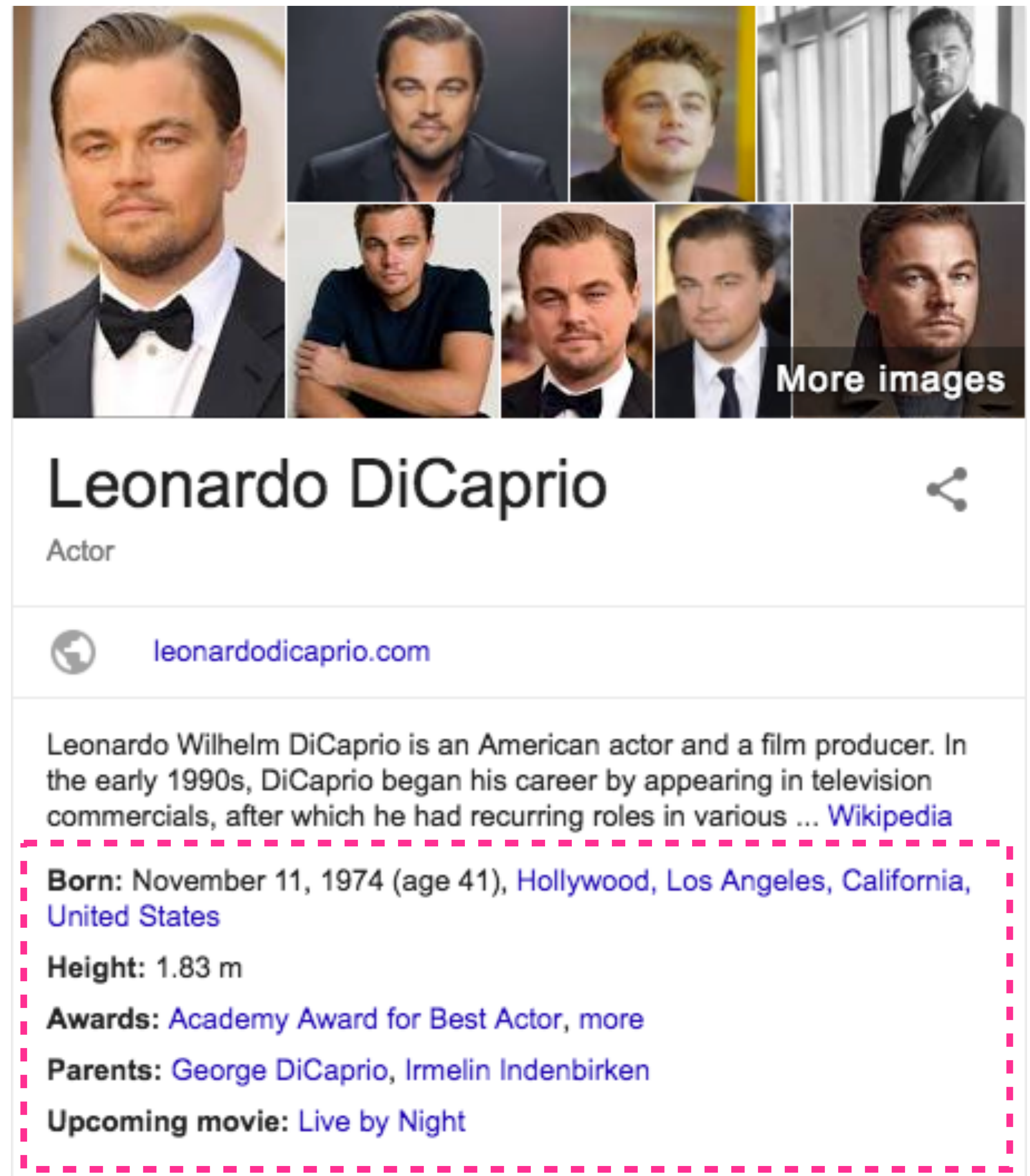
PROBABILISTIC KNOWLEDGE GRAPH CONSTRUCTION:
COMPOSITIONAL AND INCREMENTAL APPROACHES

Dongwoo Kim with Lexing Xie and Cheng Soon Ong
CIKM 2016

WHAT IS KNOWLEDGE GRAPH (KG)?

