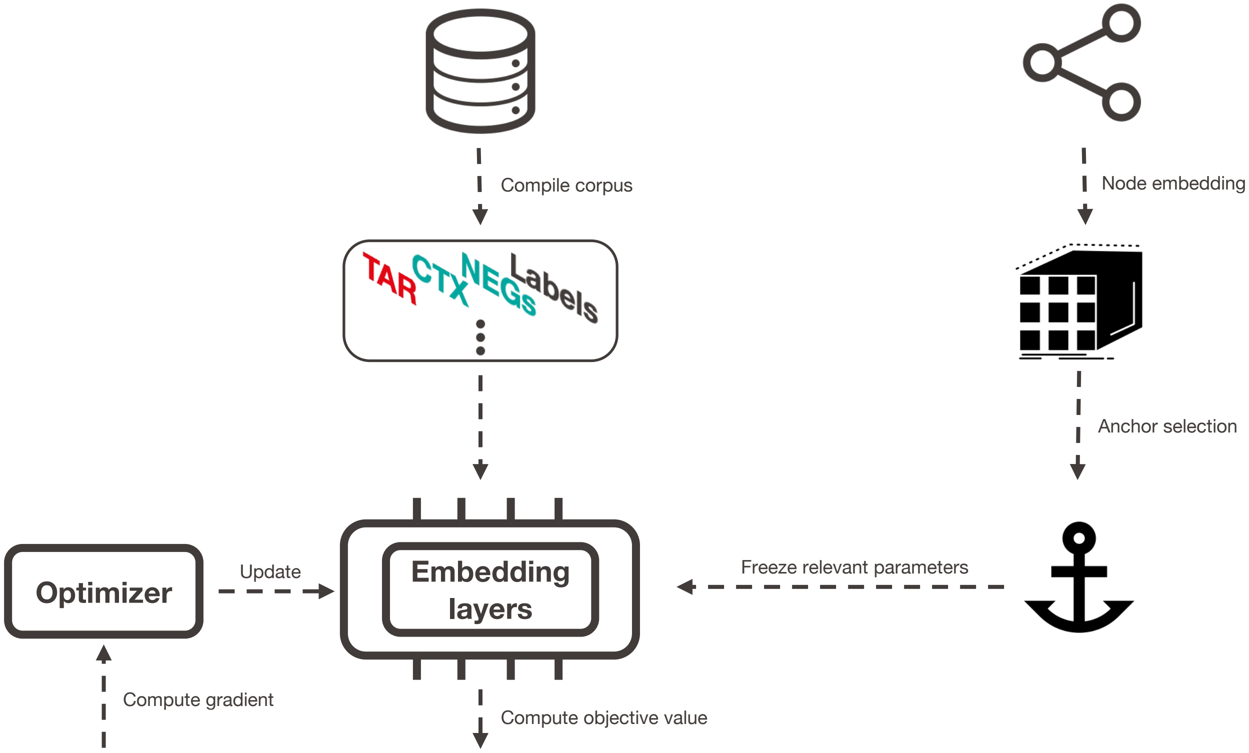
GMM for anchors is non-learnable:

$$\mu: node\ vector$$
$$\Sigma: Diagonal\ matrix\ with\ entries = 1$$

Anchor selection:

- k-means++
- Sampling with density: k-NN estimates the density
- High-degree nodes: Hubs as the anchors
- Sampling with the degree

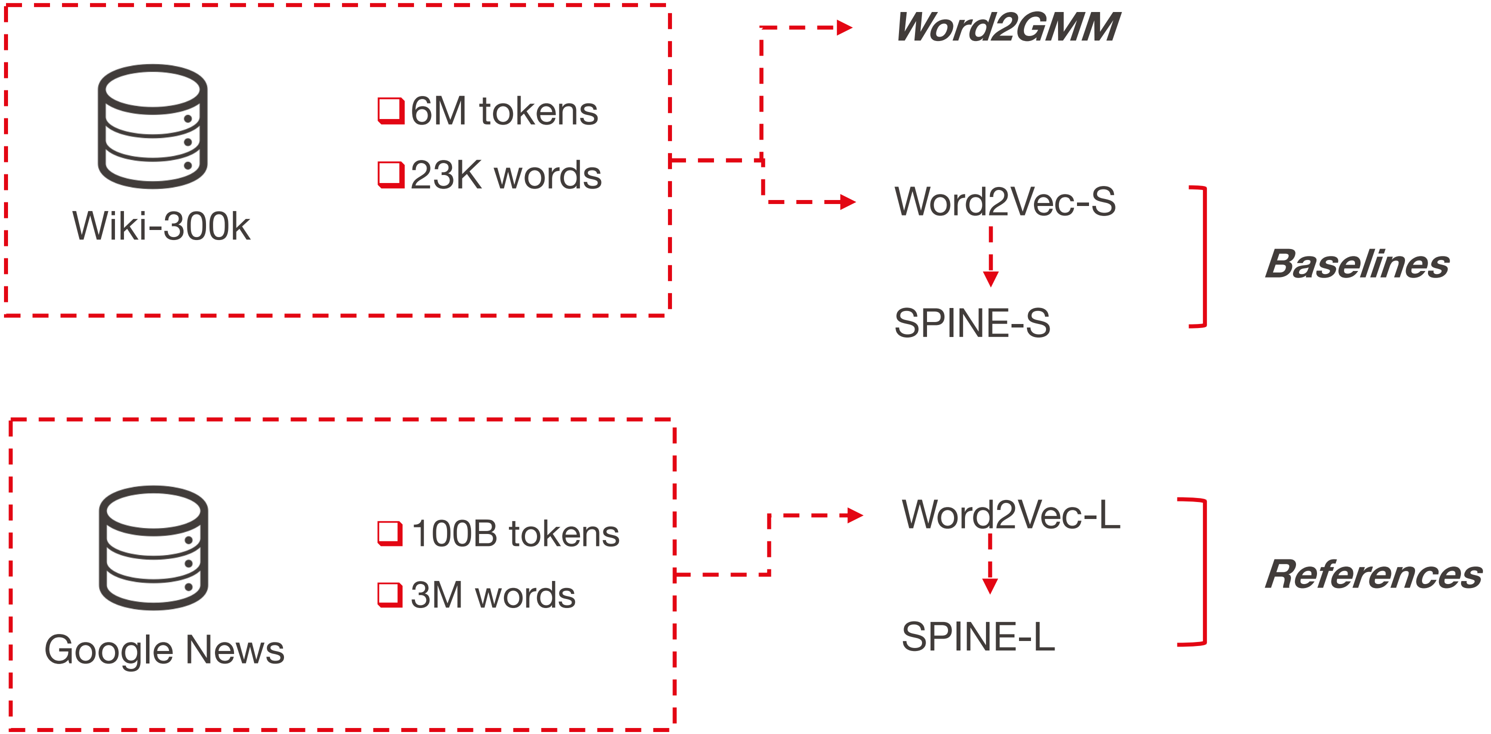
Method - Word2GMM



$$\log \sigma(-\ell_2^2(a^i, \mu^i, \Sigma, b^j, \eta^j, \Sigma)) + \sum_{k=1}^{K} \mathbb{E}_{w_k \sim p_n(w)}[\log \sigma(\ell_2^2(a^i, \mu^i, \Sigma, b^k, \eta^k, \Sigma)))]$$

https://go.epfl.ch/word2gmm

Experiment - Setup



SPINE: https://github.com/jacobdanovitch/SPINE

Experiment - Word similarity

Similarity

Word2GMM: $-l_2^2$

Word2Vec, SPINE: cosine similarity

Datasets

7947 valid pairs of words from 13 datasets.

