EMNLP 2015
CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING
LISBON

CORE: Context-Aware Open Relation Extraction with Factorization Machines

Fabio Petroni

SAPIENZA
UNIVERSITÀ DI ROMA

Luciano Del Corro

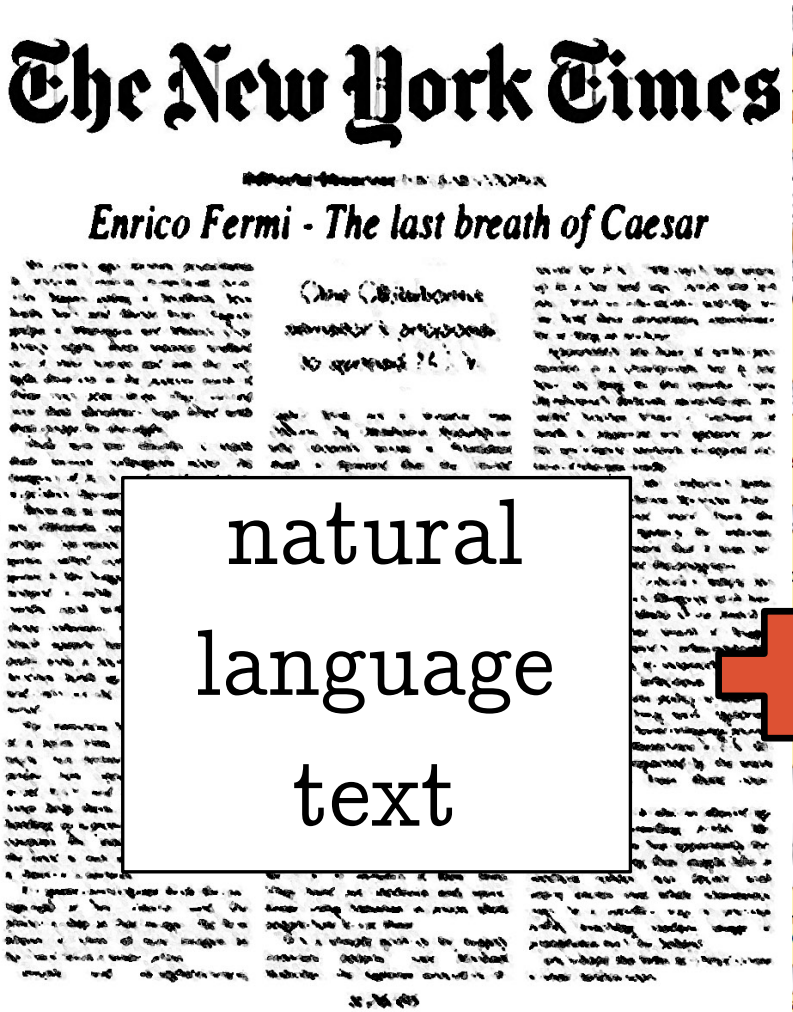
max planck institut informatik

