Connecting Large-Scale Knowledge Bases and Natural Language

Antoine Bordes & Jason Weston

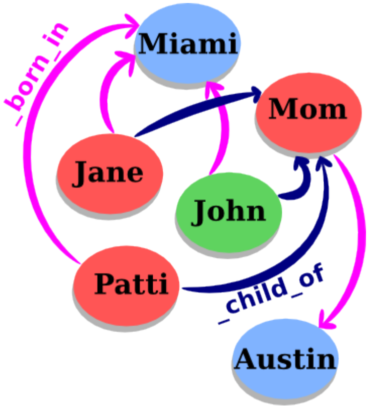
CNRS - Univ. Tech. Compiègne & Google

UW - MSR Summer Institute 2013,

Alderbrook Resort, July 23, 2013

Knowledge Bases (KBs)

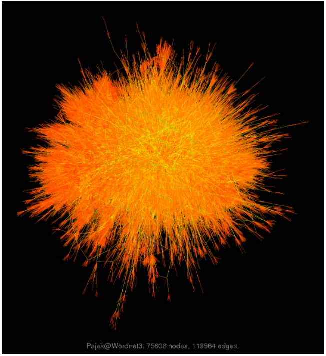
- Data is structured as a graph.
- Each node = an entity.
- Each edge = a relation.
- A relation = (*sub*, *rel*, *obj*):
  - *sub* =*subject*,
  - *rel* = *relation type*,
  - *obj* = *object*.
- Nodes w/o features.



Example of KB: WordNet

- WordNet: dictionary where each entity is a sense (synset).

- Popular in NLP.
- Statistics:
  - 117k entities;
  - 20 relation types;
  - 500k relations.
- Examples:
  (car_NN_1, _has_part, _wheel_NN_1)
  (score_NN_1, _is_a, _rating_NN_1)
  (score_NN_2, _is_a, _sheet_music_NN_1)



Example of KB: Freebase

- Freebase: huge collaborative (hence noisy) KB.

- Part of the Google Knowledge Graph.
- Statistics:
  - > 80M of entities;
  - > 20k relation types;
  - > 1.2B relations.
- Examples:

  (Lil Wayne, _born_in, New Orleans)

  (Seattle, _contained_by, USA)

  (Machine Learning, _subdiscipline, Artificial Intelligence)

Connecting KBs and Natural Language

- **Why?**
  - Text → KB: information extraction;
  - KB → Text: interpretation (NER, semantic parsing), summary.

- **Main issue:** KBs are hard to manipulate.
  - Very large dimensions: $10^5 – 10^8$ entities, $10^7 – 10^9$ rel. types;
  - Sparse: few valid links;
  - Noisy/incomplete: missing/wrong relations/entities.

- **How?**
  1. Encode KBs into low-dimensional vector spaces;
  2. Use these representations as KB data in text applications.

Menu

- Context
- Modeling Knowledge Bases
  - Embedding-based Models
  - Experiments on Freebase
- Connecting KBs & Natural Language
  - Wordnet+Text for Word-Sense Disambiguation
  - Freebase+Text for Relation Extraction
- Conclusion
  - What now?

Statistical Relational Learning

- **Framework:**
  - n_s subjects $\{sub_i\}_{i \in [\![1;n_s]\!]}$
  - n_r relation types $\{rel_k\}_{k \in [\![1;n_r]\!]}$
  - n_o objects $\{obj_j\}_{j \in [\![1;n_o]\!]}$

  → For us, $n_s = n_o = n_e$ and $\forall i \in [\![1;n_e]\!], sub_i = obj_i$.

  - A relation exists for (sub_i, rel_k, obj_j) if $rel_k(sub_i, obj_j) = 1$
- **Goal:** We want to model, from data,

$$\mathbb{P}[rel_k(sub_i, obj_j) = 1]$$

(equivalent to approximate the binary tensor $\mathbf{X} \in \{0,1\}^{n_s \times n_o \times n_r}$)

Energy-based Learning

Two main ideas:

1. Models based on low-dimensional continuous vector embeddings for entities and relation types, learned to define a similarity criterion.
2. Stochastic training with sub-sampling of unknown relations.