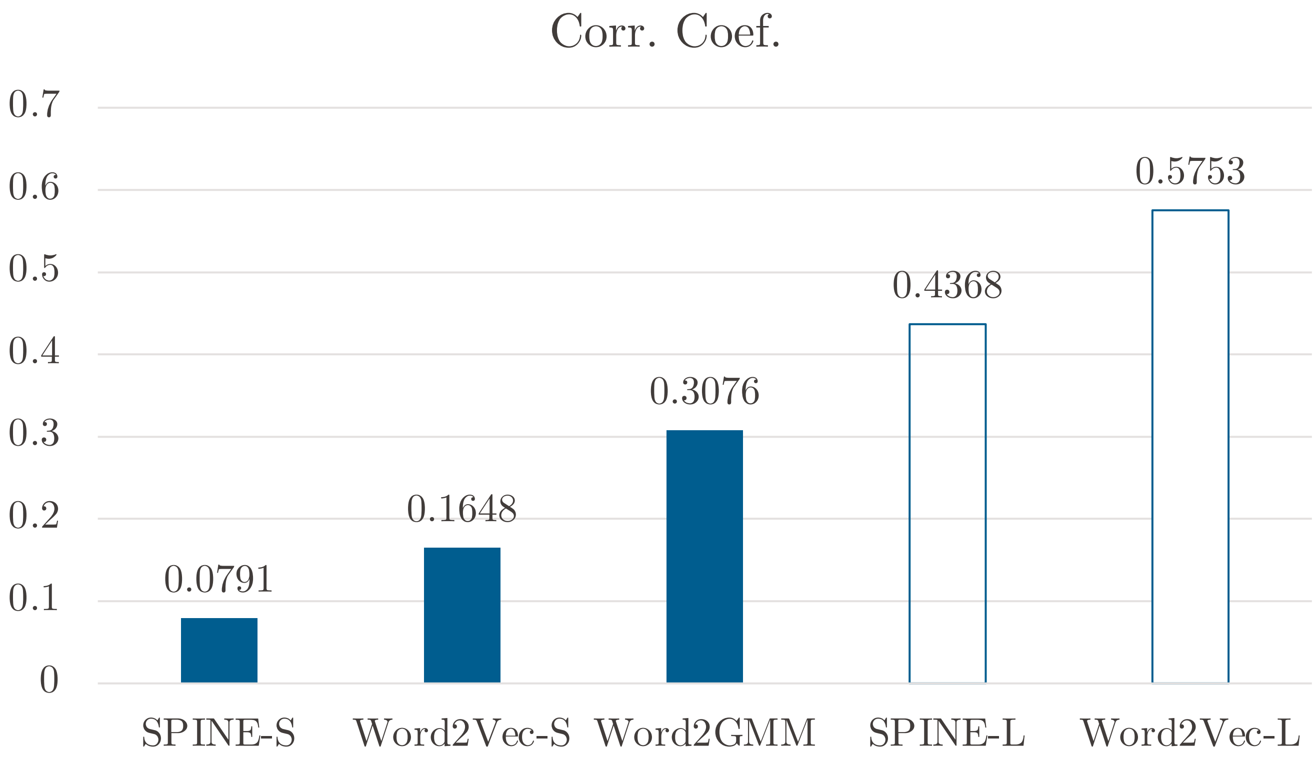
Metric

Spearman's rank correlation coefficient r_s.

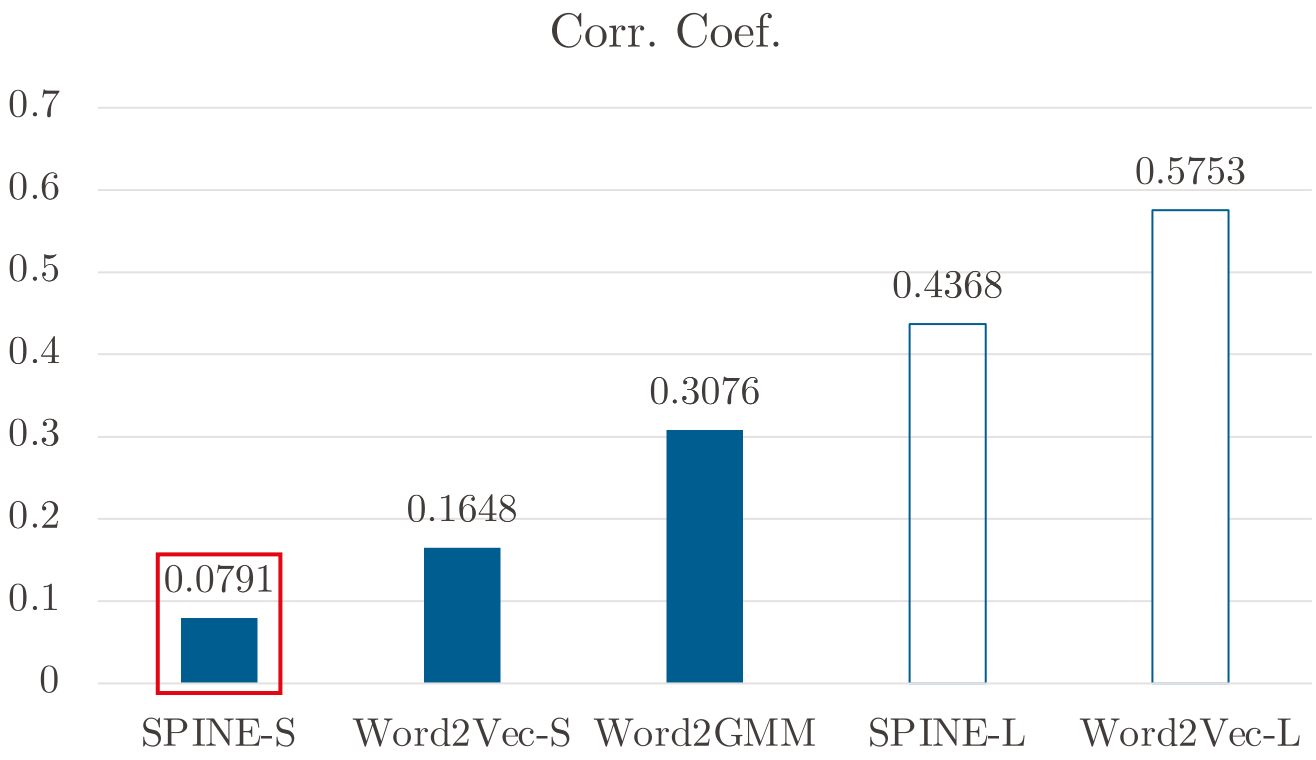
$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Weighted sum of r_s from all datasets.

