Multi-Relational Latent Semantic Analysis

Kai-Wei Chang

Joint work with
Scott Wen-tau Yih, Chris Meek

Microsoft Research

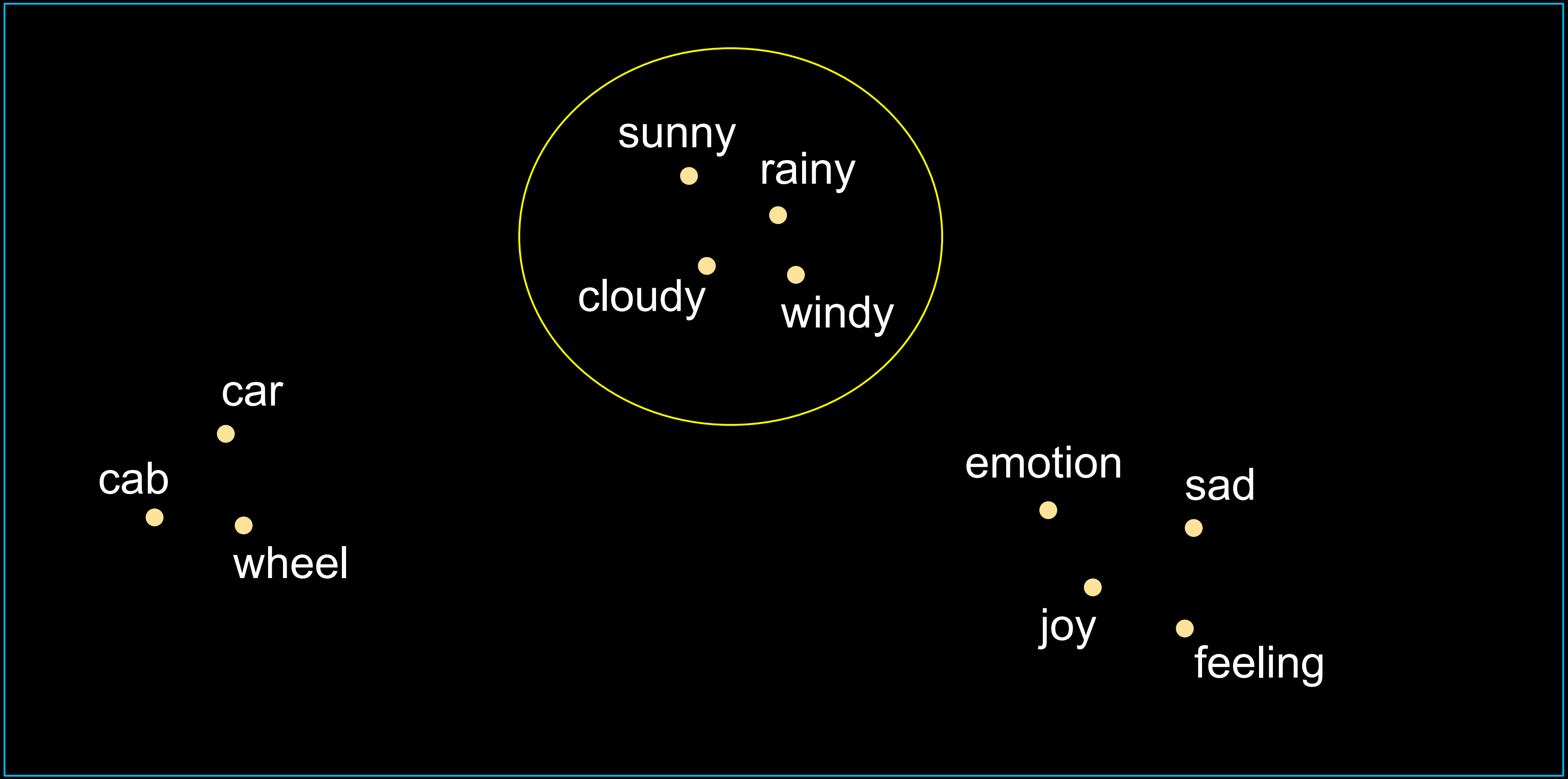
Natural Language Understanding

- Build an intelligent system that can interact with human using natural language
- Research challenge
  - Meaning representation of text
  - Support useful inferential tasks
- Semantic word representation is the foundation
  - Language is compositional
  - Word is the basic semantic unit

Continuous Semantic Representations

- A lot of popular methods for creating word vectors!
  - Vector Space Model [Salton & McGill 83]
  - Latent Semantic Analysis [Deerwester+ 90]
  - Latent Dirichlet Allocation [Blei+ 01]
  - Deep Neural Networks [Collobert & Weston 08]
- Encode term co-occurrence information
- Measure semantic similarity well

Continuous Semantic Representations



sunny

rainy

cloudy

windy

car

cab

wheel

emotion

sad

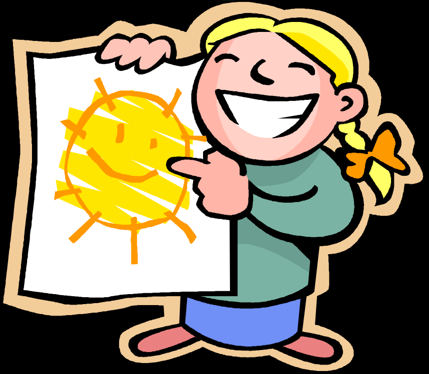
joy

feeling

Semantics Needs More Than Similarity

Tomorrow will be **rainy**.

