

Towards Holistic Concept Representations: Embedding Relational Knowledge, Visual Attributes, and Distributional Word Semantics

Steffen Thoma, **Achim Rettinger**, Fabian Both
rettinger@kit.edu, http://www.aifb.kit.edu/web/Achim_Rettinger/en

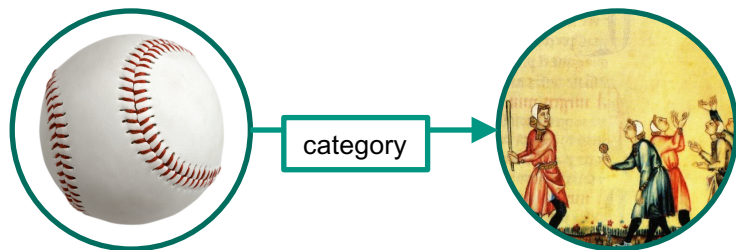
ADAPTIVE DATA ANALYTICS GROUP
INSTITUTE OF APPLIED INFORMATICS AND FORMAL DESCRIPTION METHODS (AIFB)



What is captured in entity-
embeddings
learned from KGs?

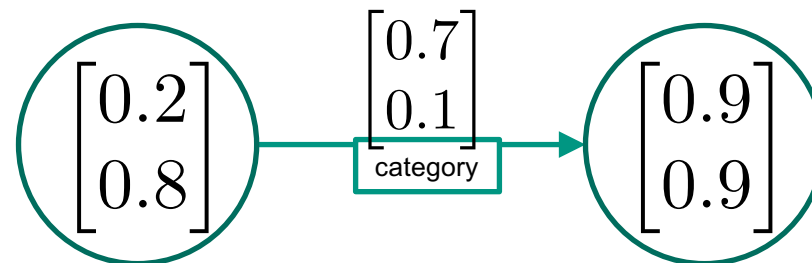
■ Latent Feature Models

- Latent Translation Models
- Tensor Decomposition
- Multi Layer Perceptrons
- Latent Graphical Models



■ Approaches

- **TransE**
- TransH
- TransR
- Rescal
- HoIE,
- ComplEx,
- RDF2Vec
-



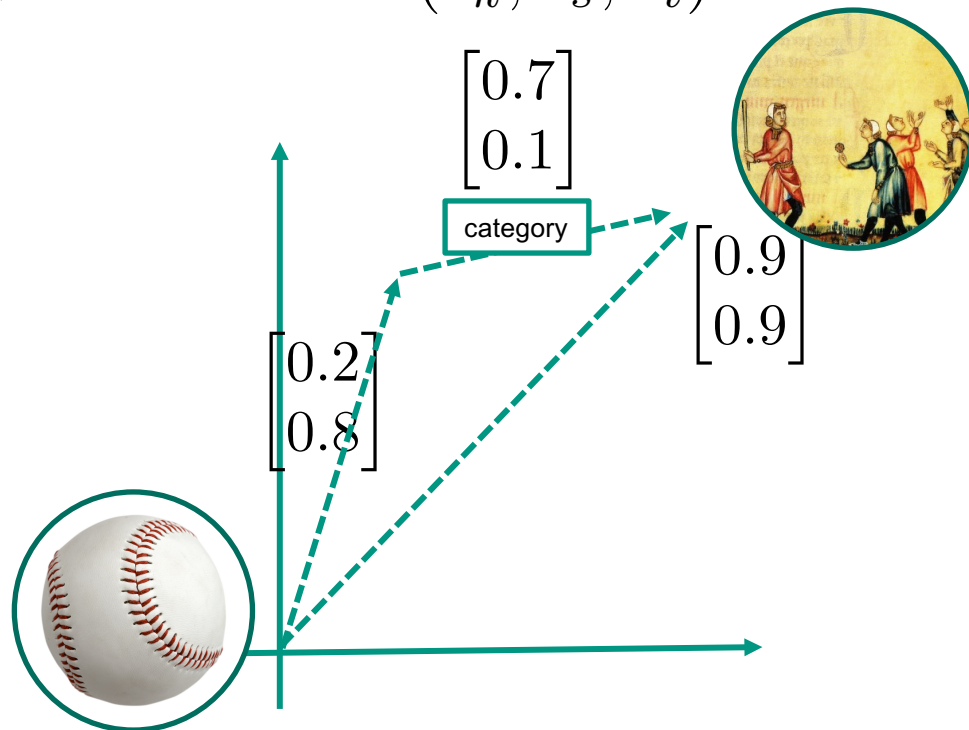
$$score(\mathbf{r}_k, \mathbf{e}_i, \mathbf{e}_j)$$

Latent Distance Models – TransE – Model

[Bor13]

$$\text{score}^{\text{TransE}}(\mathbf{r}_k, \mathbf{e}_i, \mathbf{e}_j)^{\text{known}} > \text{score}^{\text{TransE}}(\mathbf{r}_k, \mathbf{e}_s, \mathbf{e}_t)^{\text{corrupted}}$$