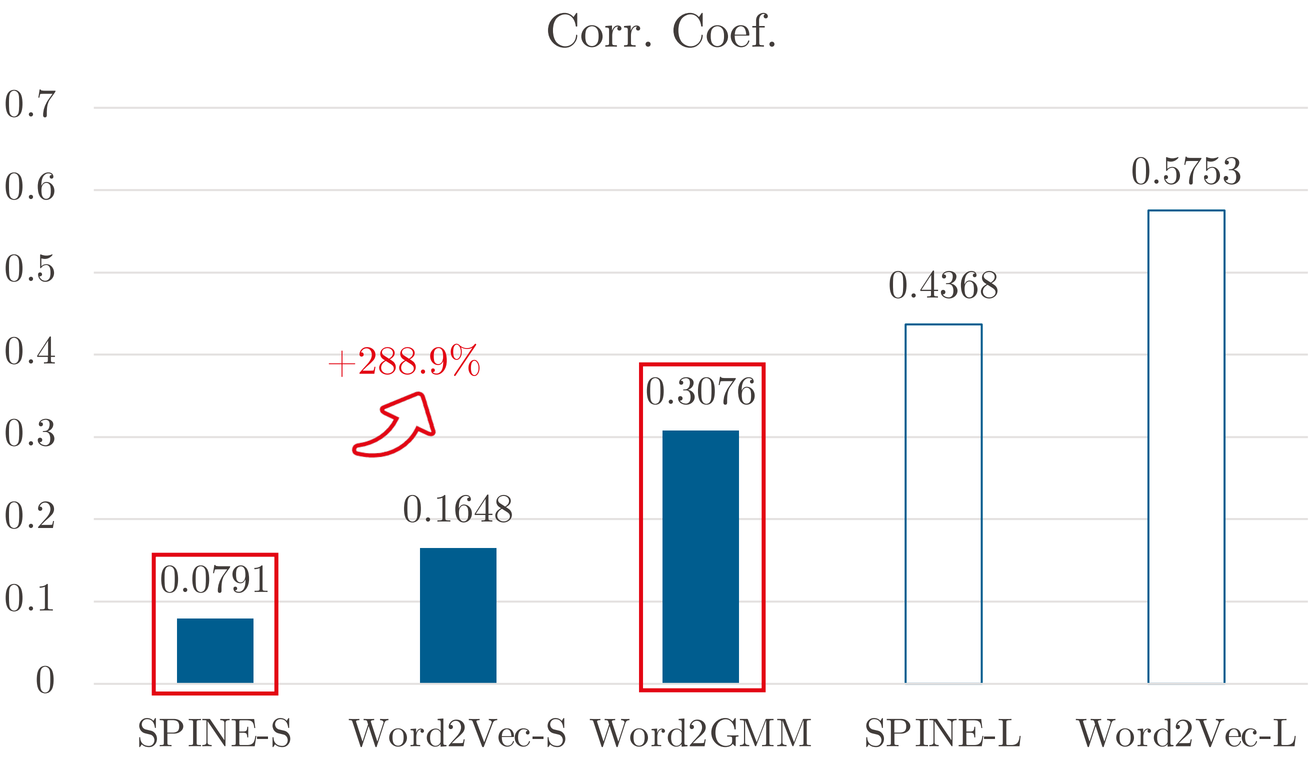
Experiment - Word similarity



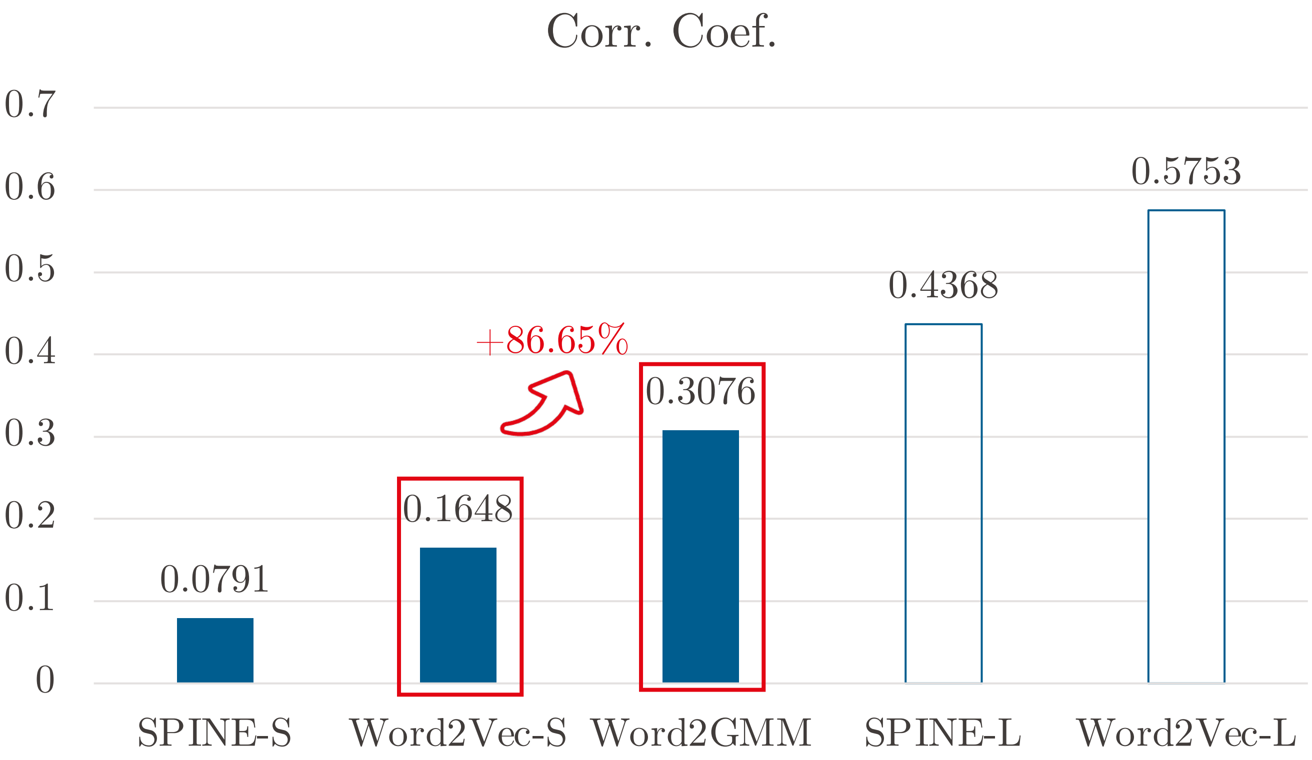
Experiment - Word similarity



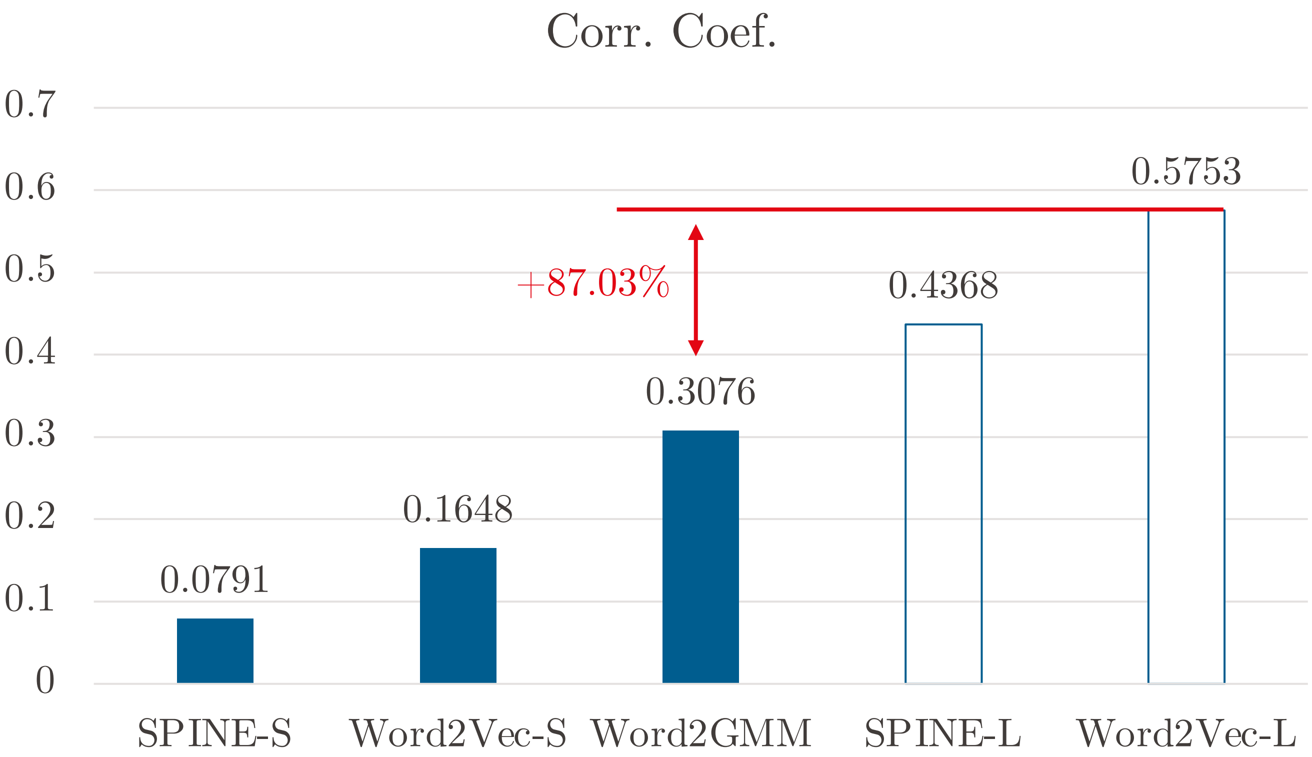
Experiment - Word similarity



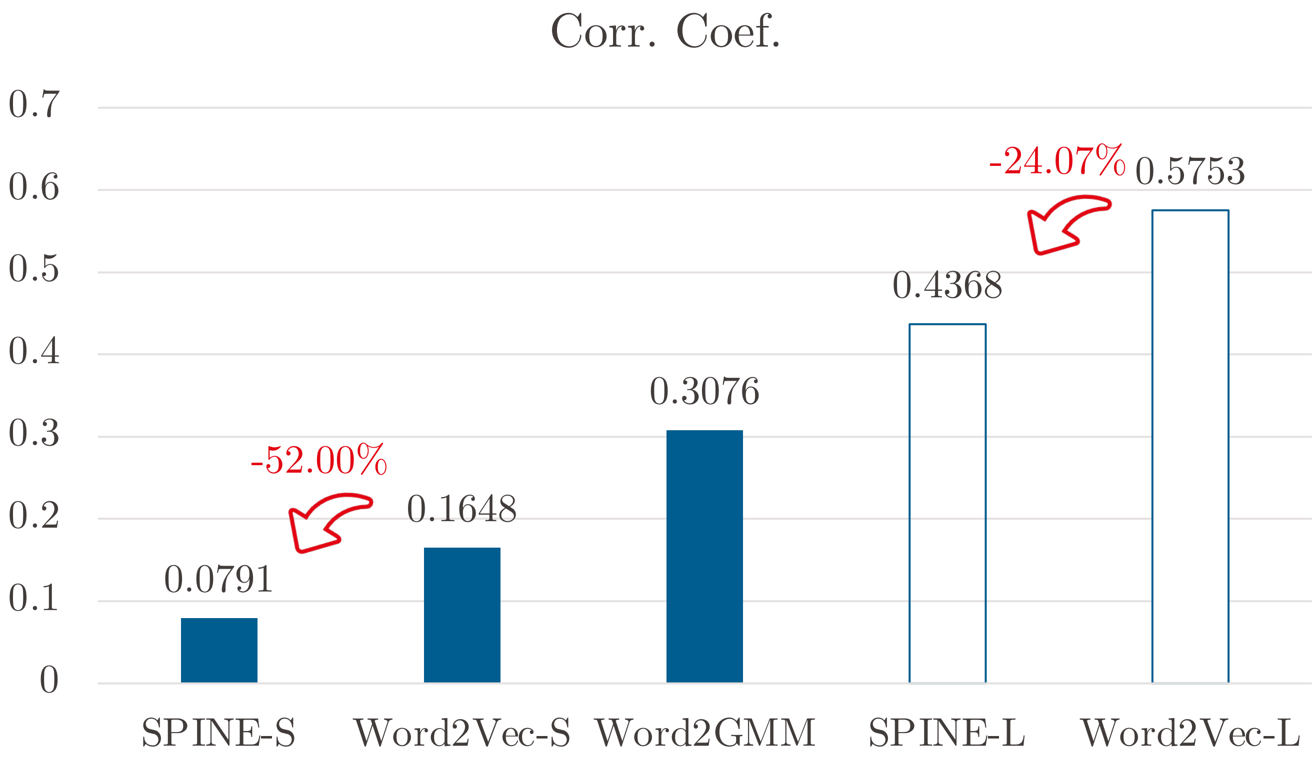
Experiment - Word similarity



Experiment - Word similarity



Experiment - Word similarity



EPFL

Experiment - Word similarity

Word closeness

EX Word2GMM Word2Vec-L

build: completion, construction, bas, infrastructure, designated.
construct, develop, built, rebuild, establish.

film: movie, films, bros, sitcoms, cartoon, starring.
movie, films, filmmaker, filmmakers, filmmaking.

social: societal, sustainability, deliberation, openness, norms.
societal, Carmeta Albarus Lindo, socio, media optimization SMO, cultural.

Experiment - Interpretability

❑ Qualitative assessment

Word2GMM: 5 top-activated nodes

Word2Vec, SPINE: 5 top-activated words

❑ Error analysis

The situation when Word2GMM fails to give interpretation.

❑ Activation pattern analogy

Activation pattern analogy with respect to word similarity.