- KG widely used in various tasks such as search and Q&A
- **Triple** (a basic unit of KG):
  - Consists of source and target entities and a relation
  - e.g: <DiCaprio, starred in, The Revenant>



INCOMPLETE KNOWLEDGE GRAPH

- KG is constructed from external sources
  - Wikipedia, Web pages…
- Many existing KGs are incomplete
  - 93.8% of persons from Freebase have no place of birth (Min et al, 2013)
  - Hard to obtain from external sources
- Need alternative way to complete KG

In this talk, I will discuss

- **Knowledge graph completion** task
  - where our goal is to identify which triple is positive or negative
- Specifically, discuss two approaches for knowledge base completion:
  1. **Predict** unknown triples based on observed triples
  2. **Manually label** unknown triples with **human experts**
- These two approaches are **complementary** to each other

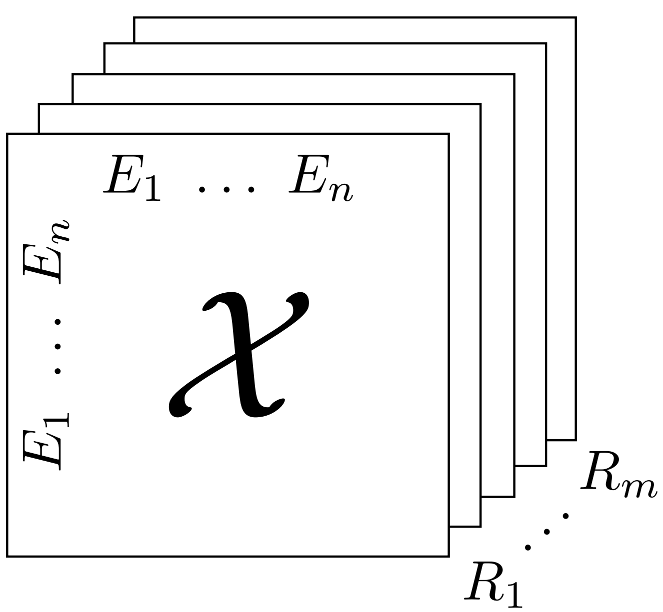
1. Unknown Triple Prediction

- Unknown triples may be inferred from existing triples
  - <DiCaprio, starred in, The Revenant>

    → <DiCaprio, job, actor>
- Statistical relational model
  - Goal: predict unknown triples from known triples
  - Find latent semantics of entities and relations

Tensor Representation of KG

- KG can be represented as 3-dimensional binary tensor
- or a collection of matrices each of which represent one relation
    - Example: <DiCaprio, starred in, The Revenant>
      <DiCaprio, not starred in, Star Trek>



$$x_{ikj} = 1 \left\{ \begin{array}{l} \text{i = DiCaprio} \\ \text{j = The Revenant} \\ \text{k = starred in} \end{array} \right.$$

$$x_{ikj'} = 0 \left\{ \begin{array}{l} \text{i = DiCaprio} \\ \text{j' = Star Trek} \\ \text{k = starred in} \end{array} \right.$$

Bilinear tensor factorisation (RESCAL) (Nickel, et. al, 2011)

- Latent space model
    - Represent an entity as $\mathbf{D}$-dimensional vector
    - Represent a relation as $\mathbf{D}\mathrm{x}\mathbf{D}$ matrix
- Better performance on knowledge graph to compare with general purpose tensor factorisation methods
  (Dong, et. al, 2014)

LATENT REPRESENTATION OF ENTITIES AND RELATIONS

Leonardo DiCaprio

$$e_{\mathrm{D}} = \begin{bmatrix} 0.8 \\ 0.1 \end{bmatrix}$$

Receive Award

$$R_{\mathrm{RA}} = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.1 \end{bmatrix}$$

Oscar

$$e_{\mathrm{O}} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$

Yoshua Bengio

$$e_{\mathrm{B}} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

<Bengio, Receive, Oscar>

$$x_{\mathrm{B \cdot RA \cdot O}} \approx e_{\mathrm{B}}^{\top} R_{\mathrm{RA}} e_{\mathrm{O}}$$

<DiCaprio, Receive, Oscar>

$$x_{\mathrm{D \cdot RA \cdot O}} \approx e_{\mathrm{D}}^{\top} R_{\mathrm{RA}} e_{\mathrm{O}}$$

Latent Representation of Entities and Relations

Leonardo DiCaprio

$$e_{\mathrm{D}} = \begin{bmatrix} 0.8 \\ 0.1 \end{bmatrix}$$

Yoshua Bengio

$$e_{\mathrm{B}} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

Receive Award

$$R_{\mathrm{RA}} = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.1 \end{bmatrix}$$

Oscar

$$e_{\mathrm{O}} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$

<Bengio, Receive, Oscar> < <DiCaprio, Receive, Oscar>

$$e_{\mathrm{B}}^{\top} R_{\mathrm{RA}} e_{\mathrm{O}} = 0.16 \quad < \quad e_{\mathrm{D}}^{\top} R_{\mathrm{RA}} e_{\mathrm{O}} = 0.60$$