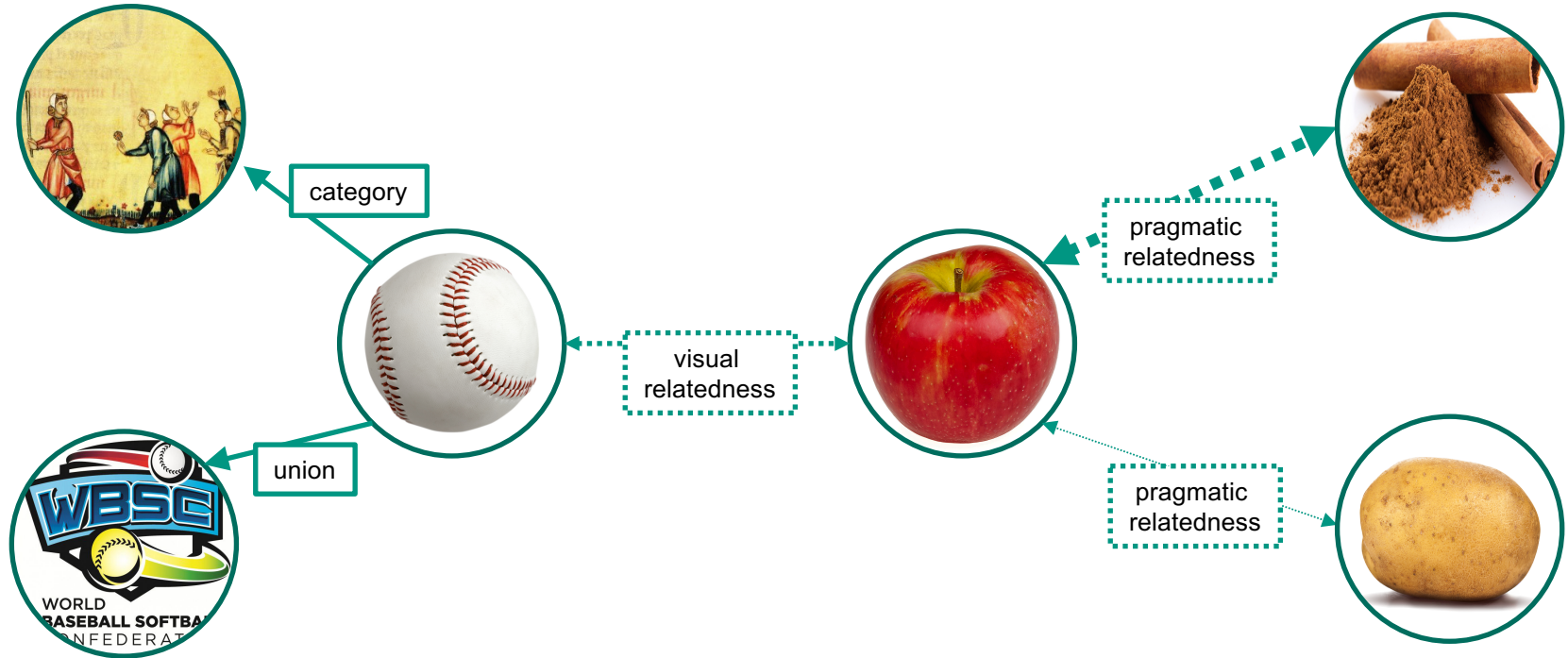
$$\begin{aligned} \text{score}^{\text{TransE}}(\mathbf{r}_k, \mathbf{e}_i, \mathbf{e}_j) \\ &= -d(\mathbf{e}_i + \mathbf{r}_k, \mathbf{e}_j) \\ &= -\|\mathbf{e}_i + \mathbf{r}_k - \mathbf{e}_j\|_2 \end{aligned}$$



What is captured in entity-embeddings learned from a KG?

They capture abstract relational context.

Is there other types of context that could complement entity embeddings?



Motivation

Other media (images, text documents) contain additional information:

Example Baseball:

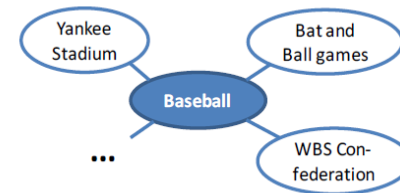
■ **Visual** – Shape, Color, Background



■ **Textual** – Co-occurrence Correlation

“... 26th pitcher in **baseball** history to have 40 games with at least 10 strikeouts ...”

■ **Knowledge Graph** - Relational Knowledge



Is there other types of context
that could complement entity
embeddings?

How do we collect such
diverse content with a
common encoding?

Yes. Context from
the visual and
lingual modality.

Visual Features



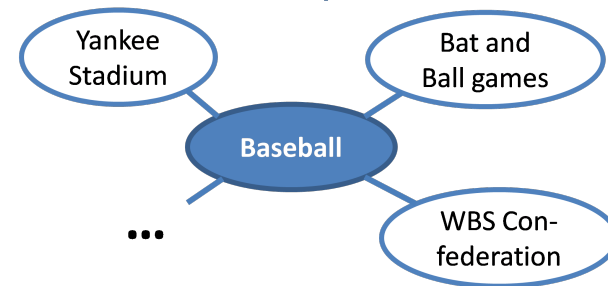
Word Embeddings



“... 26th pitcher in **baseball** history to have 40 games with at least 10 strikeouts ...”

“It's not some shocking **baseball** miracle.”

KG-Entity Embeddings

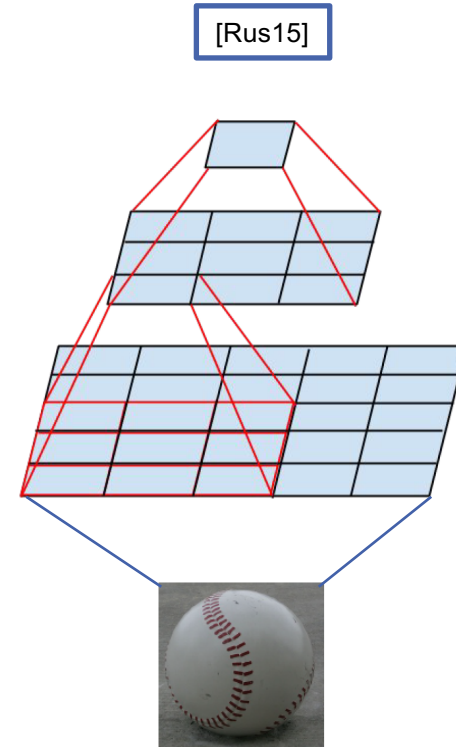


Visual Embedding – Inception V3

- Deep Convolutional Neural Networks
 - Optimized on object recognition

- Abstract visual features

Higher level layers correspond to more abstract features



Schematic Convolutional Net, abstracting visual features

Text Embedding – Word2Vec



[Mik13]