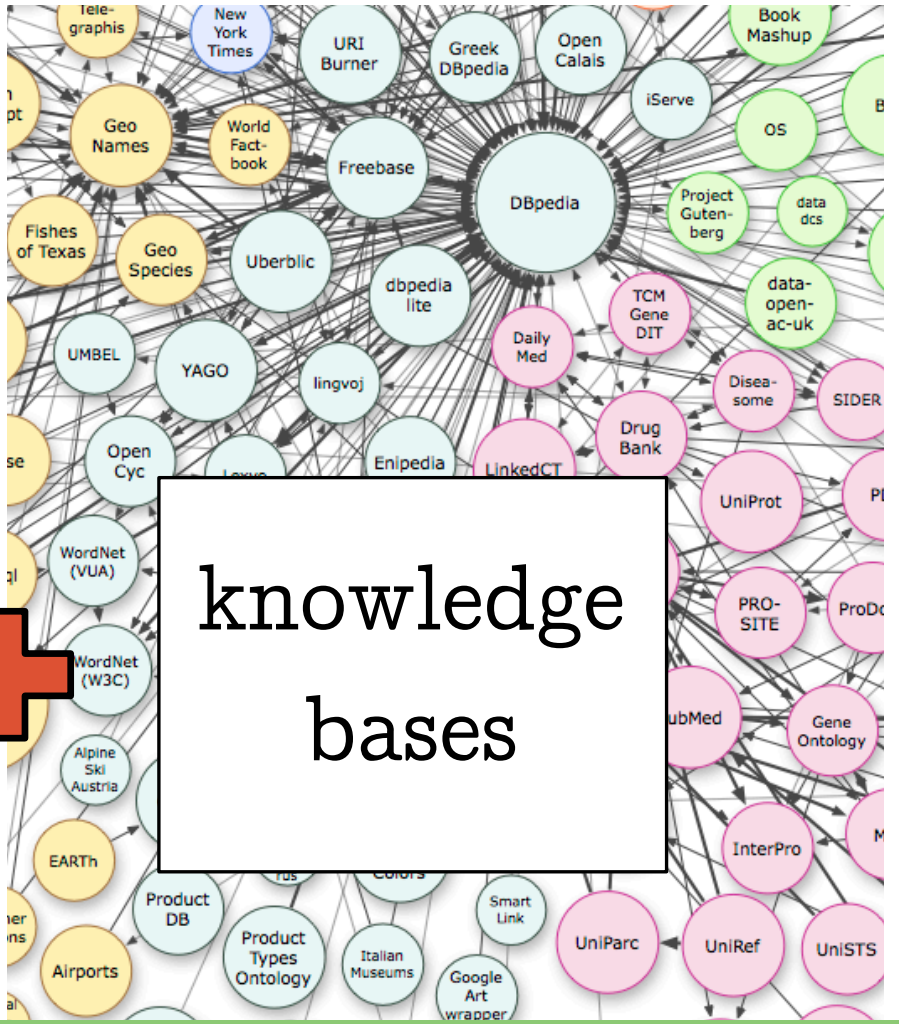
Rainer Gemulla

UNIVERSITY OF MANNHEIM

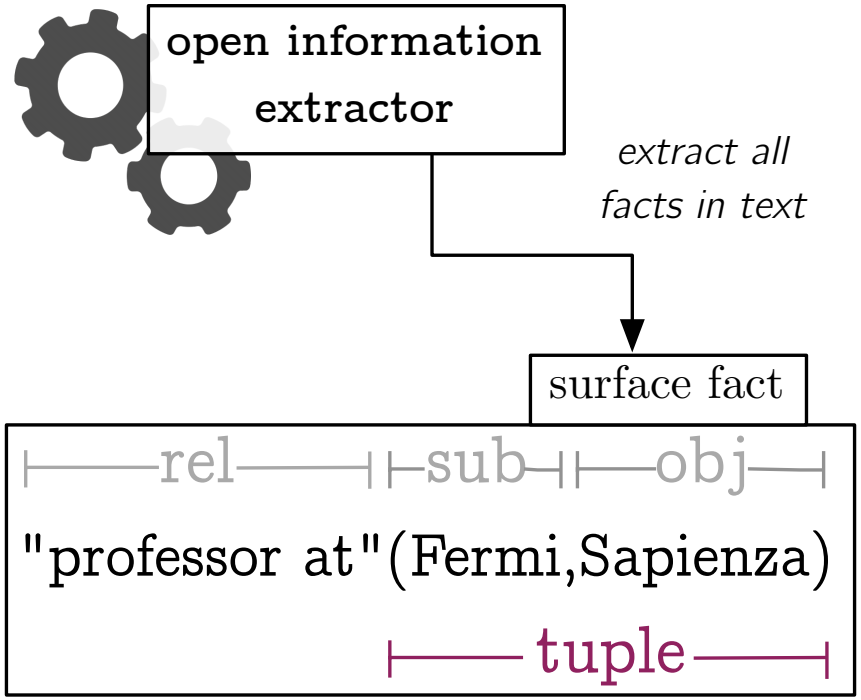
Open relation extraction

- Open relation extraction is the task of extracting new facts for a potentially unbounded set of relations from various sources