→ *DiCaprio is more likely to get the Oscar award!*

GOAL OF STATISTICAL RELATIONAL MODELS

Leonardo DiCaprio

$$e_D = \begin{bmatrix} ? \\ ? \end{bmatrix}$$

Receive Award

$$R_{RA} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

Oscar

$$e_O = \begin{bmatrix} ? \\ ? \end{bmatrix}$$

Yoshua Bengio

$$e_B = \begin{bmatrix} ? \\ ? \end{bmatrix}$$

→ *The goal of statistical relational model is to infer latent vectors & matrices from known triples*

→ *Predict new positive triples from inferred latent space*

PROBABILISTIC BILINEAR (COMPOSITIONAL) MODEL

- We propose **probabilistic RESCAL** (PRESCAL)
- Place a probability distribution over entity and relation matrices

$$e_i \sim \mathcal{N}_D(0, \sigma_e^2 I) \qquad R_k \sim \mathcal{N}_{D^2}(0, \sigma_R^2 I)$$

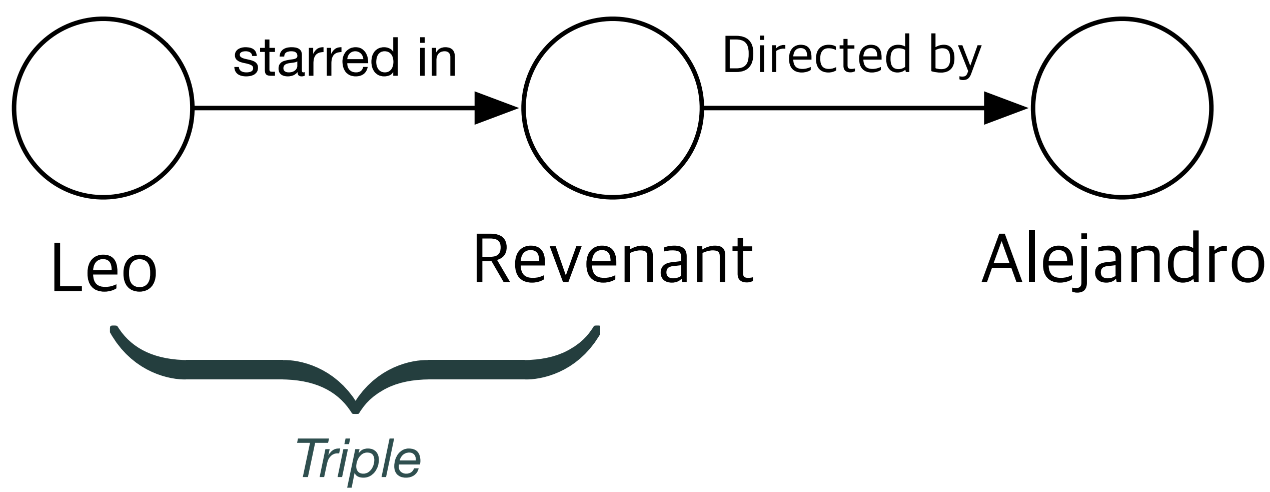
- Place a probability distribution over triples