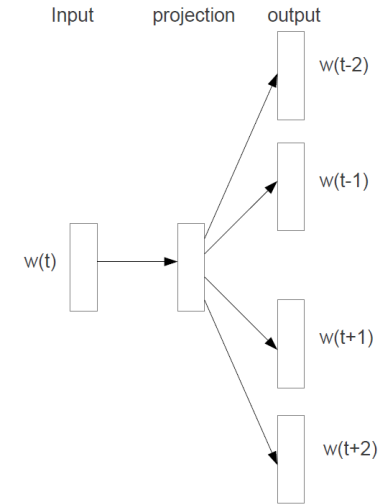
Words represented as vectors

„King“ →

0.2	0.1	1.5	0.3	...					
-----	-----	-----	-----	-----	--	--	--	--	--

Arithmetic operations


$$\text{vector}(\text{king}) - \text{vector}(\text{man}) + \text{vector}(\text{woman}) = \text{vector}(\text{queen})$$




Skip-gram: Predicting surrounding words

Visual Features



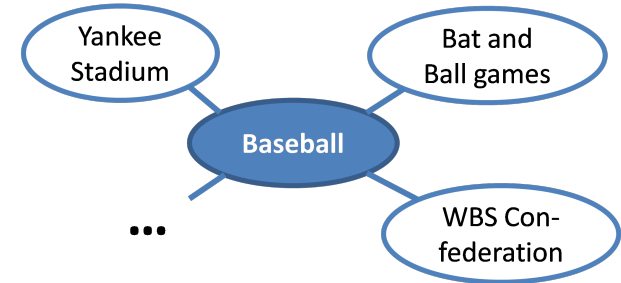
Word Embeddings



"... 26th pitcher in **baseball** history to have 40 games with at least 10 strikeouts ..."

"It's not some shocking **baseball** miracle."

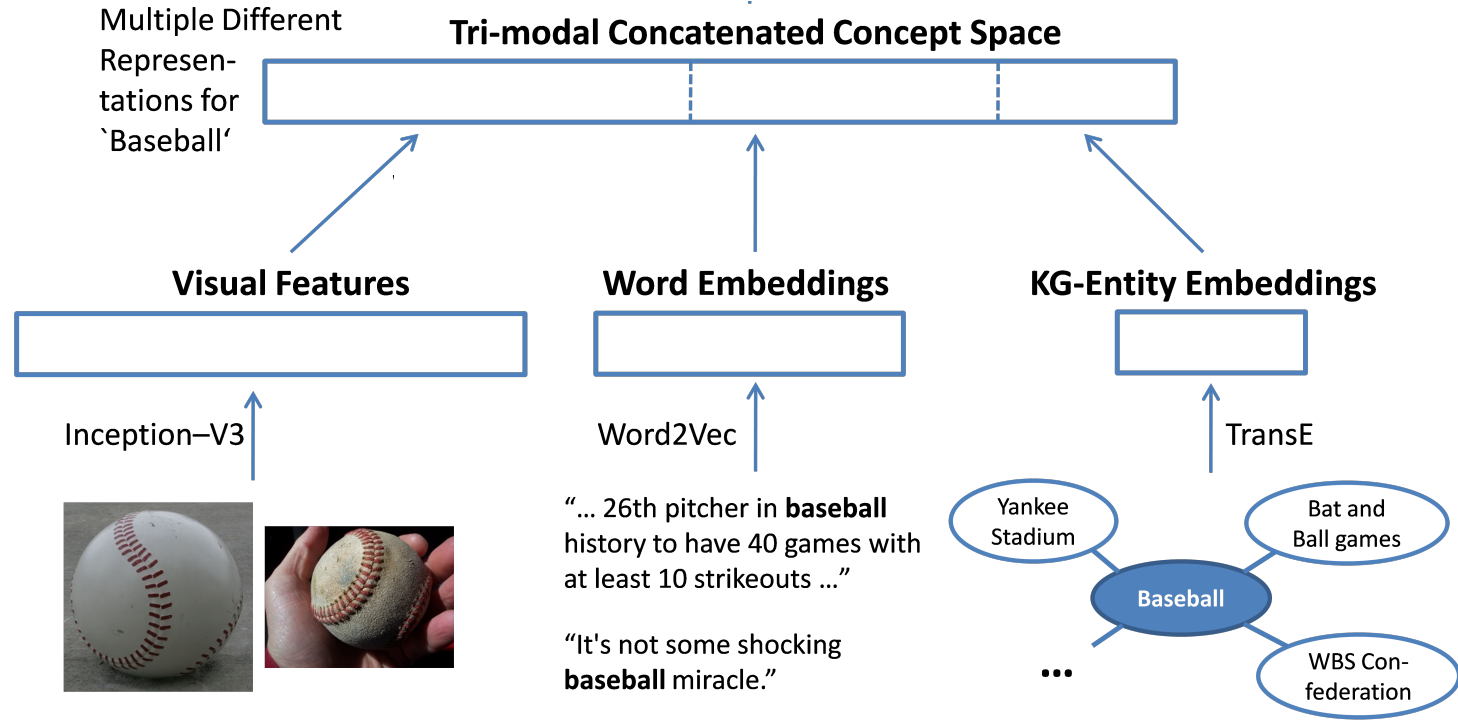
KG-Entity Embeddings



How do we collect such
diverse content with a
common encoding?

Multiple Embeddings

How do we align the
embeddings across
modalities?



Shared Concept Space

Alignment of concepts from model space to shared space.