Tomorrow will be **sunny**.



~~*similar*(rainy, sunny)?~~

antonym(rainy, sunny)?

Leverage Linguistic Resources

- Can't we just use the existing linguistic resources?
  - Knowledge in these resources is never complete
  - Often lack of degree of relations
- Create a continuous semantic representation that
  - Leverages existing rich linguistic resources
  - Discovers new relations
  - Enables us to measure the degree of multiple relations (not just similarity)

Roadmap

- Introduction
- Background
  - Latent Semantic Analysis (LSA)
  - Polarity Inducing LSA (PILSA)
- Multi-Relational Latent Semantic Analysis (MRLSA)
  - Encoding multi-relational data in a tensor
  - Tensor decomposition & measuring degree of a relation
- Experiments

Roadmap

- Introduction
- Background
  - Latent Semantic Analysis (LSA)
  - Polarity Inducing LSA (PILSA)
- Multi-Relational Latent Semantic Analysis (MRLSA)
  - Encoding multi-relational data in a tensor
  - Tensor decomposition & measuring degree of a relation
- Experiments

Latent Semantic Analysis [Deerwester+ 1990]

- Data representation
  - Encode single-relational data in a matrix
    - Co-occurrence (e.g., from a general corpus)
    - Synonyms (e.g., from a thesaurus)
- Factorization
  - Apply SVD to the matrix to find latent components
- Measuring degree of relation
  - Cosine of latent vectors

Encode Synonyms in Matrix

- Input: Synonyms from a thesaurus
- Joyfulness: joy, gladden
- Sad: sorrow, sadden

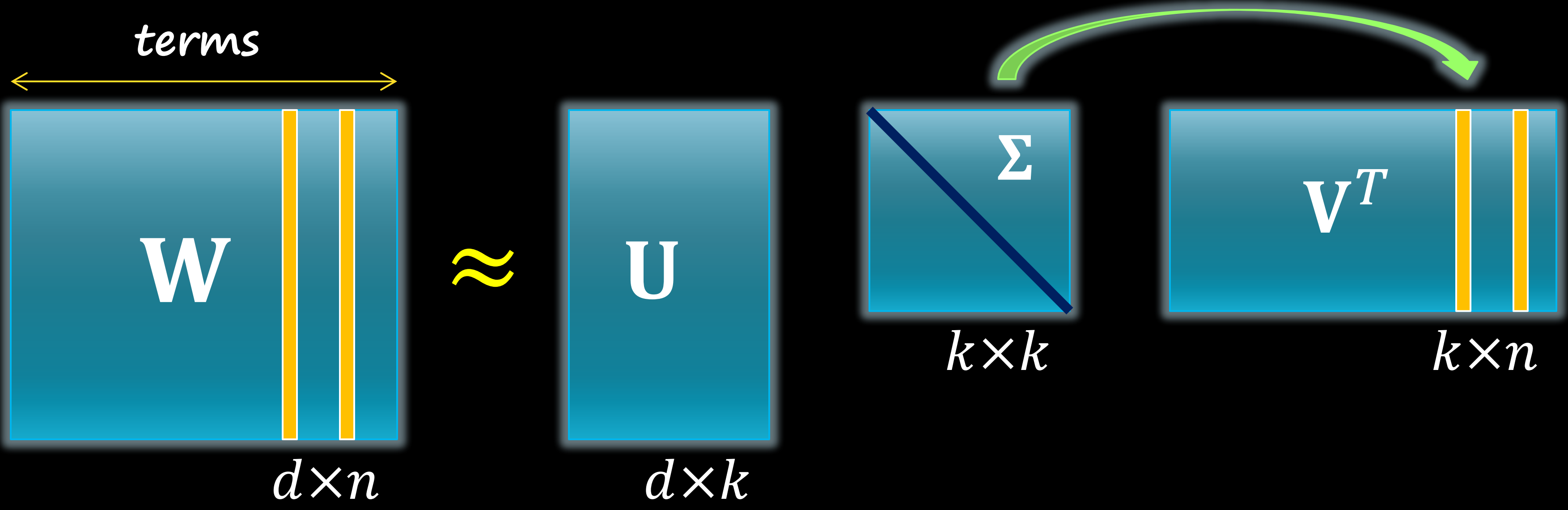
Target word: row-vector

Term: column-vector

	joy	gladden	sorrow	sadden	goodwill
Group 1: "joyfulness"	1	1	0	0	0
Group 2: "sad"	0	0	1	1	0
Group 3: "affection"	0	0	0	0	1

Cosine Score

Mapping to Latent Space via SVD



- SVD generalizes the original data
  - Uncovers relationships not explicit in the thesaurus
  - Term vectors projected to k-dim latent space
- Word similarity: cosine of two column vectors in ΣV^T