$$x_{ijk} \sim \mathcal{N}(e_i^\top R_k e_j, \sigma_x^2)$$

- Measure **how likely a triple is positive or not**

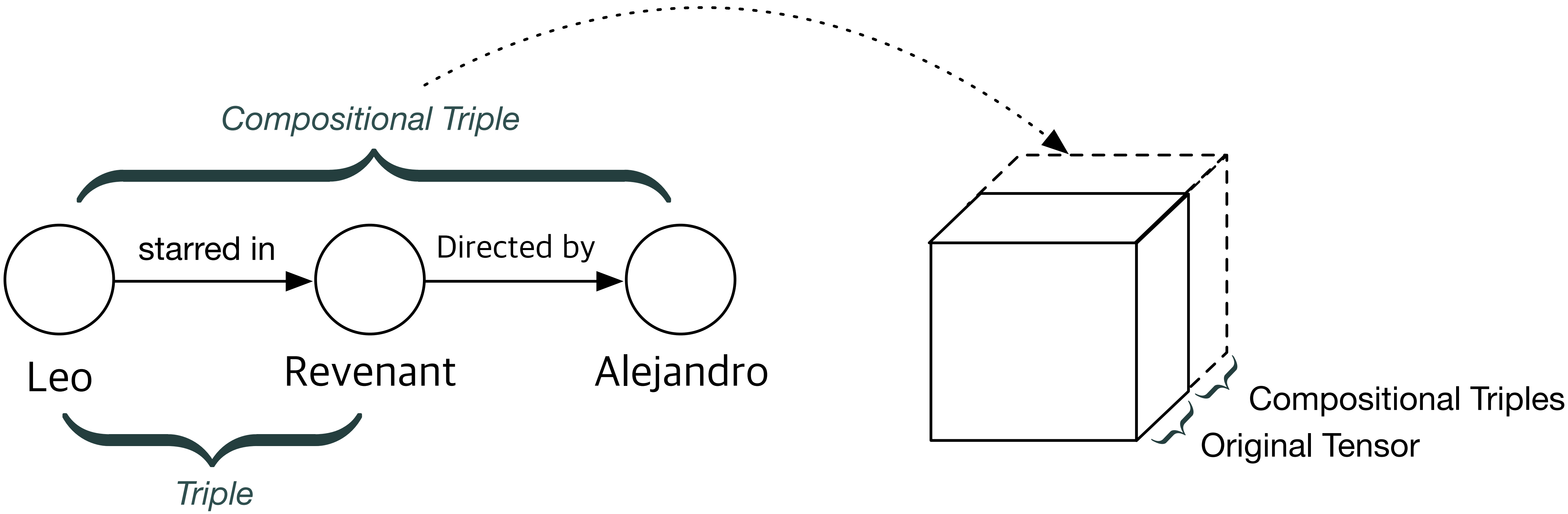
KNOWLEDGE COMPOSITION

- Another Limitation of RESCAL: The model ignores the **path information** on KG during factorisation



- But, the factorisation model ignores the compositional relation between Leonardo DiCaprio and Alejandro G. Iñárritu

KG AUGMENTED WITH COMPOSITIONAL TRIPLES



- We propose a probabilistic compositional RESCAL
- Additional data with compositional relations
- Explicitly include path information into factorisation