- Textual

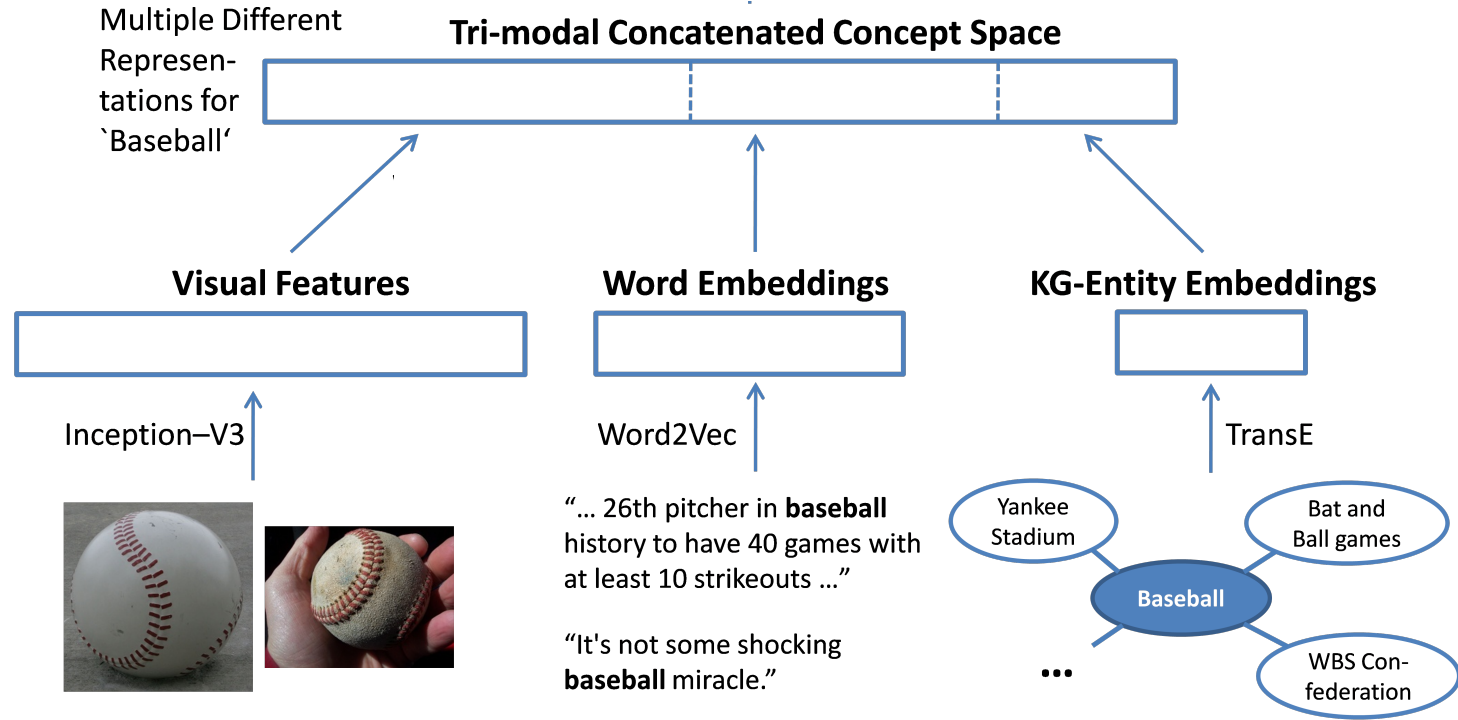
- Direct matching to words in model

- KG (DBPedia)

- Get most probable URI (entity) for a given word

- Visual

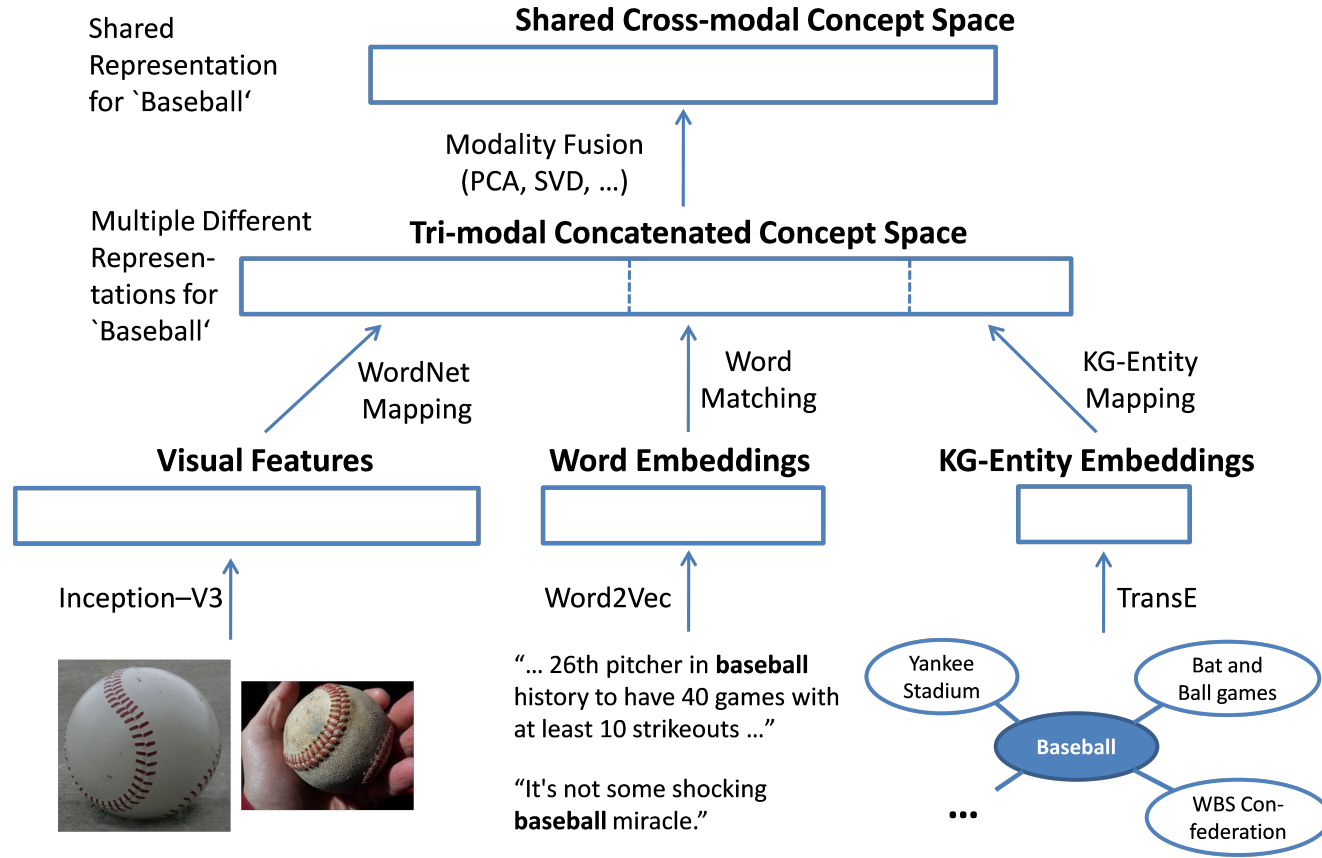
- Use WordNet hierarchy to get from image categories (synsets) to words