Problem: Handling Two Opposite Relations

Synonyms & Antonyms

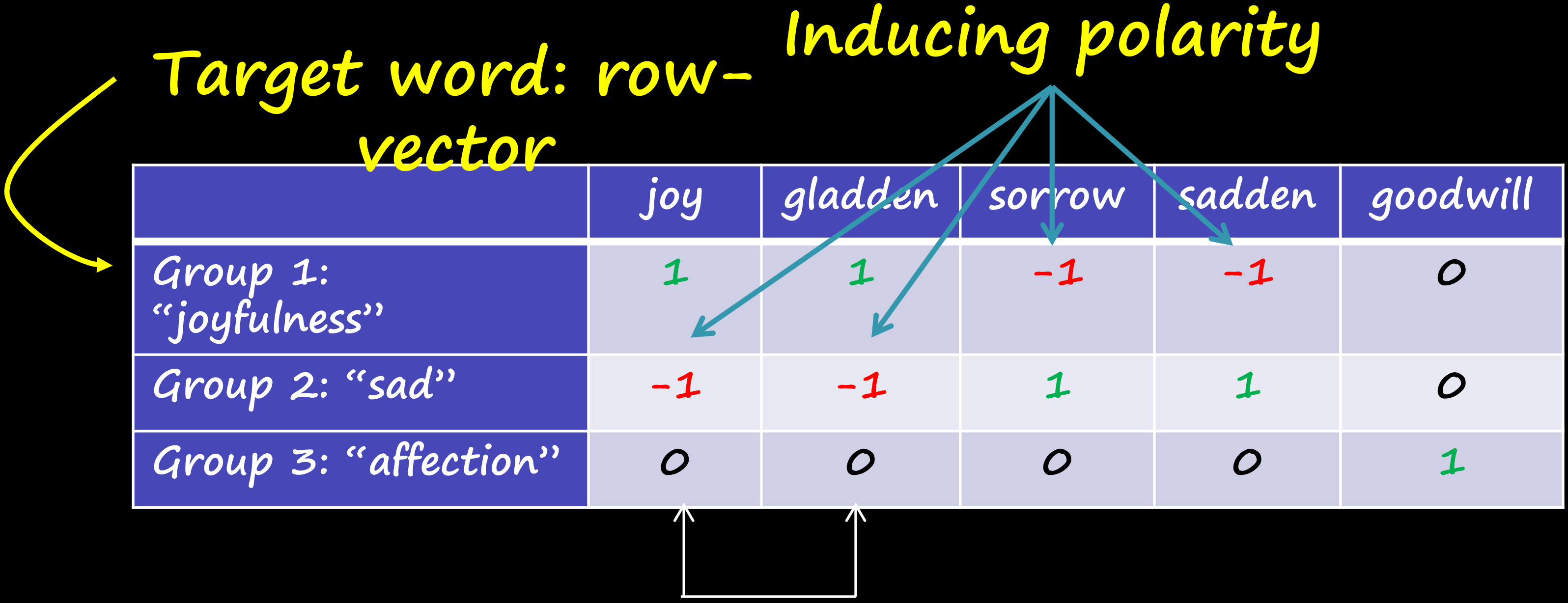
- LSA cannot distinguish antonyms [Landauer 2002]
  - "Distinguishing synonyms and antonyms is still perceived as a difficult open problem." [Poon & Domingos 09]

Polarity Inducing LSA [Yih, Zweig, Platt 2012]

- Data representation
  - Encode two opposite relations in a matrix using "polarity"
    - Synonyms & antonyms (e.g., from a thesaurus)
- Factorization
  - Apply SVD to the matrix to find latent components
- Measuring degree of relation
  - Cosine of latent vectors