Learning Representations

- Subjects and objects are represented by vectors in $\mathbb{R}^d$.
  - $\{sub_i\}_{i \in [\![1;n_s]\!]} \quad \rightarrow \quad [\mathbf{s}^1, \ldots, \mathbf{s}^{n_s}] \in \mathbb{R}^{d \times n_s}$
  - $\{obj_j\}_{j \in [\![1;n_o]\!]} \quad \rightarrow \quad [\mathbf{o}^1, \ldots, \mathbf{o}^{n_o}] \in \mathbb{R}^{d \times n_o}$

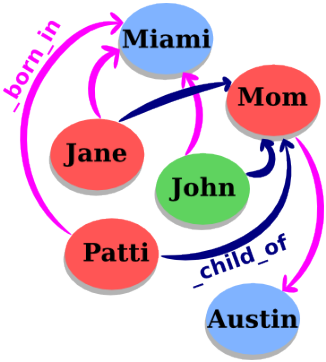
  For us, $n_s = n_o = n_e$ and $\forall i \in [\![1;n_e]\!], \mathbf{s}_i = \mathbf{o}_i$.

- Rel. types = similarity operators between subjects/objects.
  - $\{rel_k\}_{k \in [\![1;n_r]\!]} \quad \rightarrow \quad$ operateurs $\{\boldsymbol{r}_k\}_{k \in [\![1;n_r]\!]}$

- Learning similarities depending on $rel \rightarrow d(sub, rel, obj)$.
  (we can retrieve a probability using a transfer function)

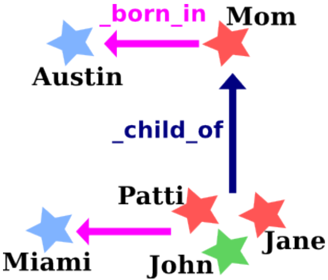
Modeling Relations as Translations

Intuition: we would like that $\boldsymbol{s} + \boldsymbol{r} \approx \boldsymbol{o}$.



Modeling Relations as Translations

Intuition: we would like that $\boldsymbol{s} + \boldsymbol{r} \approx \boldsymbol{o}$.



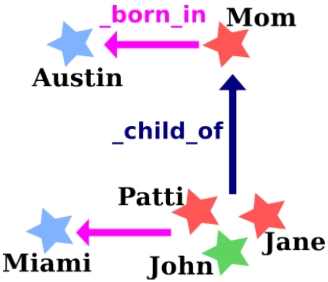
Modeling Relations as Translations

Intuition: we would like that $\boldsymbol{s} + \boldsymbol{r} \approx \boldsymbol{o}$.

We define the similarity measure:

$$d(sub, rel, obj) = ||\boldsymbol{s} + \boldsymbol{r} - \boldsymbol{o}||_2^2$$

We learn $\boldsymbol{s}$,$\boldsymbol{r}$ and $\boldsymbol{o}$ that verify that.



Stochastic Training

- Learning by stochastic gradient descent: one observed (true?) relation after the other.

- For each relation from the training set:
  1. we sub-sample unobserved relations (false?).
  2. we check if the similarity of the true relation is lower.
  3. **if not**, we update parameters of the considered relations.

- Stopping criterion: performance on a validation set.

Menu

- Context
- Modeling Knowledge Bases
  - Embedding-based Models
  - Experiments on Freebase
- Connecting KBs & Natural Language
  - Wordnet+Text for Word-Sense Disambiguation
  - Freebase+Text for Relation Extraction
- Conclusion
  - What now?

Chunks of Freebase

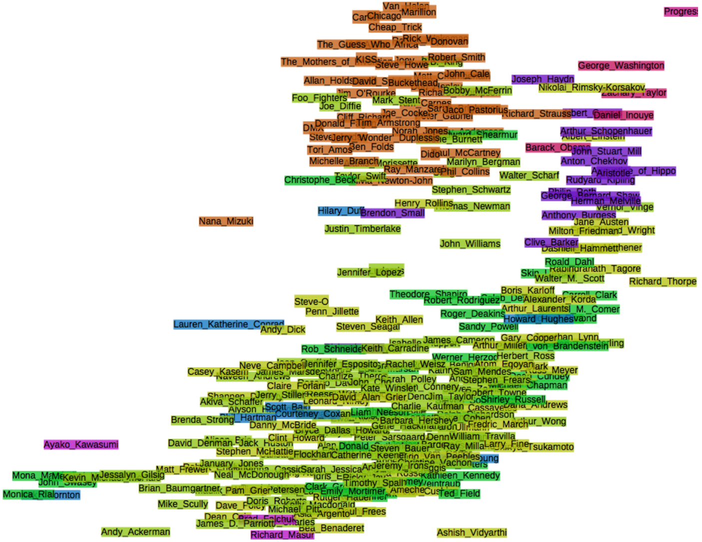
- **Data statistics**:

	Entities (n_e)	Rel. (n_r)	Train. Ex.	Valid. Ex.	Test Ex.
Freebase15k	14,951	1,345	483,142	50,000	59,071
Freebase1M	1×10^9	23,382	17.5×10^9	50,000	177,404

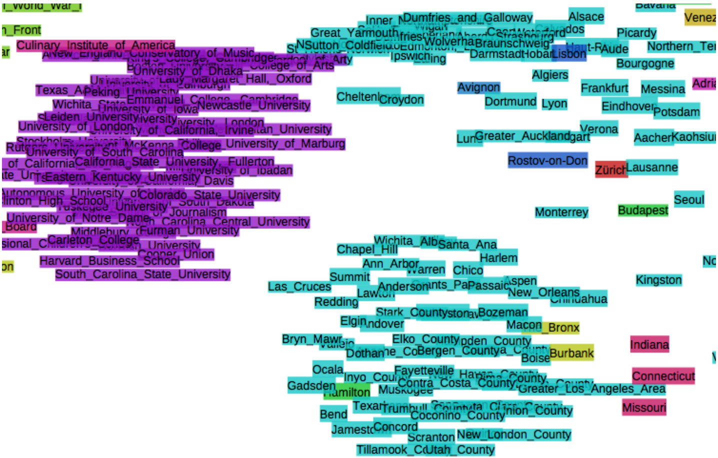
- **Experimental setup**:
  - Embedding dimension: 50.
  - Training time:
    - on Freebase15k: ≈5h (on 1 CPU),
    - on Freebase1M : ≈1j (on 16 cores).

Visualization of 1,000 Entities

Visualization of 1,000 Entities - Zoom 1



Visualization of 1,000 Entities - Zoom 2

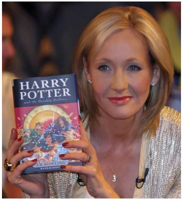


Visualization of 1,000 Entities - Zoom 3

Link Prediction

"Who influenced J.K. Rowling?"

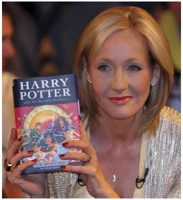
J. K. Rowling `_influenced_by` ?



Link Prediction

"Who influenced J.K. Rowling?"

J. K. Rowling `_influenced_by` G. K. Chesterton
J. R. R. Tolkien
C. S. Lewis
Lloyd Alexander
Terry Pratchett
Roald Dahl
Jorge Luis Borges
Stephen King
Ian Fleming



Link Prediction

"Which genre is the movie WALL-E?"

WALL-E `_has_genre` ?

Link Prediction

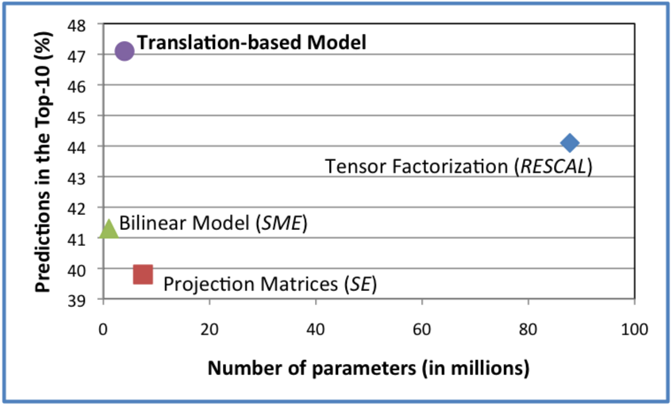
"Which genre is the movie WALL-E?"

WALL-E `_has_genre`

- Animation
- Computer animation
- Comedy film
- Adventure film
- Science Fiction
- Fantasy
- Stop motion
- Satire
- Drama

Link Prediction

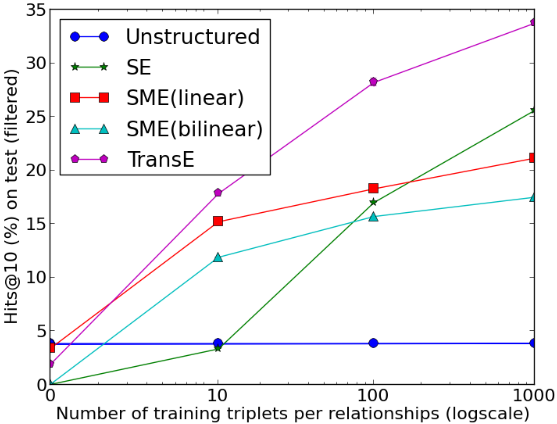
On Freebase15k:



On Freebase1M, our model predicts 34% in the Top-10.