How do we align the
embeddings across
modalities?

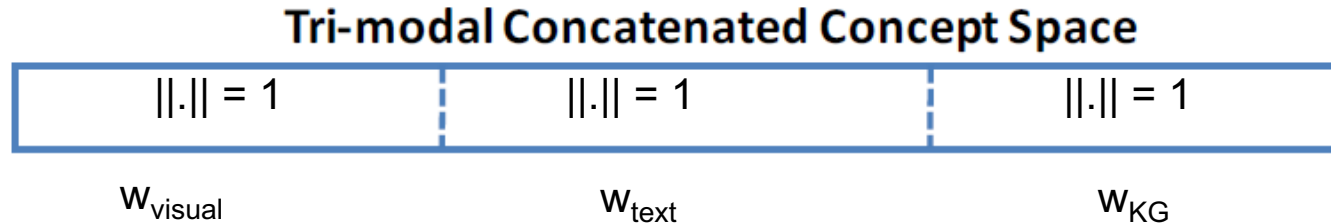
How do we identify
complementary
information?

Match them across
modalities

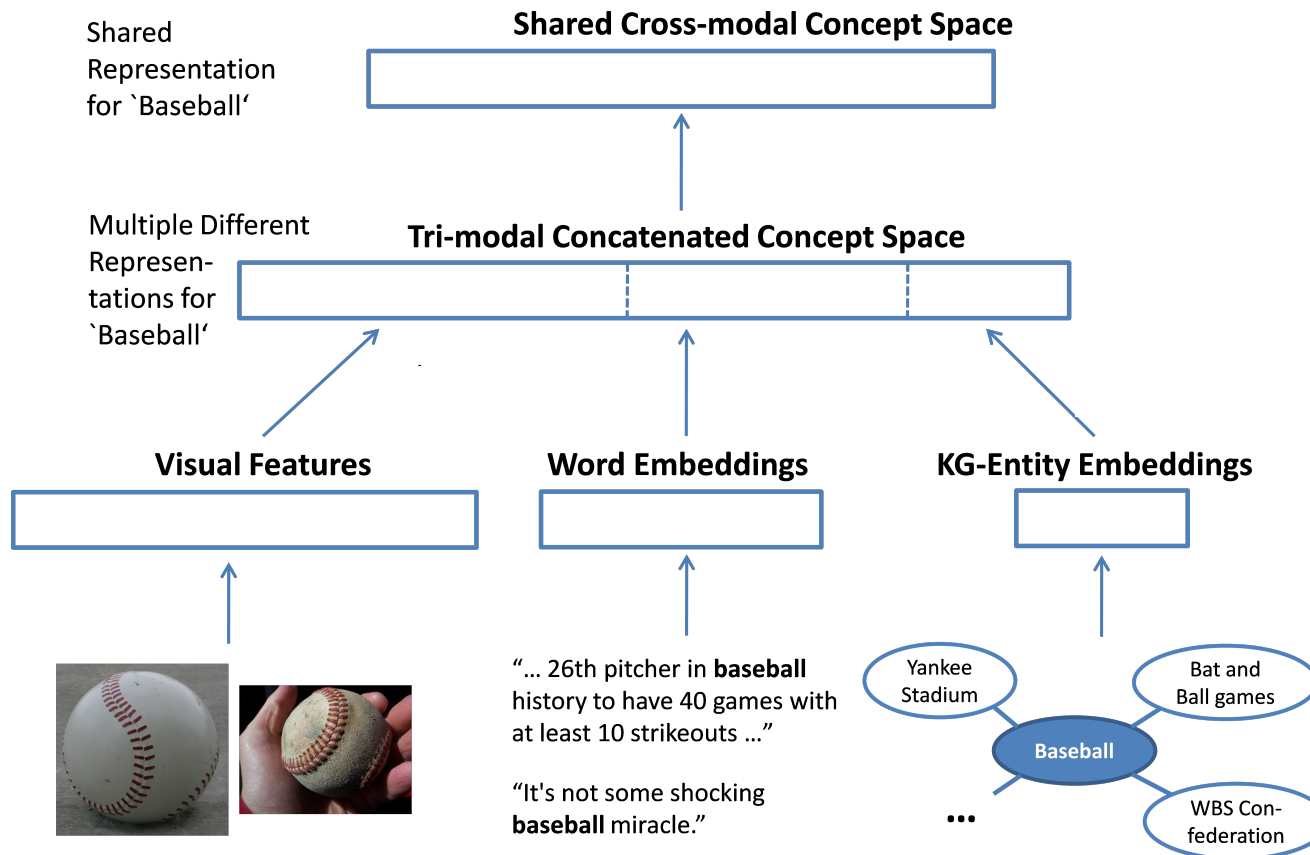


Fusion techniques

- Crucial: **normalization** and **weighting** before combination



- Shared Cross-modal Concept Space
 - PCA, SVD, Autoencoder



How do we identify
complementary
information?