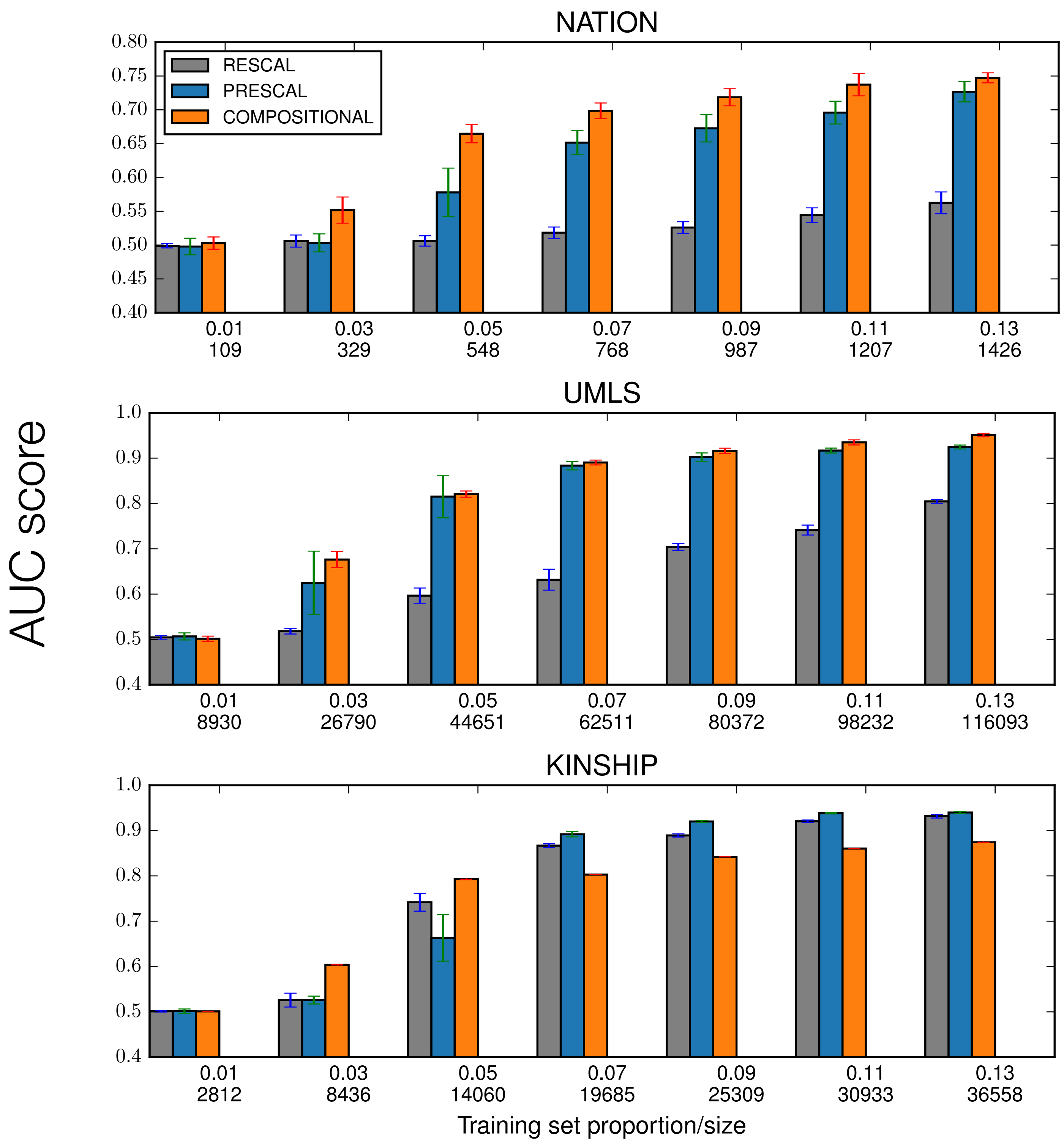
Tensor Factorisation on Real Datasets

	# Relations	# Entities	# Positvie Triples	Sparsity
KINSHIP	26	104	10790	0.038
UMLS	49	135	6752	0.008
NATION	56	14	2024	0.184

Datasets

- Split each dataset into training / testing sets, measure predictive performance of test set given training set
- Examples:
  - KINSHIP: <person A, father of, person B>
  - UMLS: <disease E, affects, function F>
  - NATION: <nation C, supports, nation D>

PREDICTION RESULT



- High AUC indicates better predictive performance
- Probabilistic models outperforms non-probabilistic model
- Compositional models outperform for the UMLS and NATION datasets

2. Completion with Human Experts

- Some domains may not have enough information to construct a KG (e.g. Bio-informatics, medical data)
  - Without enough known triples, RESCAL models do not work well
- Each unknown entry of KG may be uncovered by domain experts
  - Requires some amount of budget to uncover
  - Too many triples to be evaluated, e.g. UMLS = 900,000 triples
- Goal: how to **choose a triple to be evaluated next**
  - **to maximise the number of positive triples** given a limited budget

GREEDY INCREMENTAL COMPLETION WITH STATISTICAL MODEL

- Using prediction of bilinear model
- Greedy approach for incremental completion
  1. Infer latent semantics of KG given observed triples
  2. Query an unknown triple that has the highest expected value with the current approximation
  $$\arg\max_{ijk} \mathbb{E}[e_i^\top R_k e_j]$$
  3. Update late semantics based on the observation from queried result
  4. Repeat 2-3 until reach the budget limitation