Encode Synonyms & Antonyms in Matrix

- Joyfulness: joy, gladden; sorrow, sadden
- Sad: sorrow, sadden; joy, gladden

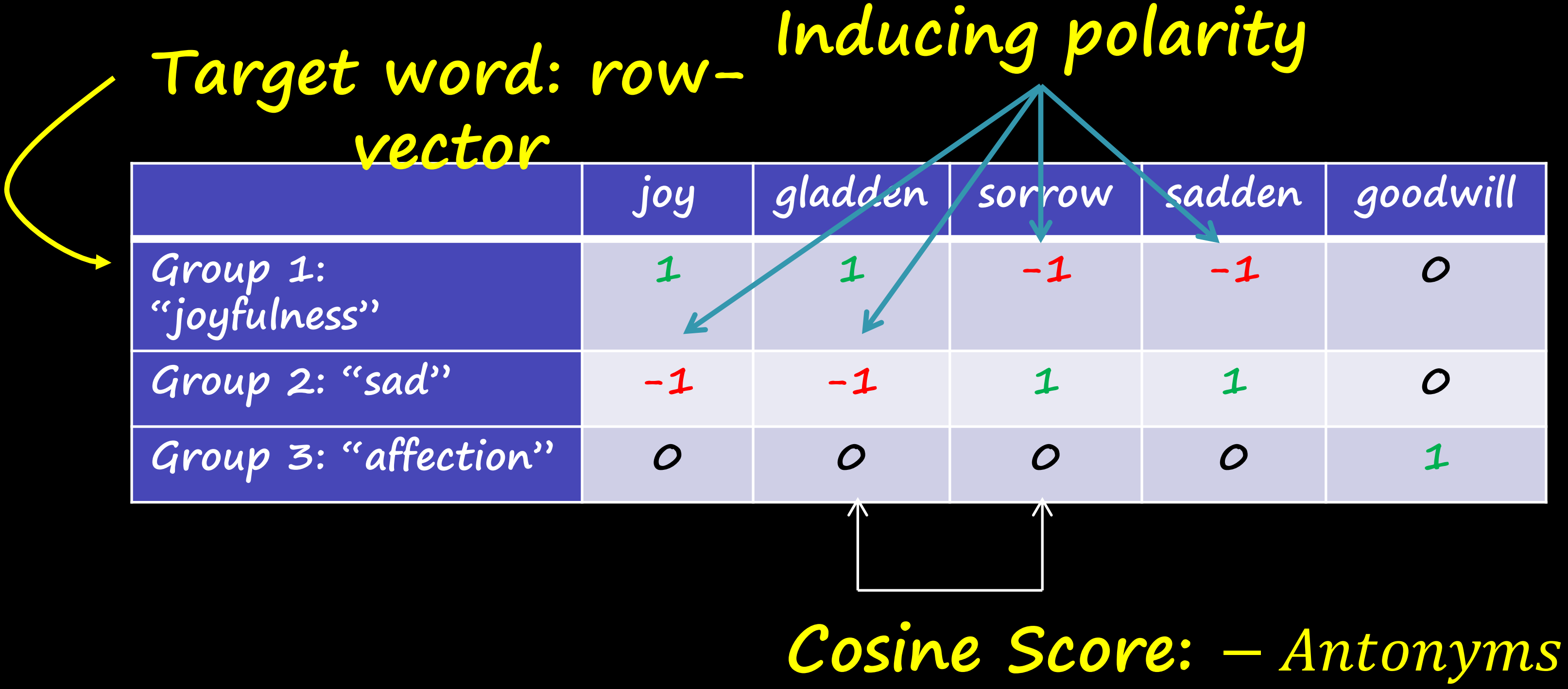


	joy	gladden	sorrow	sadden	goodwill
Group 1: "joyfulness"	1	1	-1	-1	0
Group 2: "sad"	-1	-1	1	1	0
Group 3: "affection"	0	0	0	0	1

Cosine Score: $+\mathit{Synonyms}$

Encode Synonyms & Antonyms in Matrix

- Joyfulness: joy, gladden; sorrow, sadden
- Sad: sorrow, sadden; joy, gladden



Inducing polarity

Target word: row-vector

	joy	gladden	sorrow	sadden	goodwill
Group 1: "joyfulness"	1	1	-1	-1	0
Group 2: "sad"	-1	-1	1	1	0
Group 3: "affection"	0	0	0	0	1

Cosine Score: $-$ *Antonyms*

Problem: How to Handle More Relations?

- Limitation of the matrix representation
  - Eac relation bet
  - Tw trick

Encode multiple relations in a 3-way tensor (3-dim array)!

- Encoding other binary relations
  - Is-A (hyponym) – ostrich *is a* bird
  - Part-whole – engine is a *part of* car

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Encode Multiple Relations in Tensor

- Represent word relations using a tensor
  - Each slice encodes a relation between **terms** and **target words.**

	joy	gladden	sadden	feeling
joyfulness	1	1	0	0
gladden	1	1	0	0
sad	0	0	1	0
anger	0	0	0	0

Synonym layer

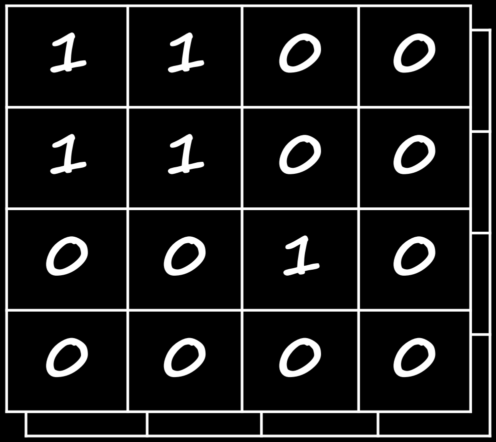
	joy	gladden	sadden	feeling
joyfulness	0	0	0	0
gladden	0	0	1	0
sad	1	0	0	0
anger	0	0	0	0

Antonym layer

Construct a tensor with two slices

Encode Multiple Relations in Tensor

- Can encode multiple relations in the tensor



1	1	0	0
1	1	0	0
0	0	1	0
0	0	0	0

	joy	gladden	sadden	feeling
joyfulness	0	0	0	1
gladden	0	0	0	0
sad	0	0	0	1
anger	0	0	0	1