Dimensionality
reduction techniques

How do we measure if
those embeddings are
more holistic in terms of
covered context?

Empirical Analysis – Word Similarity

Examples for word similarity:

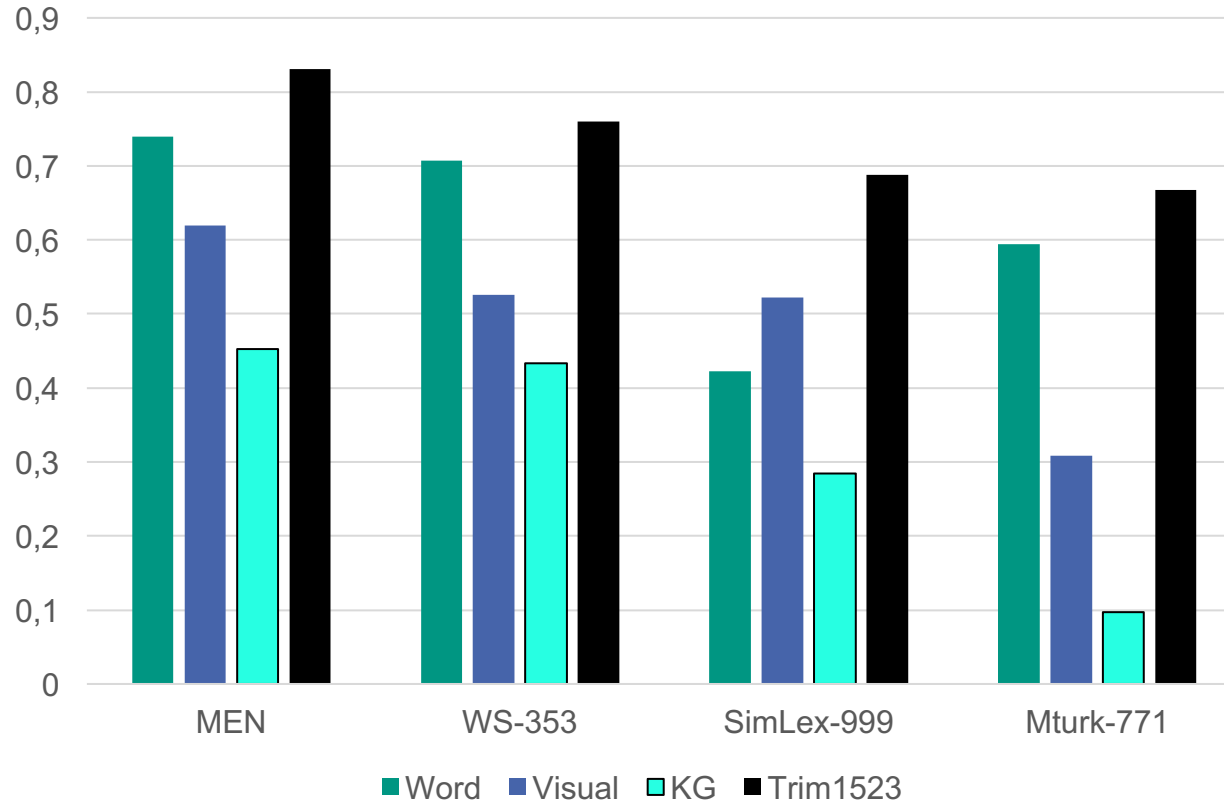
(sun, sunlight) → 50.0 (high similarity)

(happy, kiss) → 26.0 (medium similarity)

(bakery, zebra) → 0.0 (low similarity)

Datasets : MEN, WSS-353, SIMLEX-999, Mturk-771

Empricial Analysis – Word Similarity – Rank Correlation

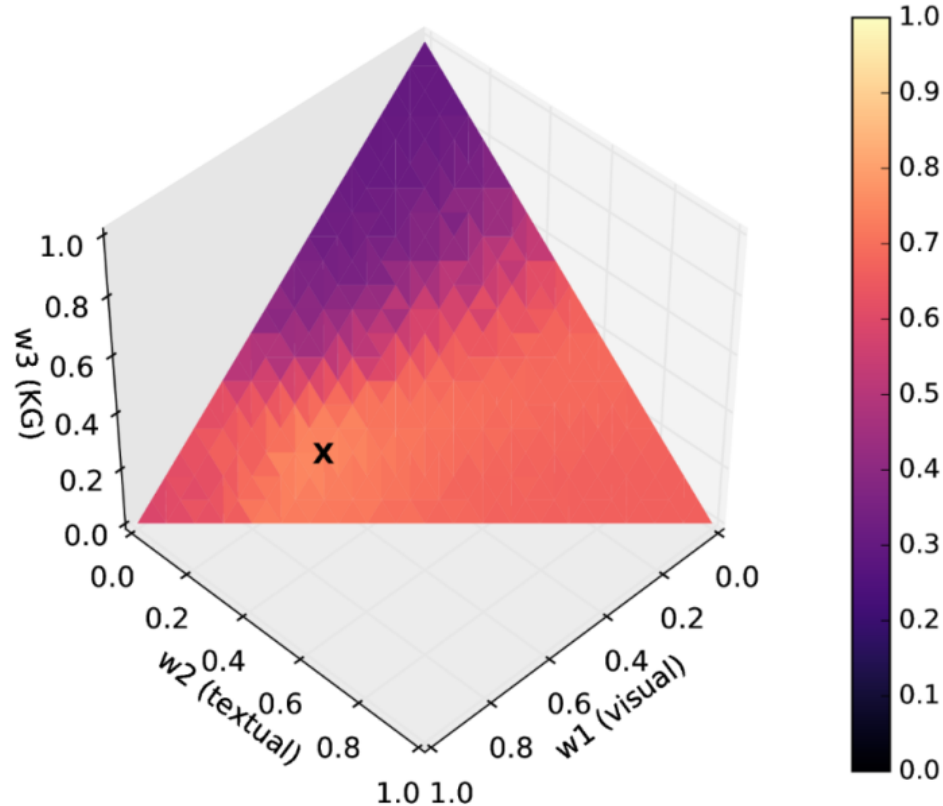


How do we measure if those embeddings are more holistic in terms of covered context?