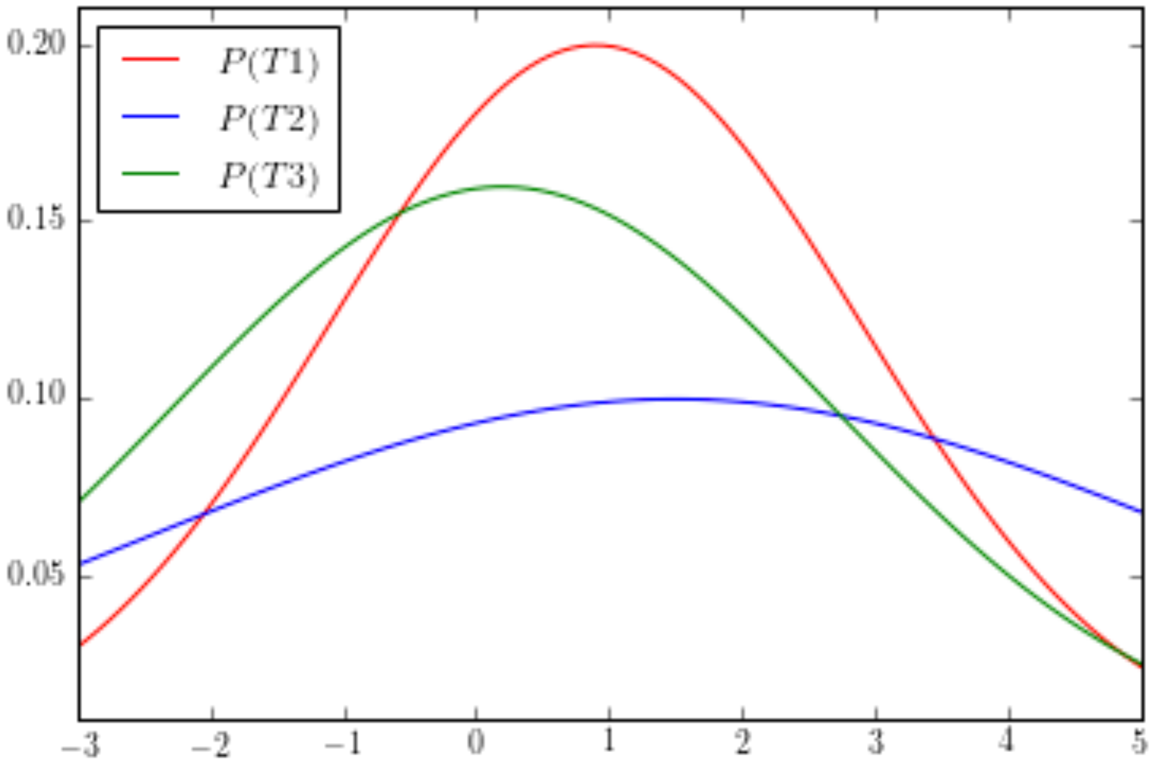
EXPLOITATION-EXPLORATION

- Greedy approach always query the maximum triple: may not efficiently explorer the latent space

$$\arg\max_{ijk} \mathbb{E}[e_i^\top R_k e_j]$$

- We employ randomised probability matching (RPM) algorithm (Thompson, 1933; a.k.a Thompson sampling)
  - Optimal balance between exploitation and exploration
  - Instead of finding maximum triple, consider the **uncertainty** of unknown triples

Randomised Probability Matching



T1 = <DiCaprio, starred in, The Revenant>

T2 = <Benjio, receive, Oscar>

T3 = <DiCaprio, profession, Actor>

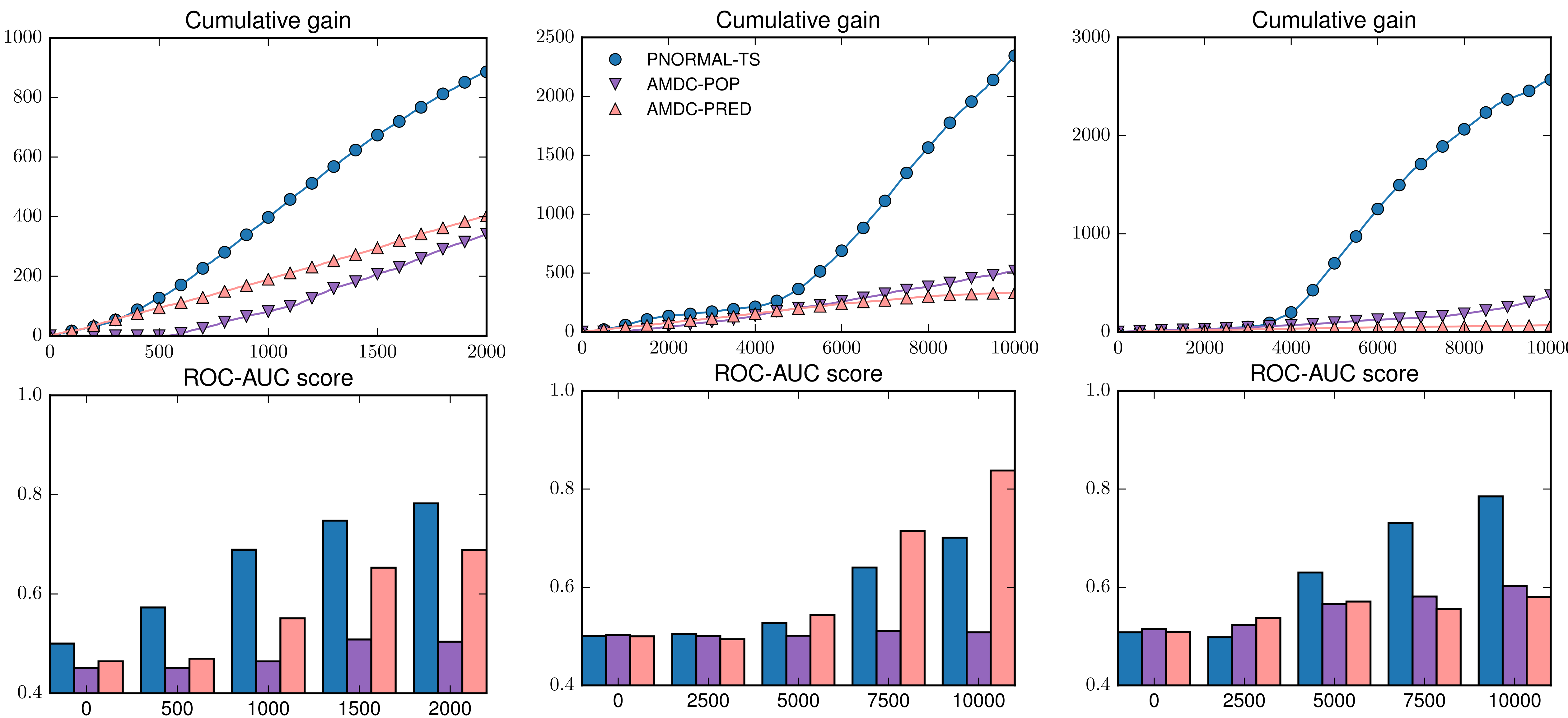
$$E[T_1] = 0.9 \quad E[T_2] = 1.5 \quad E[T_3] = 0.2$$