Learn Unknown Relation Types

Learning embeddings of 40 unkowns relation types.



Menu

- Context
- Modeling Knowledge Bases
  - Embedding-based Models
  - Experiments on Freebase
- Connecting KBs & Natural Language
  - Wordnet+Text for Word-Sense Disambiguation
  - Freebase+Text for Relation Extraction
- Conclusion
  - What now?

Disambiguation within a Specific Framework

Disambiguation → connect free text and the KB WordNet.

Towards open-text semantic parsing:

``A musical score accompanies a television program ."

↓ ***Semantic Role Labeling***

(``A musical score", ``accompanies", ``a television program")

↓ ***Preprocessing (POS, Chunking, ...)***

`((_musical_JJ score_NN ), _accompany_VB , _television_program_NN )`

↓ ***Word-sense Disambiguation***

`((_musical_JJ_1 score_NN_2), _accompany_VB_1, _television_program_NN_1)`

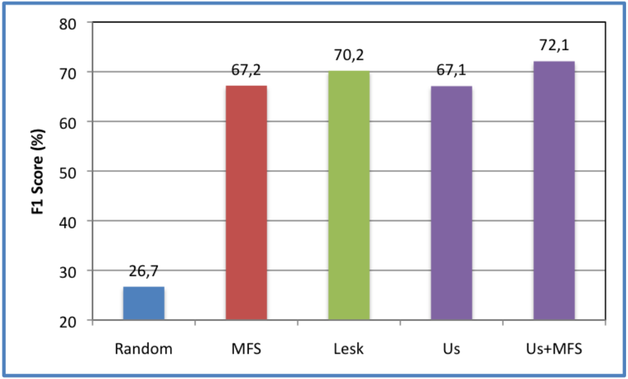
Joint Modeling of Text and WordNet

- Text is converted into relations (*sub*,*rel*,*obj*).
- We learn a vector for any symbol: words, entities and relation types from WordNet.
- Our system can label 37,141 words with 40,943 synsets.

	Train. Ex.	Test Ex.	Labeled?	Symbol
WordNet	146,442	5,000	No	synsets
Wikipedia	2,146,131	10,000	No	words
ConceptNet	11,332	0	Non	words
Ext. WordNet	42,957	5,000	Yes	words+synsets
Unamb. Wikip.	981,841	0	Yes	words+synsets
TOTAL	3,328,703	20,000	-	-

Experimental Results

F1-score on 5,000 test sentences to disambiguate.



Enrich WordNet

We create similarities going beyond WordNet.

"what does an army attack?"

army_NN_1 `attack_VB_1` ?

Enrich WordNet

We create similarities going beyond WordNet.

"what does an army attack?"

army_NN_1 `attack_VB_1` troop_NN_4
armed_service_NN_1
ship_NN_1
territory_NN_1
military_unit_NN_1

Enrich WordNet

We create similarities going beyond WordNet.

"Who or what earns money"

? `earn_VB_1` money_NN_1

Enrich WordNet

We create similarities going beyond WordNet.

"Who or what earns money"

person_NN_1	earn_VB_1	money_NN_1
business_firm_NN_1		
family_NN_1		
payoff_NN_3		
card_game_NN_1		

Menu

- Context
- Modeling Knowledge Bases
  - Embedding-based Models
  - Experiments on Freebase
- Connecting KBs & Natural Language
  - Wordnet+Text for Word-Sense Disambiguation
  - Freebase+Text for Relation Extraction
- Conclusion
  - What now?

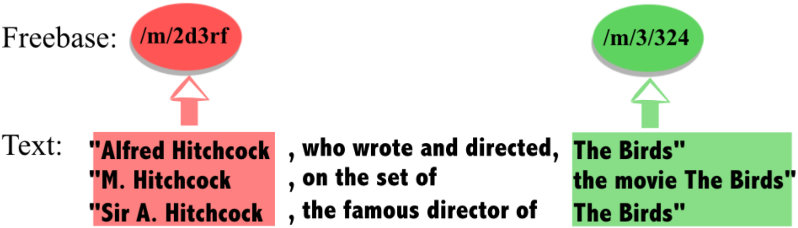
Relation Extraction

Given a bunch of sentences.