Is every modality contributing information?

Word similarity assessed by humans

Empirical Analysis – Word Similarity – Influence of Modalities



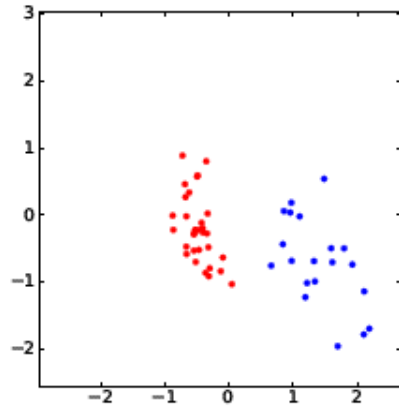
Is every modality
contributing
information?

How do the embedding
spaces differ?

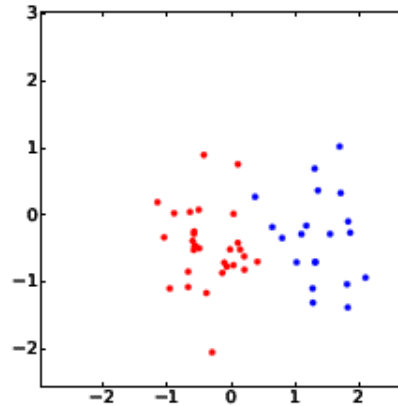
Yes.

Empirical Analysis – Entity Segmentation

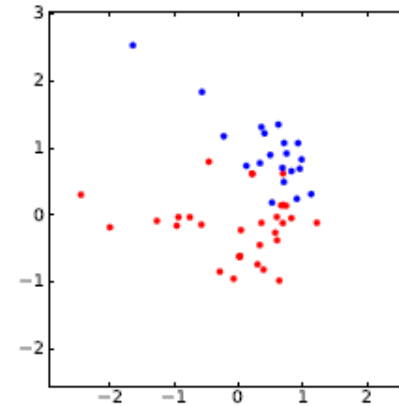
First two PCA components for various birds (blue) and land vehicles (red)



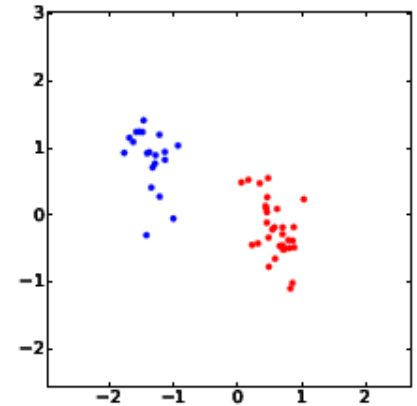
(a) Textual



(b) KG



(c) Visual



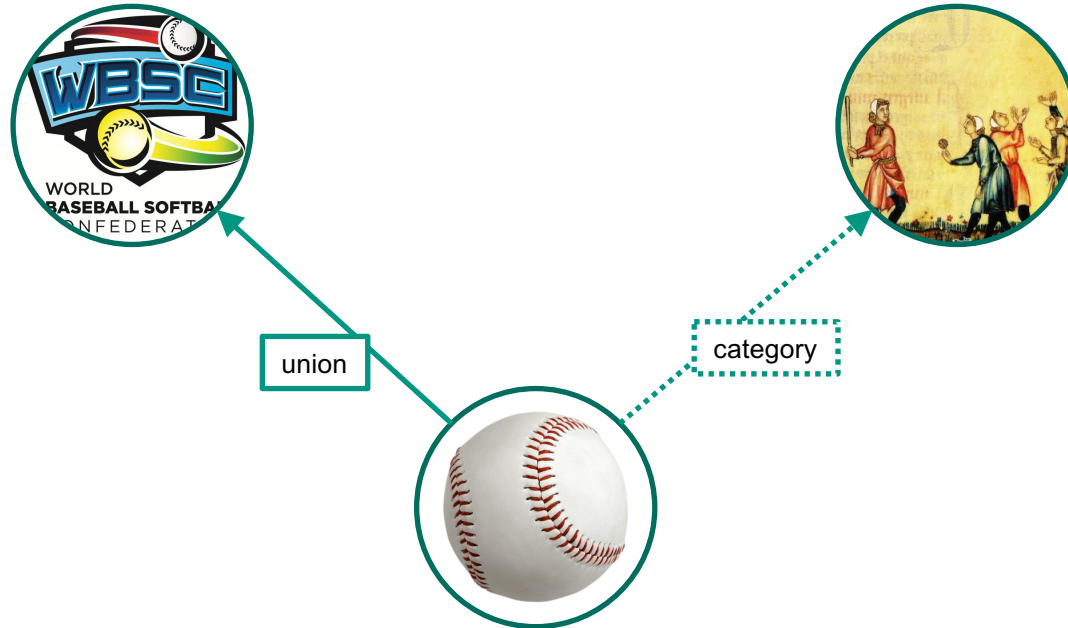
(d) CONC

How do the embedding
spaces differ?

Do knowledge graph
tasks benefit?