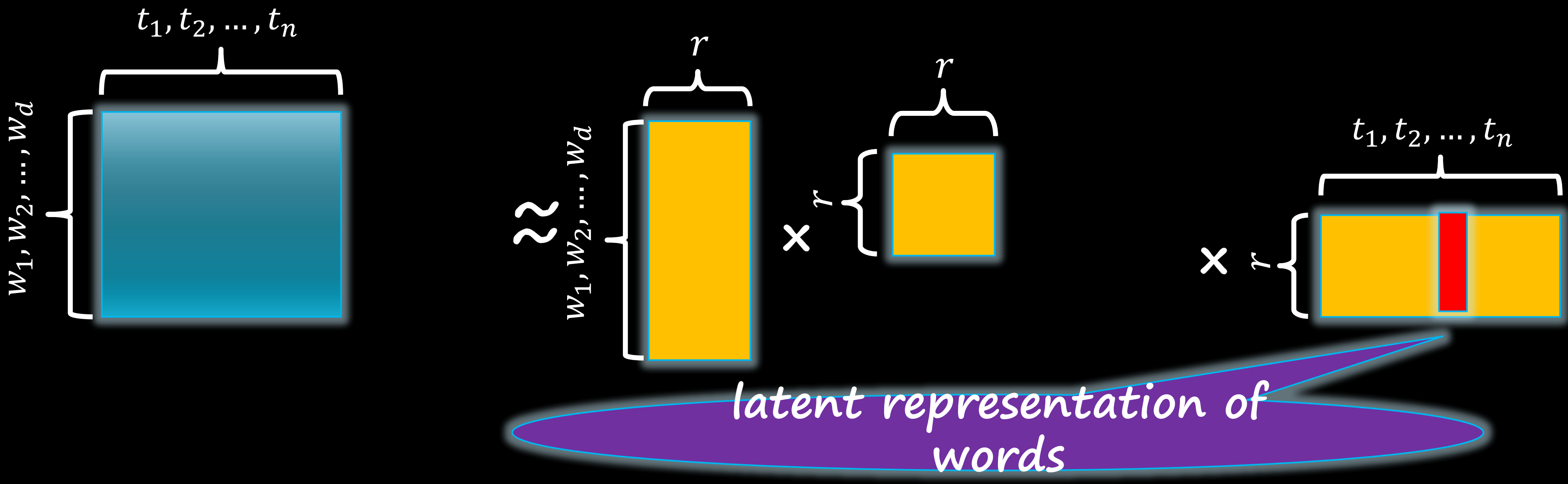
Hyponym layer

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Tensor Decomposition – Analogy to SVD

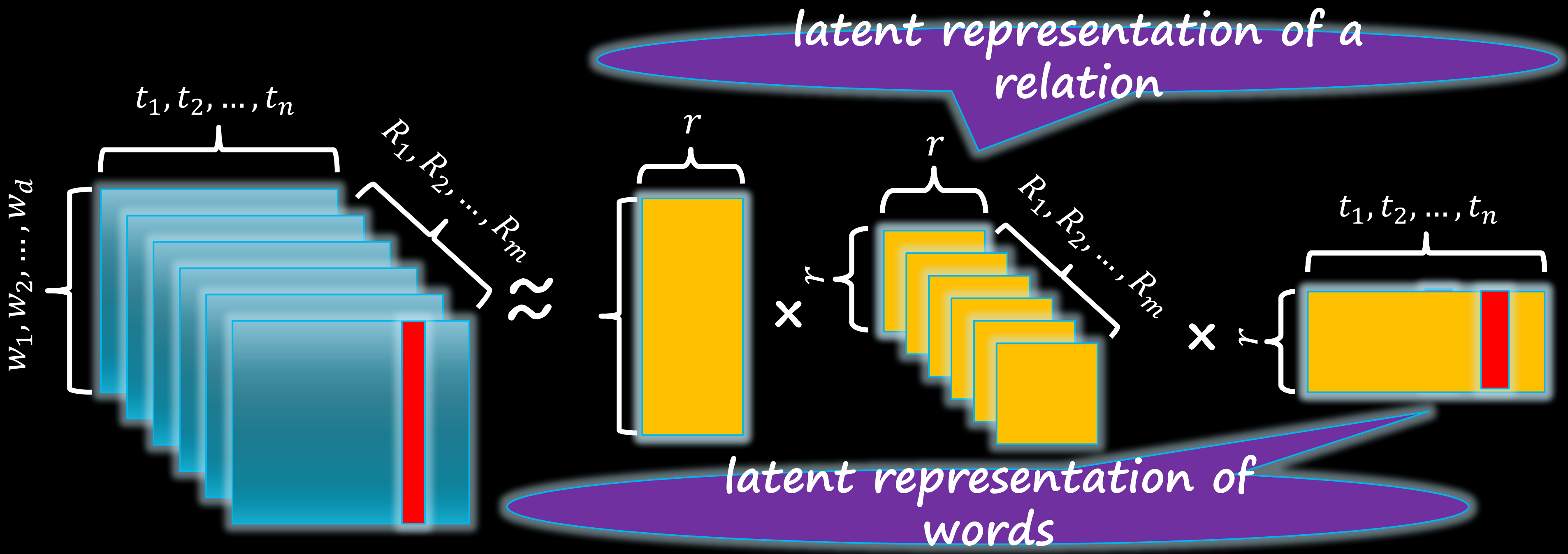
- Derive a low-rank approximation to generalize the data and to discover unseen relations
- Apply Tucker decomposition and reformulate the results



latent representation of words

Tensor Decomposition – Analogy to SVD

- Derive a low-rank approximation to generalize the data and to discover unseen relations
- Apply Tucker decomposition and reformulate the results



latent representation of a relation

latent representation of words

Multi-Relational LSA

- Data representation
  - Encode multiple relations in a tensor
    - Synonyms, antonyms, hyponyms (is-a), … (e.g., from a linguistic knowledge base)
- Factorization
  - Apply tensor decomposition to the tensor to find latent components
- Measuring degree of relation
  - Cosine of latent vectors after projection

Measure Degree of Relation

- Similarity
  - Cosine of the latent vectors
- Other relation (both symmetric and asymmetric)
  - Take the latent matrix of the *pivot* relation (synonym)
  - Take the latent matrix of the relation
  - Cosine of the latent vectors after projection