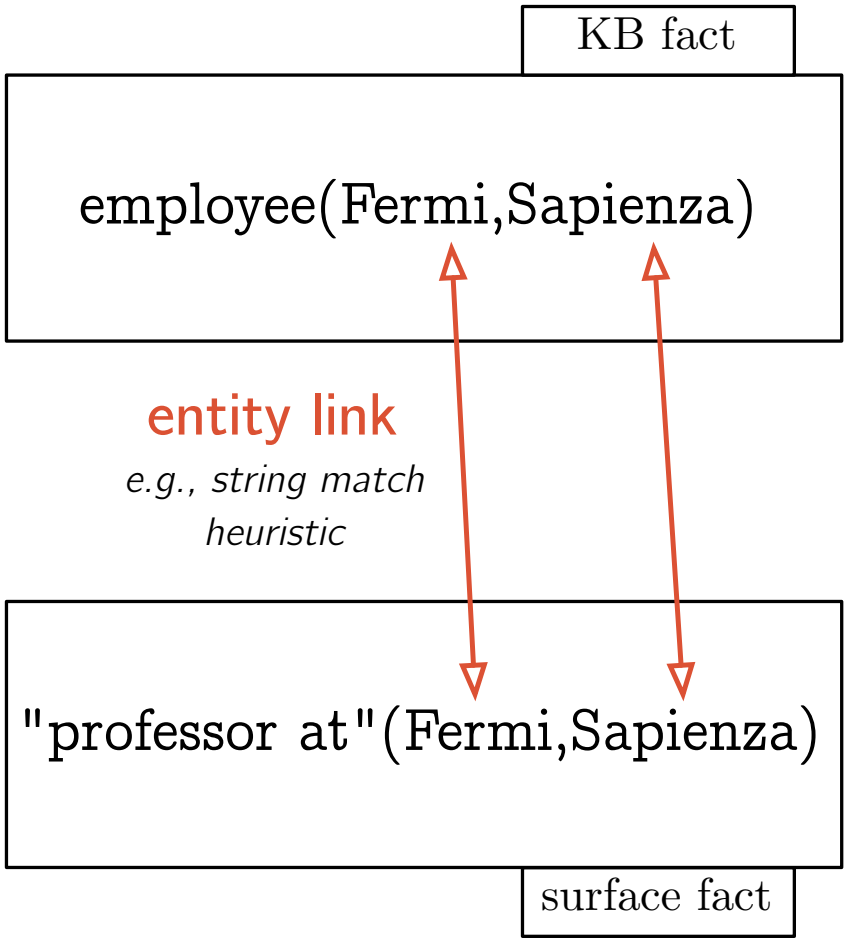


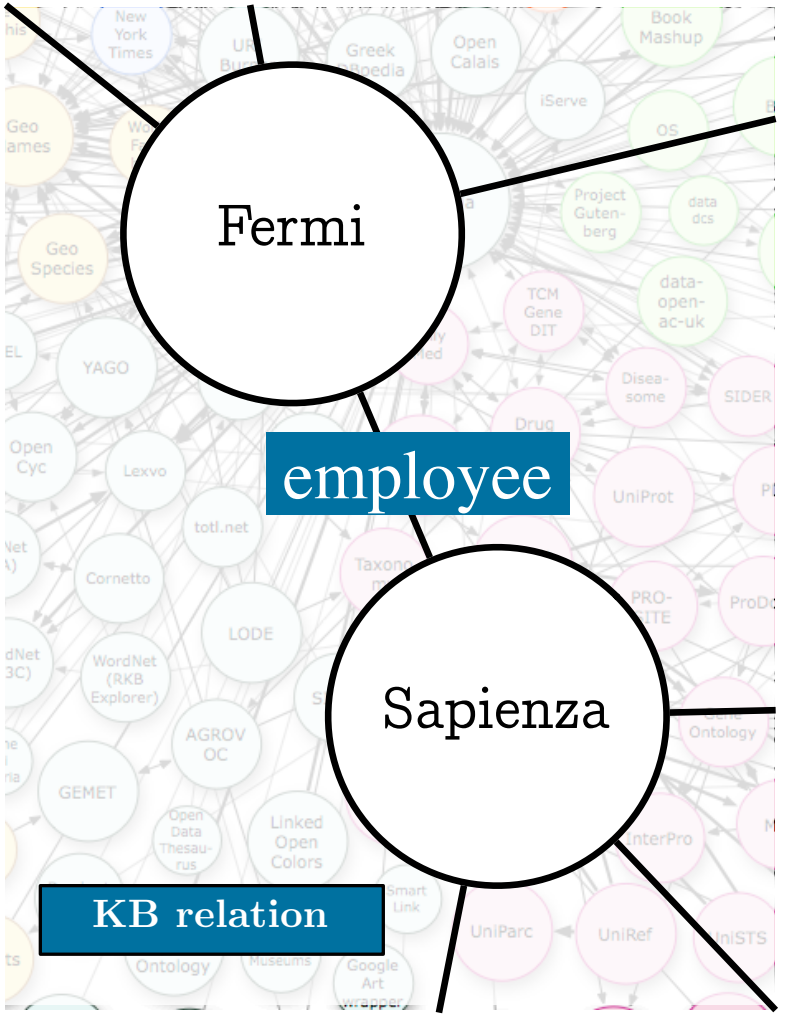