Entity segmentation

Empirical Analysis – Entity-Type Prediction



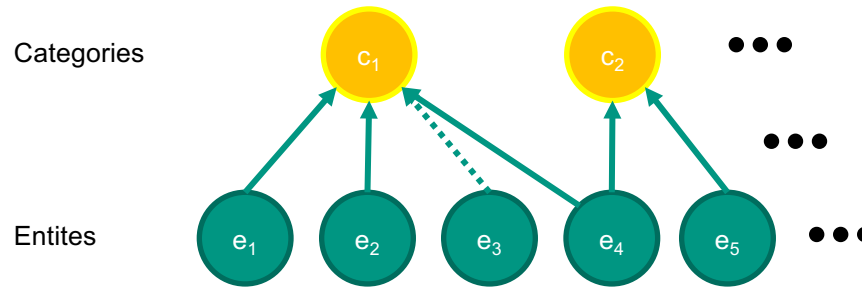
Empirical Analysis – Entity-Type Prediction

Hierarchic Construction (HC)

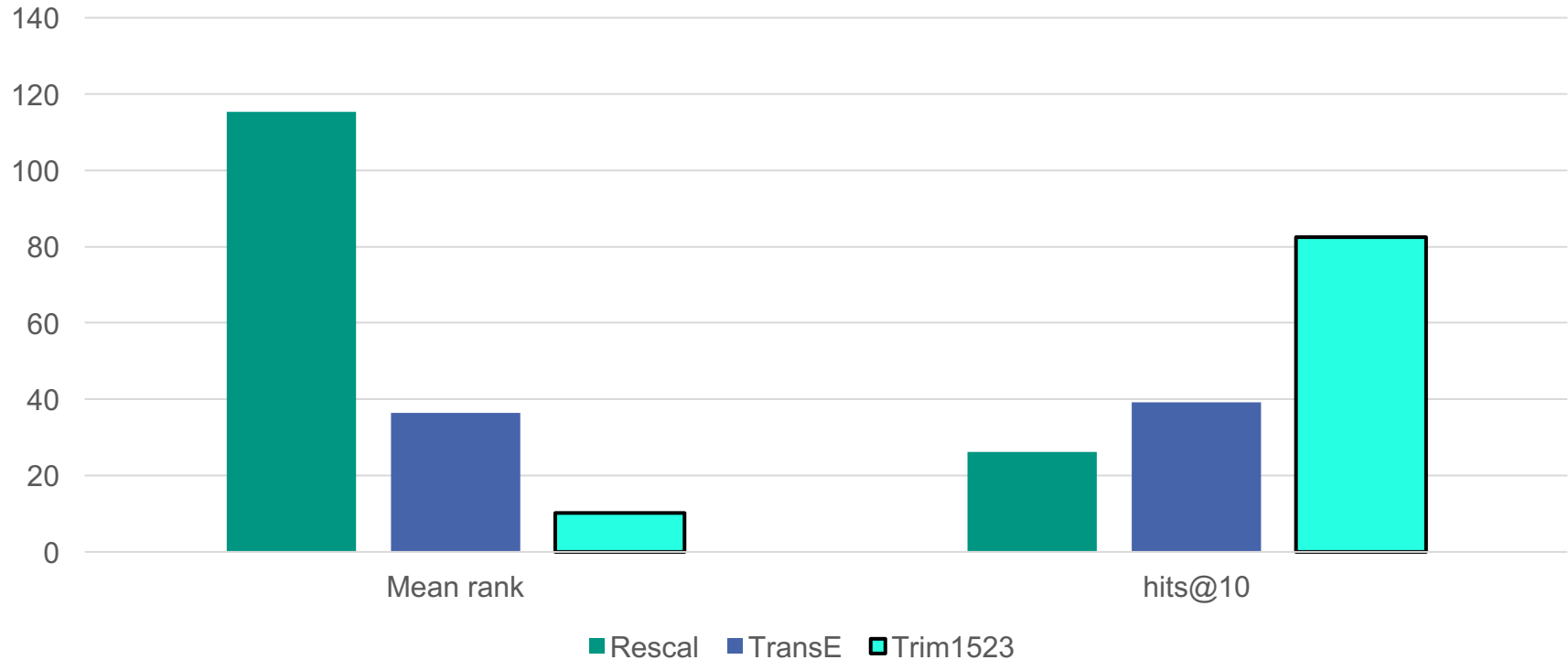
Constructing categorial embeddings from the multi-modal embeddings:

$$c_j = \frac{1}{N} \sum e_i, \forall c_j \text{ iff } (e_i, c_j) \text{ exists}$$

In each evaluation run: leave out the edges (e_i, c_j) which have to be predicted e.g. e_3 is left out for building c_1 as this connection exists and has to be predicted.



Empirical Analysis – Entity-Type Prediction



What are the lessons learned?

Visual common-sense knowledge and distributional text semantics complements entity embeddings.

Cross-modal concept representations show a significantly better performance on various benchmarks.

So, shouldn't everyone
try cross-modal concept
embeddings?

?

Future Challenges

1. How to scale to the size of KGs?
2. How to learn the most general-purpose entity representations? How to represent them?
3. Which modalities and data sources should/can be exploited?
4. Can you transfer knowledge back to single-modal embeddings? [Bot17]
5. Early-fusion techniques better?

References (related work)

[Rus15] Russakovsky, Deng, Su, Krause, Satheesh, Ma, Huang, Karpthy, Khosla, Bernstein, Berg, Fei-Fei: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) 115(3), 211–252 (2015)

[Bor13] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, Oksana Yakhnenko: *Translating Embeddings for Modeling Multi-relational Data*. NIPS 2013: 2787-2795

[Nic16] Maximilian Nickel, Kevin Murphy, Volker Tresp, Evgeniy Gabrilovich: *A Review of Relational Machine Learning for Knowledge Graphs*. Proceedings of the IEEE 104(1): 11-33 (2016)

[Mik13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)

[Bot17] Fabian Both, Steffen Thoma, Achim Rettinger: Cross-modal Knowledge Transfer: Improving the Word Embedding of Apple by Looking at Oranges. K-CAP2017, The 9th International Conference on Knowledge Capture, ACM, Dezember, 2017

Questions?
Comments?
Ideas?
Request?

Contact me:
rettinger@kit.edu