Experiment - Interpretability

❑ **Qualitative assessment** Error analysis Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher Wilde Jonas Alexis goalkeeper	football player Argentine city football player football player position in football
government	Valencia Alameda Monaco Arenas Algeria	electoral district municipality country municipality country
water	salmon lettuce fish tea rack	fish plant aquatic animal drink gadget

EPFL

Experiment - Interpretability

Qualitative assessment Error analysis Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher	football player
	Wilde	Argentine city
	Jonas	football player
	Alexis	football player
	goalkeeper	position in football
government	Valencia	electoral district
	Alameda	municipality
	Monaco	country
	Arenas	municipality
	Algeria	country
water	salmon	fish
	lettuce	plant
	fish	aquatic animal
	tea	drink
	rack	gadget

Wentao Feng

Experiment - Interpretability

Qualitative assessment | Error analysis | Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher Wilde Jonas Alexis goalkeeper	football player Argentine city football player football player position in football
government	Valencia Alameda Monaco Arenas Algeria	electoral district municipality country municipality country
water	salmon lettuce fish tea rack	fish plant aquatic animal drink gadget

EPFL

Experiment - Interpretability

Qualitative assessment | Error analysis | Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher Wilde Jonas Alexis goalkeeper	football player Argentine city football player football player position in football
government	Valencia Alameda Monaco Arenas Algeria	electoral district municipality country municipality country
water	salmon lettuce fish tea rack	fish plant aquatic animal drink gadget

Experiment - Interpretability

- **Qualitative assessment** Error analysis Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher	football player
	Wilde	Argentine city
	Jonas	football player
	Alexis	football player
	goalkeeper	position in football
government	Valencia	electoral district
	Alameda	municipality
	Monaco	country
	Arenas	municipality
	Algeria	country
water	salmon	fish
	lettuce	plant
	fish	aquatic animal
	tea	drink
	rack	gadget

Experiment - Interpretability

- **Qualitative assessment** Error analysis Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher	football player
	Wilde	Argentine city
	Jonas	football player
	Alexis	football player
	goalkeeper	position in football
government	Valencia	electoral district
	Alameda	municipality
	Monaco	country
	Arenas	municipality
	Algeria	country
water	salmon	fish
	lettuce	plant
	fish	aquatic animal
	tea	drink
	rack	gadget

Experiment - Interpretability

Qualitative assessment Error analysis Activation pattern analogy

	Word2Vec-S	Word2Vec-L
people	bassists, litre, nc dispense, daytona	capt, astronomers lakers, nec, shootout
government	wafer, quark, ibsen ounces, eocene	jacket, consortium vaccine, coupe. cigar
water	hotter, newark, bohr modernisation, lysander	microsoft, sr, bt malaysia, jan
	SPINE-S	**SPINE-L**
people	mgm, nudity, tensile semitone, secretion	viewers, readers, listeners travelers, commuters
government	katz, bess, tampa nearing, salisbury	envoy, minister, ministers parliament, ambassador
water	tarot, repel, pepys voltaire, prematurely	dam, dams, river rivers, tributary

	Word2GMM	
	Node	**Description**
people	Schumacher Wilde Jonas Alexis goalkeeper	football player Argentine city football player football player position in football
government	Valencia Alameda Monaco Arenas Algeria	electoral district municipality country municipality country
water	salmon lettuce fish tea rack	fish plant aquatic animal drink gadget

Experiment - Interpretability

Qualitative assessment **Error analysis** Activation pattern analogy

Anchor's contextual meaning is far from Wikidata's description.

EX London (Q79348): city in Pope County, Arkansas, United States.

	Node	Description
England	garner	American town
	astros	American football team
	blind	type of bet in poker
	linebacker	position in American football
	rochester	American borough

*Full description: https://www.wikidata.org/wiki/Q79348

Experiment - Interpretability

Qualitative assessment　　Error analysis　　**Activation pattern analogy**

Activation pattern

Word2GMM: nodes with values ≥ 0.01.

SPINE: axes with values ≥ 0.01.

Word2Vec: dense vector, not applicable.

Datasets

13 datasets are ordered by ground truth score decreasingly.

Similar words: first 10%　　Dissimilar words: last 10%

Metric

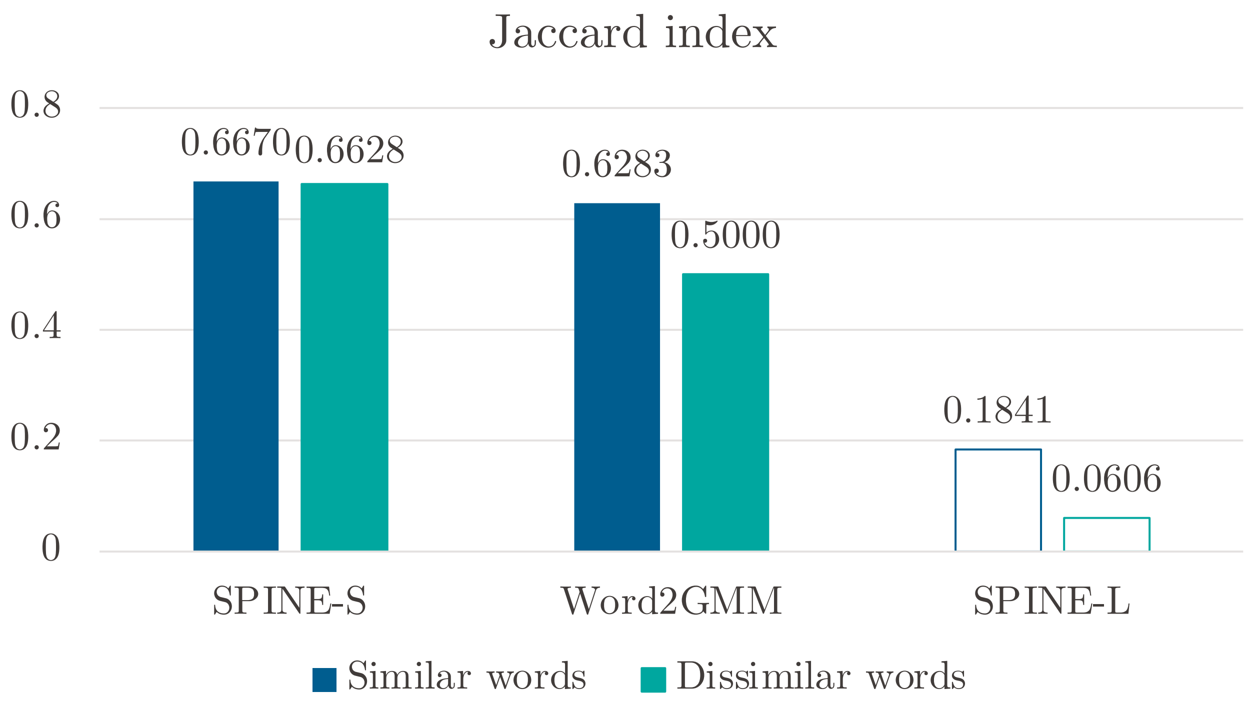
Jaccard index $J(A, B)$

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

The median of $J(A, B)$ from given set of words.