Measure Degree of Relation

Raw Representation

- $ant(\text{joy}, \text{sadden}) = \cos\left(\underline{\boldsymbol{w}_{:,\text{joy},syn}}, \underline{\boldsymbol{w}_{:,\text{sadden},ant}}\right)$

	joy	gladden	sadden	felling
joyfulness	1	1	0	0
gladden	1	1	0	0
sad	0	0	1	0
anger	0	0	0	0

Synonym layer

	joy	gladden	sadden	felling
joyfulness	0	0	0	0
gladden	0	0	1	0
sad	1	0	0	0
anger	0	0	0	0

Antonym layer

Measure Degree of Relation

Raw Representation

- $ant(\text{joy}, \text{sadden}) = \cos\left(\underline{\boldsymbol{w}_{:,\text{joy},syn}}, \underline{\boldsymbol{w}_{:,\text{sadden},ant}}\right)$

	joy	gladden	sadden	felling
joyfulness	1	1	0	0
gladden	1	1	0	0
sad	0	0	1	0
anger	0	0	0	0

Synonym layer

	joy	gladden	sadden	felling
joyfulness	0	0	0	0
gladden	0	0	1	0
sad	1	0	0	0
anger	0	0	0	0

Antonym layer

Estimate the Degree of a Relation

Raw Representation

- $Hyper(\text{joy}, \text{feeling}) = \cos\left(\underline{\boldsymbol{W}_{:,\text{joy},syn}}, \underline{\boldsymbol{W}_{:,\text{feeling},hyper}}\right)$

	joy	gladden	sadden	felling
joyfulness	1	0	0	0
gladden	1	1	0	0
sad	0	0	1	0
anger	0	0	0	0

Synonym layer

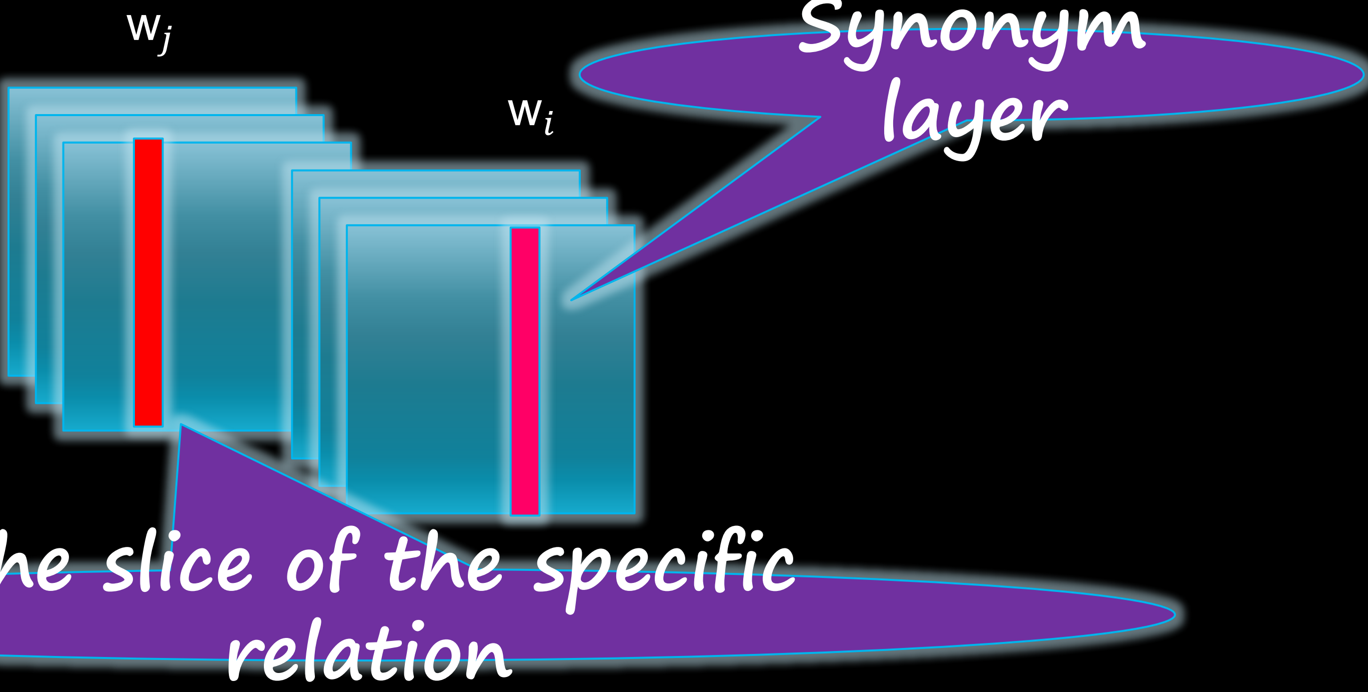
	joy	gladden	sadden	feeling
joyfulness	0	0	0	1
gladden	0	0	0	0
sad	0	0	0	1
anger	0	0	0	1