- Although T2 has the highest expectation, the variance of T2 is also high
- If we consider the **uncertainty of probability**, it might be better to choose T1
  - Chose a triple to be labelled based on a **random sample** from its distribution

EXPERIMENT: INCREMENTAL COMPLETION

- Comparison with two greedy algorithms (Kajino et al, 2015)
  - AMDC-POP: Greedy population model, always queries the highest expected triple
  - AMDC-PRED: Greedy prediction model, always queries the most ambiguous triple (expected value around 0.5)
- Performance metrics
  - Cumulative gain: **total number of positive triples** up to time t
  - AUC: predictive performance of model (used in the first task)

COMPARISON WITH GREEDY APPROACHES



NATION KINSHIP UMLS

SUMMARY

- Knowledge graph completion task can be divided into unknown triple prediction and incremental completion
- We develop probabilistic RESCAL, and incorporate compositional structure into factorisation
- In the prediction task, compositional triples help to predict unknown triples
- In the construction task, randomised probability matching efficiently uncovers the latent semantic of KG

INTUITION BEHIND BILINEAR MODEL

Indicator of good actor

relation between good actor and prestigious award in this relation

Leonardo DiCaprio

$$e_{\mathrm{D}} = \begin{bmatrix} 0.8 \\ 0.1 \end{bmatrix}$$

Receive Award

$$R_{\mathrm{RA}} = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.1 \end{bmatrix}$$