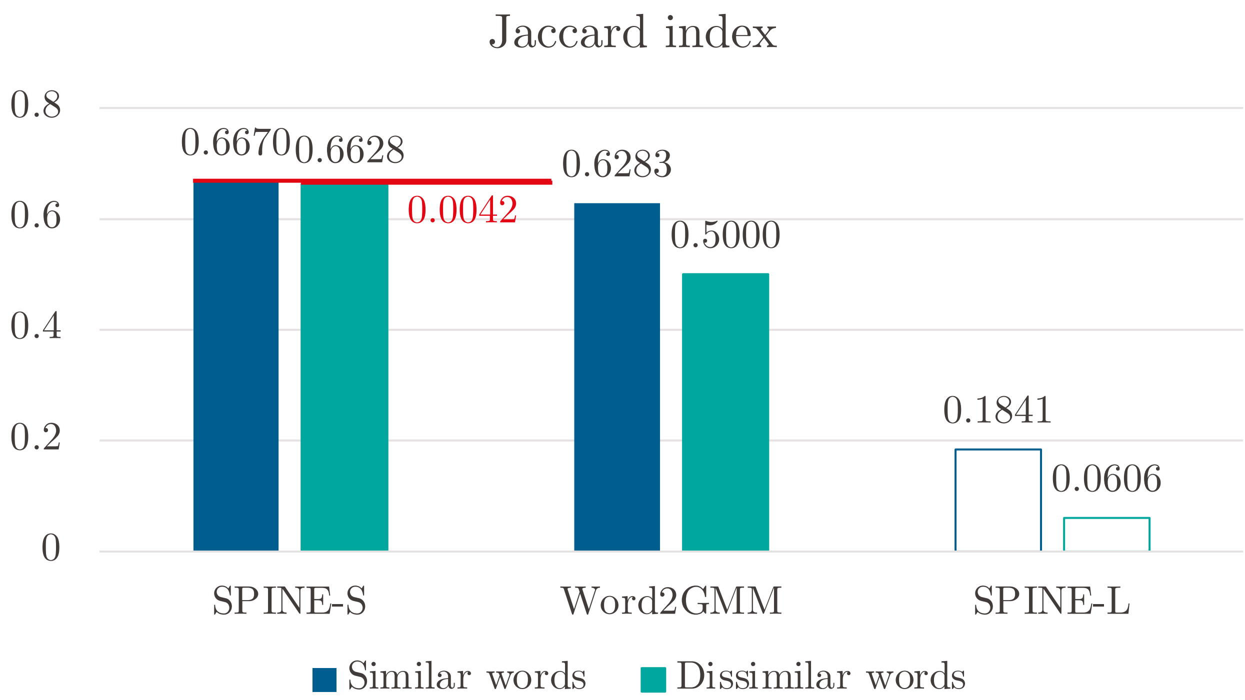
Experiment - Interpretability

❑ Qualitative assessment   Error analysis   **Activation pattern analogy**



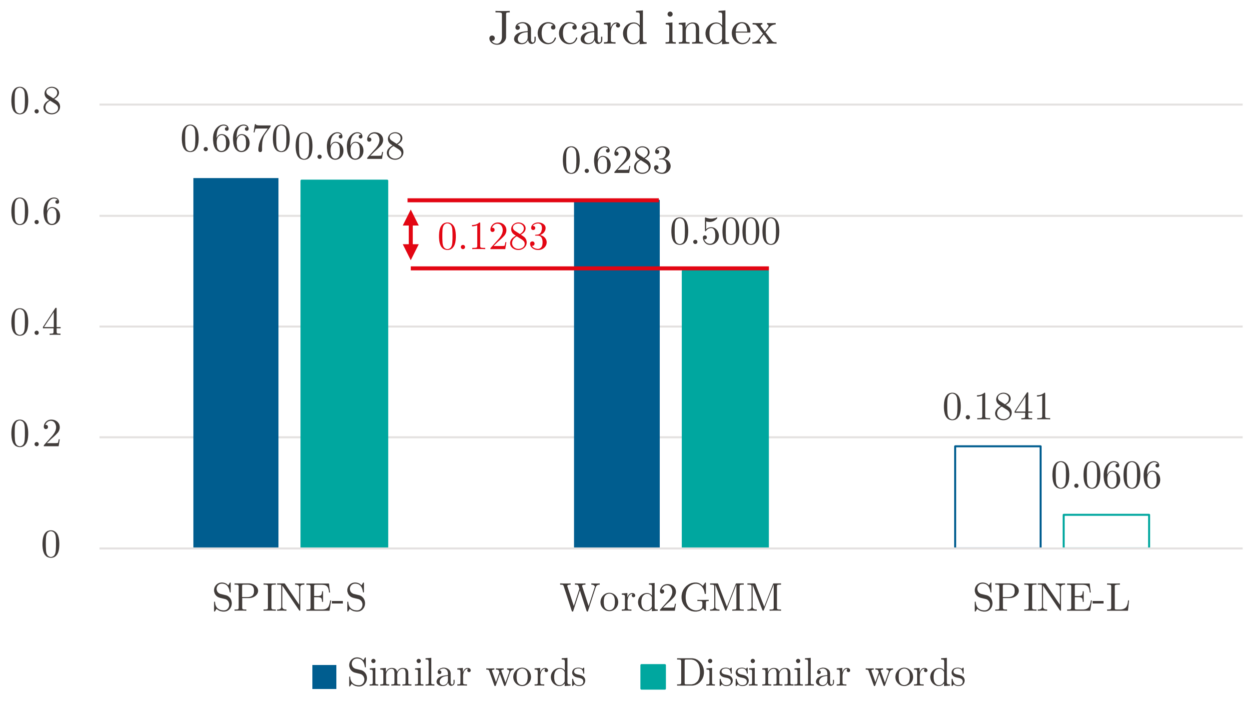
Experiment - Interpretability

- Qualitative assessment
- Error analysis
- **Activation pattern analogy**



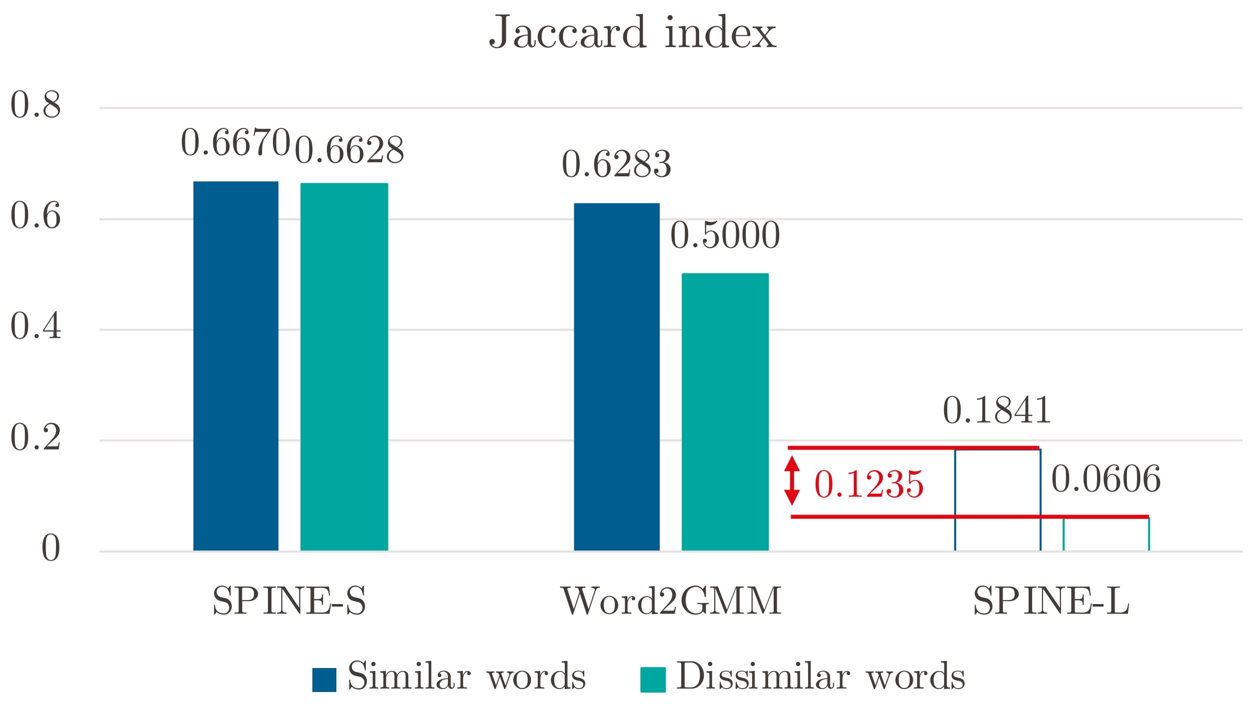
Experiment - Interpretability

❑ Qualitative assessment   Error analysis   **Activation pattern analogy**



Experiment - Interpretability

- Qualitative assessment
- Error analysis
- **Activation pattern analogy**