Hypernym layer

Measure Degree of Relation

Raw Representation

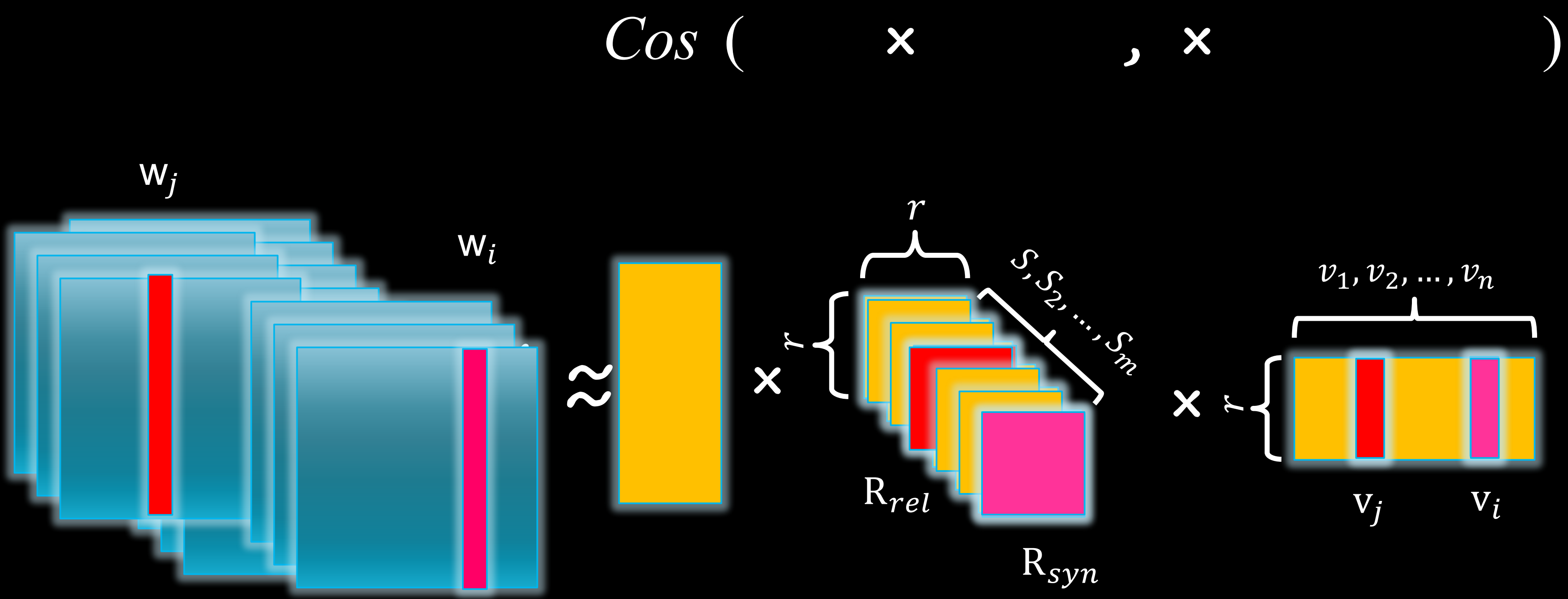
- $rel(\mathrm{w}_i, \mathrm{w}_j) = \cos\left(W_{:,\mathrm{w}_{i,syn}}, W_{:,\mathrm{w}_{j,rel}}\right)$



Measure Degree of Relation

Latent Representation

- $rel(\mathrm{w}_i, \mathrm{w}_j) = \cos\left(\boldsymbol{S}_{:,:,syn}\mathbf{V}_{i,:}^T, \boldsymbol{S}_{:,:,rel}\mathbf{V}_{j,:}^T\right)$



Roadmap

- Introduction
- Background
  - Latent Semantic Analysis (LSA)
  - Polarity Inducing LSA (PILSA)
- Multi-Relational Latent Semantic Analysis (MRLSA)
  - Encoding multi-relational data in a tensor
  - Tensor decomposition & measuring degree of a relation
- Experiments

Experiment: Data for Building MRLSA Model

- Encarta Thesaurus
  - Record synonyms and antonyms of target words
  - Vocabulary of 50k terms and 47k target words
- WordNet
  - Has synonym, antonym, hyponym, hypernym relations
  - Vocabulary of 149k terms and 117k target words
- Goals:
  - MRLSA generalizes LSA to model multiple relations