Text:	"Alfred Hitchcock	, who wrote and directed,	The Birds"
	"M. Hitchcock	, on the set of	the movie The Birds"
	"Sir A. Hitchcock	, the famous director of	The Birds"

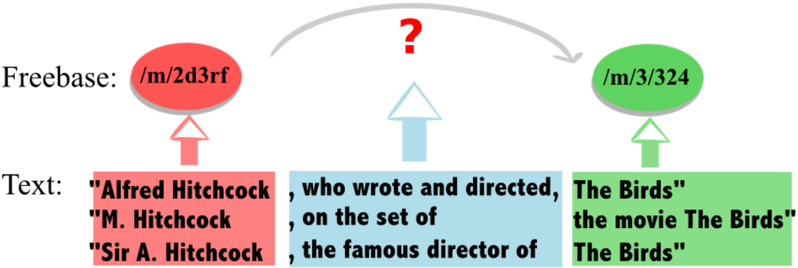
Relation Extraction

Given a bunch of sentences concerning the same pair of entities.



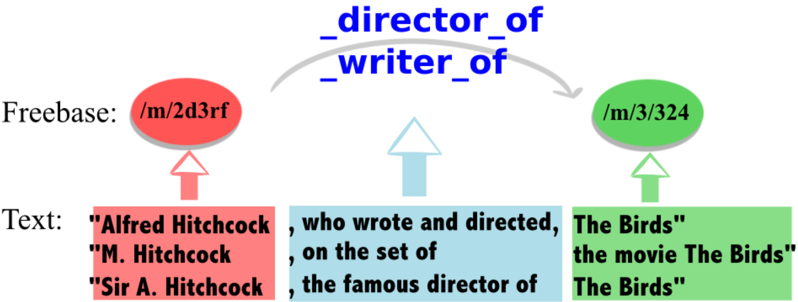
Relation Extraction

Goal: identify if there is a relation between them to add to the KB.



Relation Extraction

And from which type, to enrich an existing KB.



Jointly use Text and Freebase

- **Standard Method:** a classifier is trained to predict the relation type, given *txts* and (*sub*, *obj*):

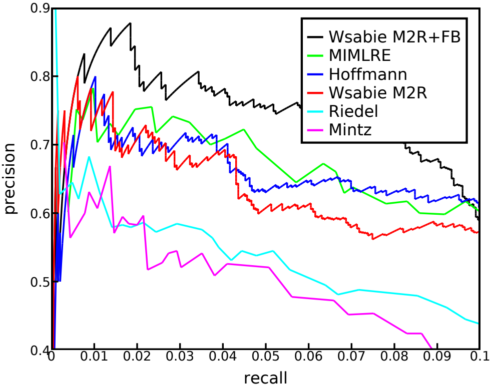
$$r(txts, sub, obj) = \arg\max_{rel'} S_{txt2rel}(txts, rel')$$

- **Idea:** extract relations by using both text + available knowledge (= current KB).
- Our model of the KB forces extracted relations to agree with it:

$$r(txts, sub, obj) = \arg\max_{rel'} \left(S_{txt2rel}(txts, rel') - d_{BC}(sub, rel', obj) \right)$$

Experiments on NYT+Freebase

We learn on New York Times papers and on Freebase.



Precision/recall curve for predicting relations.

Menu

- Context
- Modeling Knowledge Bases
  - Embedding-based Models
  - Experiments on Freebase
- Connecting KBs & Natural Language
  - Wordnet+Text for Word-Sense Disambiguation
  - Freebase+Text for Relation Extraction
- Conclusion
  - What now?

Encode KBs into vector spaces

- KBs are rich but need attention.
- Learn to project into vector spaces:
    - Ease their visualization;
    - Allows for link prediction (with/without ext. data);
    - Facilitate their use int other systems;
    - Compact format.
- Is that all?

Challenges

We're just getting started:

- How to reason: combine logic, deduction.
- Evaluate confidence in predictions.
- Summarize KBs.
- Fusion KBs.
- Connect text and KBs: mutual interactions.
- etc.

Fin

Data/code available from my webpage.

Thanks!

antoine.bordes@hds.utc.fr
http://www.hds.utc.fr/~bordesan

References

1. Learning Structured Embeddings of Knowledge Bases.
   A. Bordes, J. Weston, R. Collobert & Y. Bengio. *AAAI, 2011*.
2. Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing.
   A. Bordes, X. Glorot, J. Weston & Y. Bengio. *AISTATS, 2012*.
3. A Latent Factor Model for Highly Multi-relational Data.
   R. Jenatton, N. Le Roux, A. Bordes & G. Obozinski. *NIPS, 2012*.
4. A Semantic Matching Energy Function for Learning with Multi-relational Data.
   A. Bordes, X. Glorot, J. Weston & Y. Bengio. *MLj, 2013.*
5. Irreflexive and Hierarchical Relations as Translations.
   A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston & O. Yakhnenko. *ICML Workshop on Structured Learning, 2013*