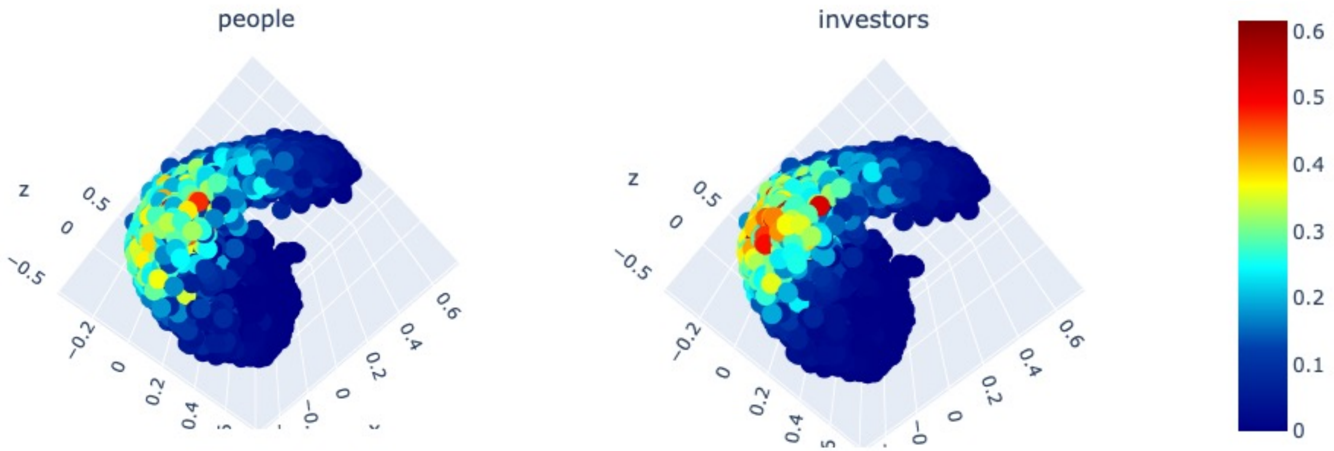


Experiment - Interpretability

Wentao Feng

- Qualitative assessment Error analysis **Activation pattern analogy**

EX



Experiment – Parameter influence

➢ Baseline

- ❑ Anchor selection: sampling with density
- ❑ Number of anchor: 128
- ❑ Number of Gaussian: 25
- ❑ Initialize covariance matrix:

  Truncated normal distribution with center $c=1$ and radius $r=0$

➢ Metric

Word similarity

Relative scores. Baseline = 1

Experiment – Parameter influence

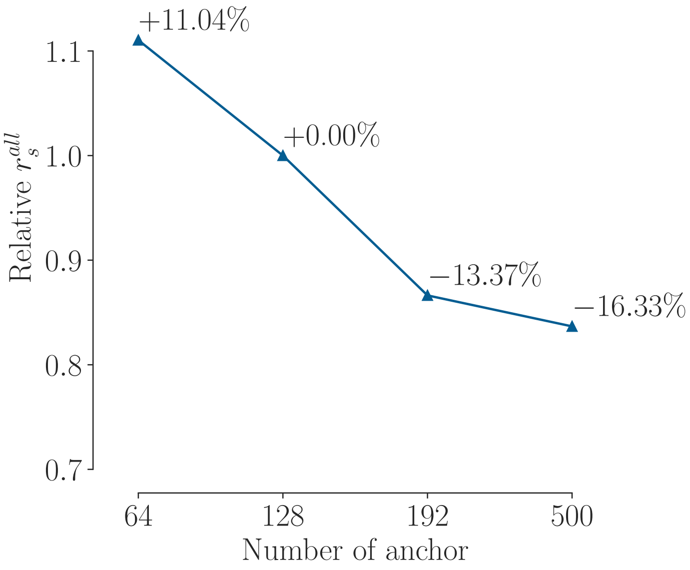
Anchor The number of Gaussian Covariance matrix