Input data: facts from natural language text

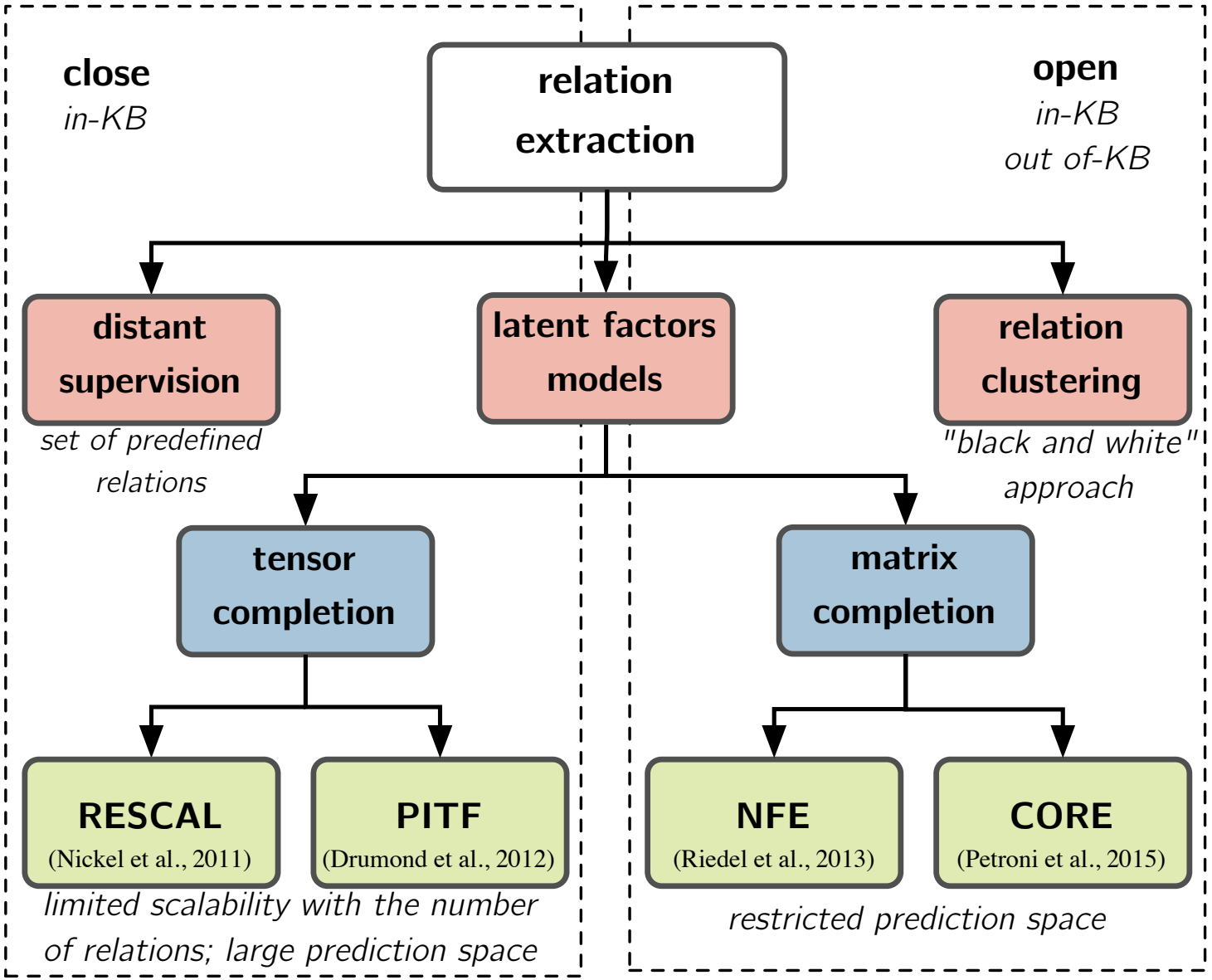


Input data: facts