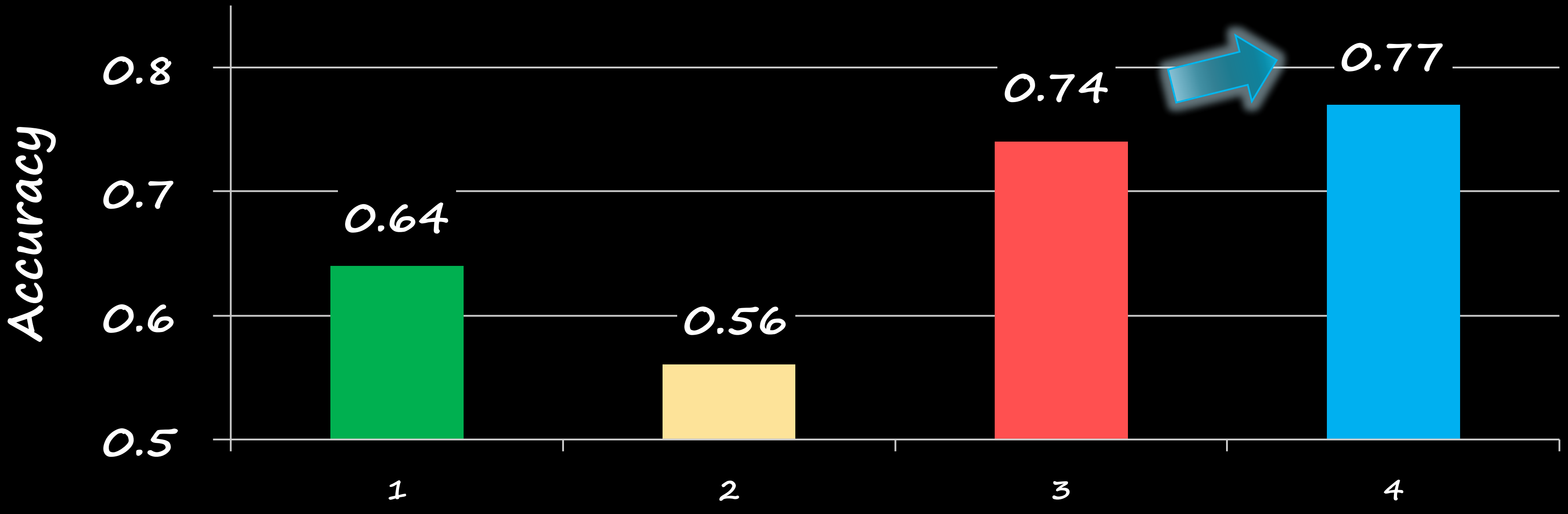
Example Antonyms Output by MRLSA

Target	High Score Words
inanimate	alive, living, bodily, in-the-flesh, incarnate
alleviate	exacerbate, make-worse, in-flame, amplify, stir-up
relish	detest, abhor, abominate, despise, loathe

** Words in blue are antonyms listed in the Encarta thesaurus.*

Results – GRE Antonym Test

- Task: GRE closest-opposite questions
  - Which is the closest opposite of *adulterate*?
    (a) renounce (b) forbid (c) purify (d) criticize (e) correct

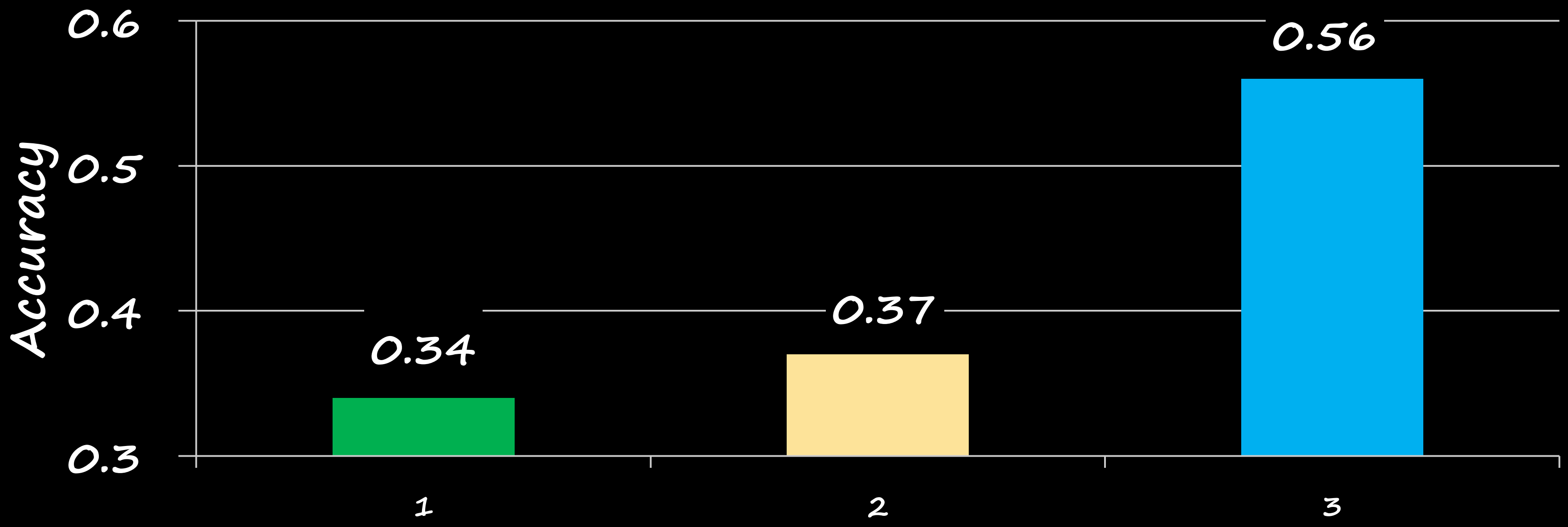


Example Hyponyms Output by MRLSA

Target	High Score Words
bird	ostrich, gamecock, nighthawk, amazon, parrot
automobile	minivan, wagon, taxi, minicab, gypsy cab
vegetable	buttercrunch, yellow turnip, romaine, chipotle, chilli

Results – Relational Similarity (SemEval-2012)

- Task: Class-Inclusion Relation (Y is-a kind of X)
  - Most/least illustrative word pairs
    (a) art:abstract (b) song:opera (c) footwear:boot (d) hair:brown



Conclusions

- Continuous semantic representation that
  - Leverages existing rich linguistic resources
  - Discovers new relations
  - Enables us to measure the degree of multiple relations
- Approaches
  - Better data representation
  - Matrix/Tensor decomposition
- Challenges & Future Work
  - Capture more types of knowledge in the model
  - Support more sophisticated inferential tasks