Oscar

$$e_{\mathrm{O}} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$

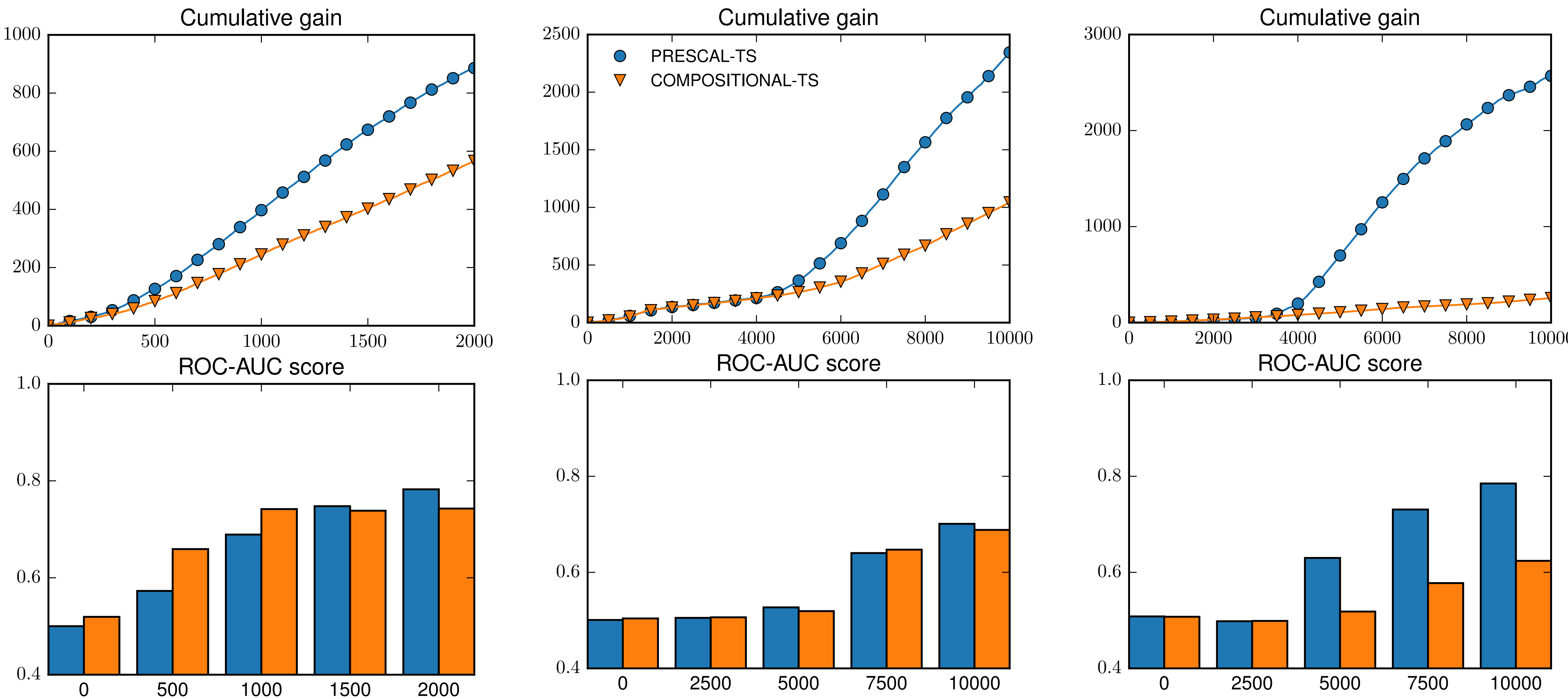
Yoshua Bengio

$$e_{\mathrm{B}} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

Indicator of prestigious award

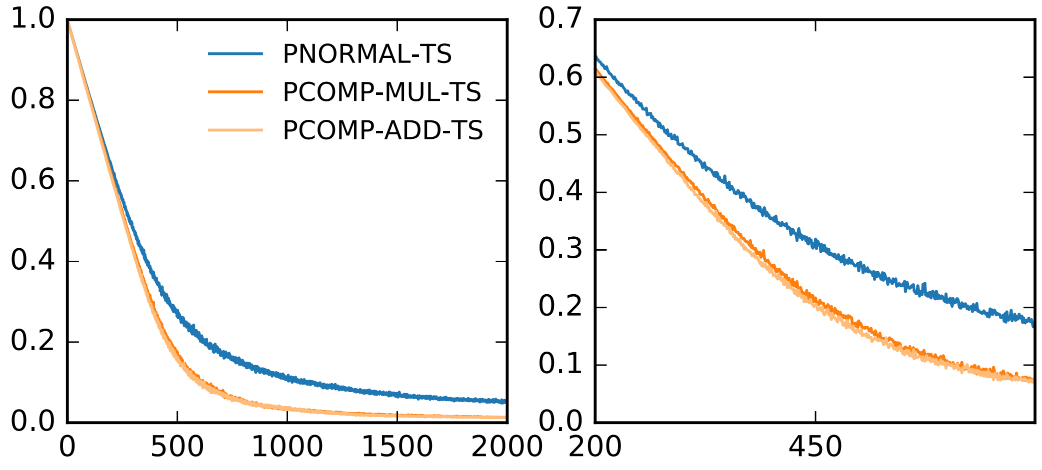
→ *Bilinear model infers latent semantics of entities and relations*

Comparison with Compositional Models

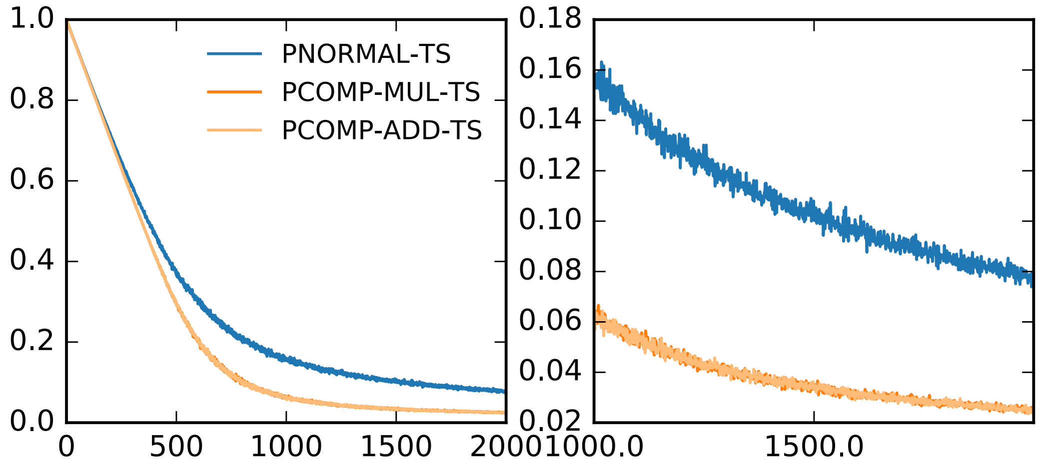


NATION KINSHIP UMLS

Posterior variance analysis



KINSHIP



UMLS

- Variance of compositional model shrinks faster than non-compositional model