from knowledge bases

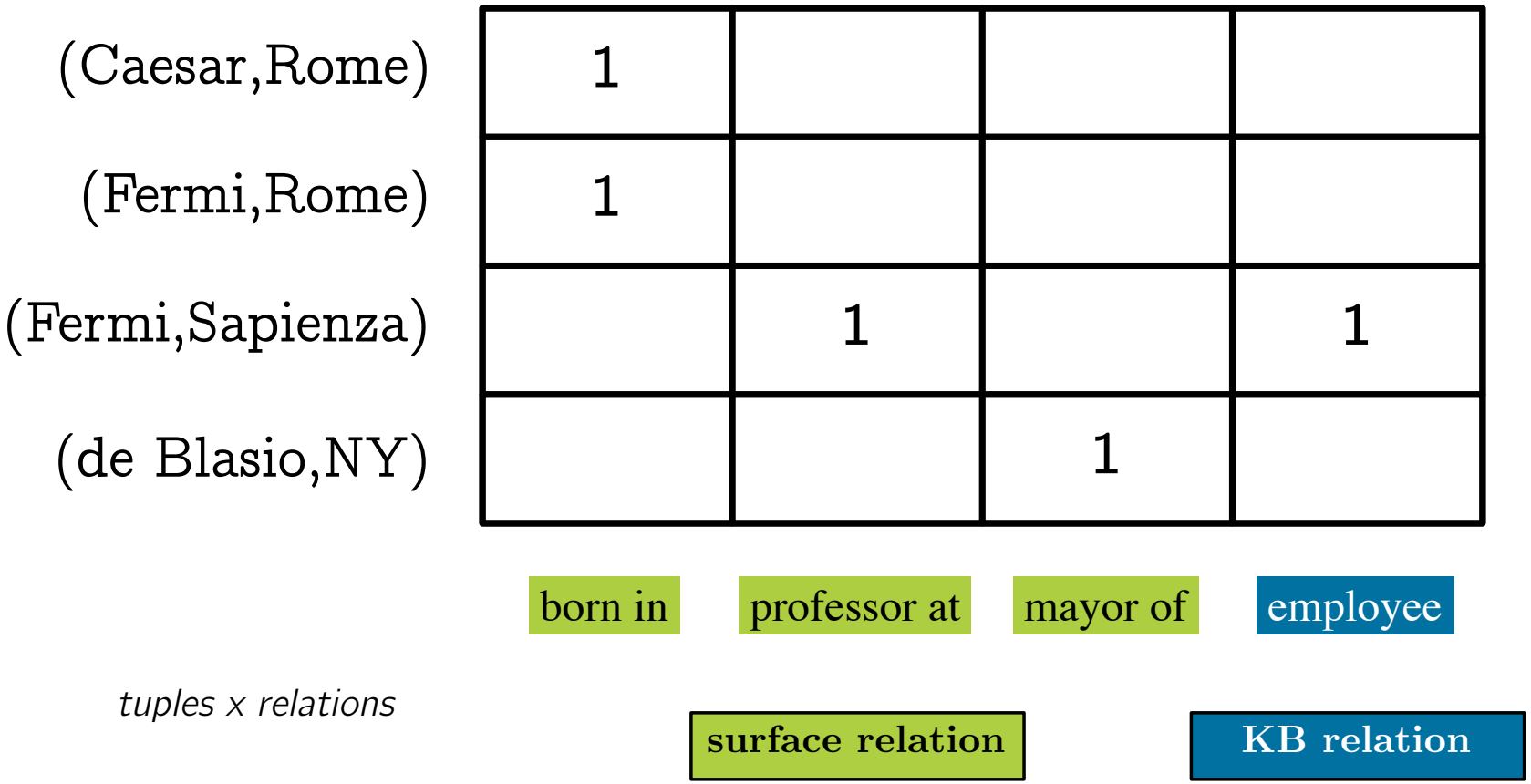




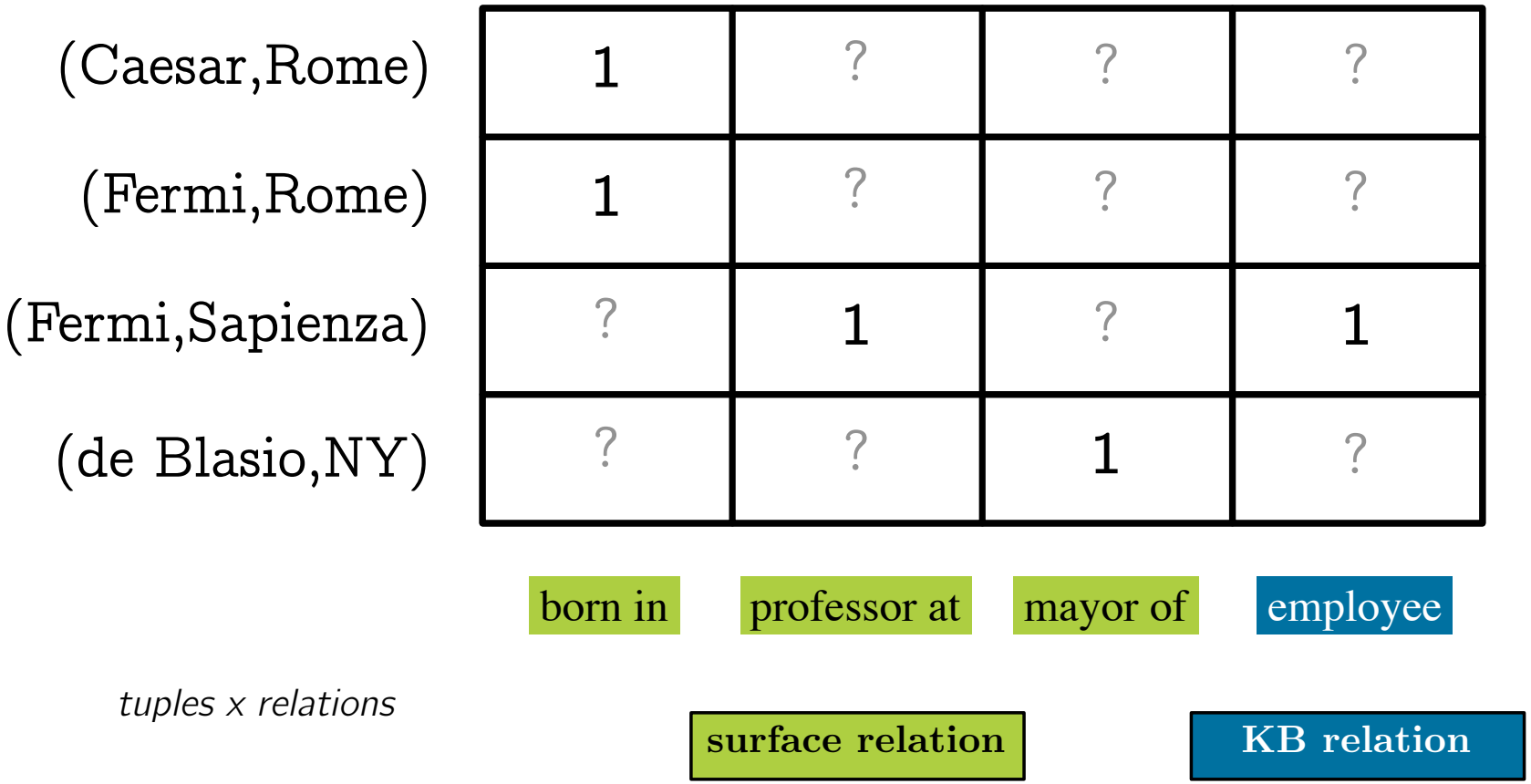