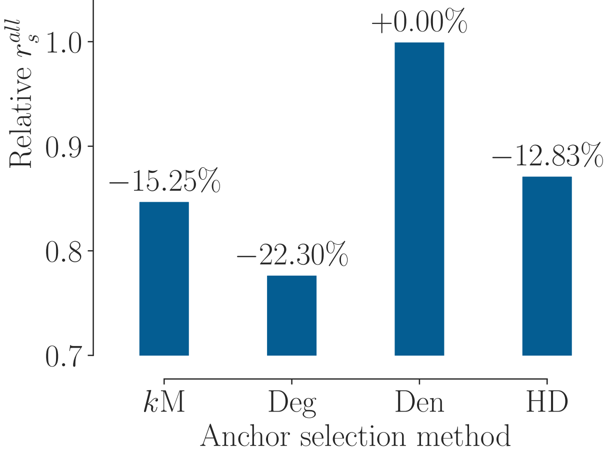
Experiment – Parameter influence

Anchor The number of Gaussian Covariance matrix

Experiment – Parameter influence

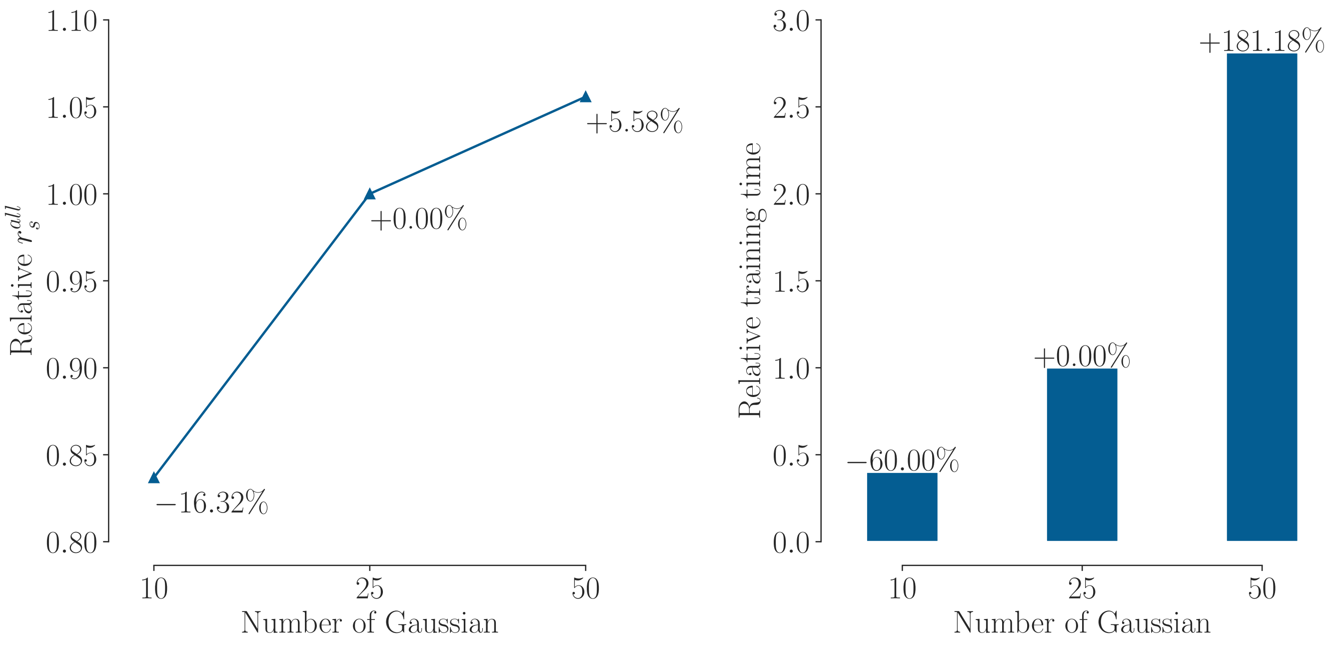
Anchor The number of Gaussian Covariance matrix





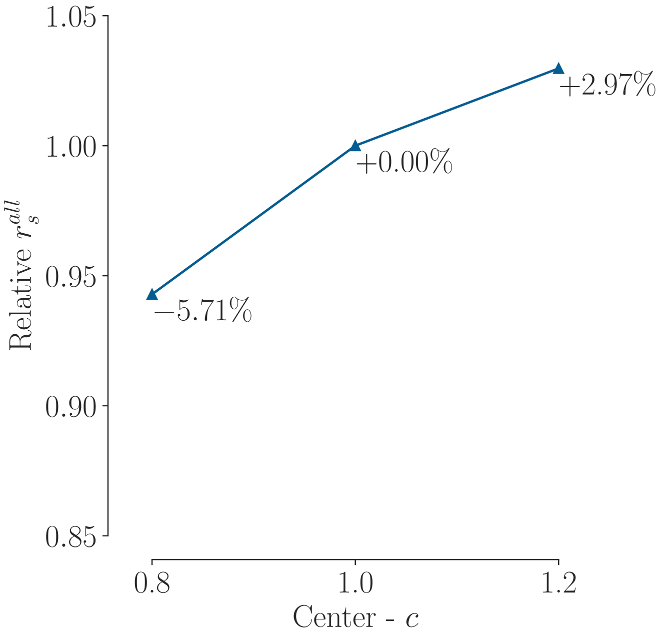
Experiment – Parameter influence

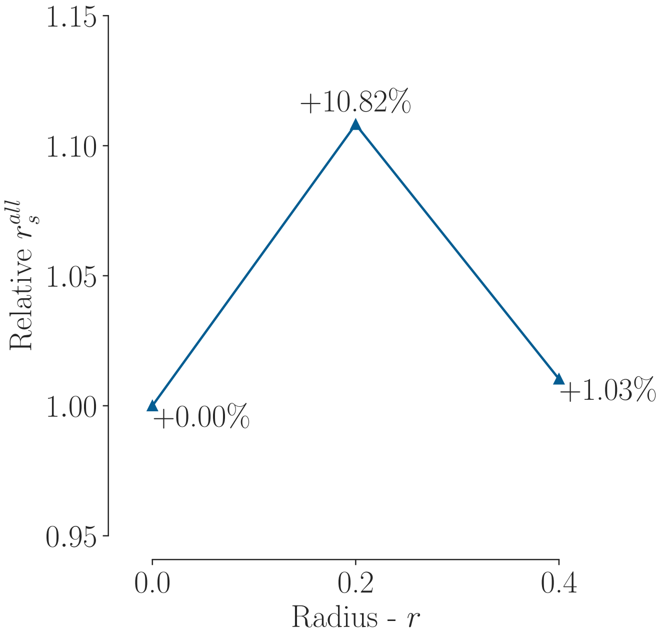
Anchor **The number of Gaussian** Covariance matrix



Experiment – Parameter influence

Anchor The number of Gaussian **Covariance matrix**





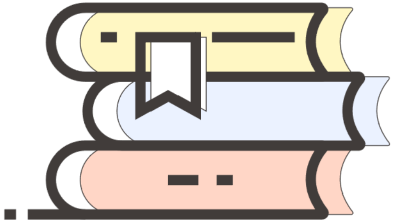
Experiment - Discussion

Pros

- Joint learning of two goals
- High-level interpretation
- Good performance on small dataset
- Large capacity of interpretability

Cons

- Sensitive to the anchors
- A little less efficient than Word2Vec



Conclusion

Conclusion - Contribution

- **Word2GMM**

  New interpretable text representations architecture.

  A novel and efficient way to use existing knowledge.

- **Evaluation**

  Comparison with classical methods.

  Analysis of interpretability with the state-of-the-art.

  Comprehension of parameter's influence.

Conclusion - Future work

Four aspects need improvement:

- ❑ Anchor selection
- ❑ Parameter fine-tuning
- ❑ A large training corpus
- ❑ Quantitative evaluation on interpretability

Thank you for your listening!

Reference

- Icons: https://www.iconfinder.com/
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." In Advances in neural information processing systems, pp. 3111-3119. 2013.
- Faruqui, Manaal, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah Smith. "Sparse overcomplete word vector representations." arXiv preprint arXiv:1506.02004 (2015).
- Subramanian, Anant, Danish Pruthi, Harsh Jhamtani, Taylor Berg-Kirkpatrick, and Eduard Hovy. "Spine: Sparse interpretable neural embeddings." In Thirty-Second AAAI Conference on Artificial Intelligence. 2018.