Relation extraction techniques taxonomy



Matrix completion for open relation extraction

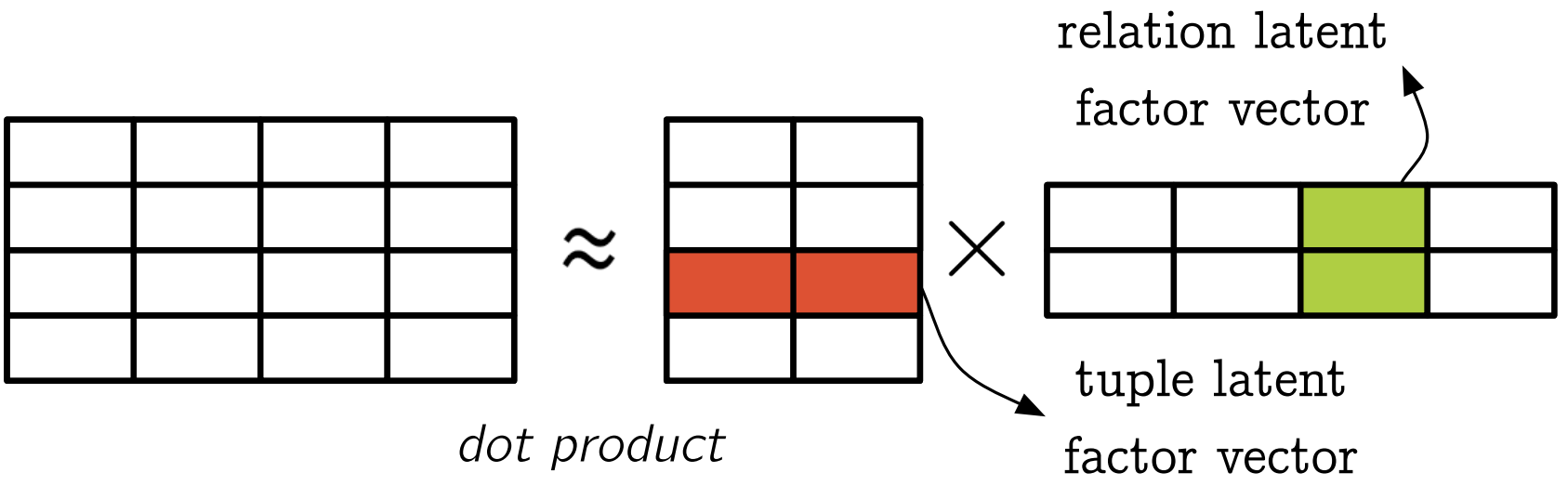


Matrix completion for open relation extraction

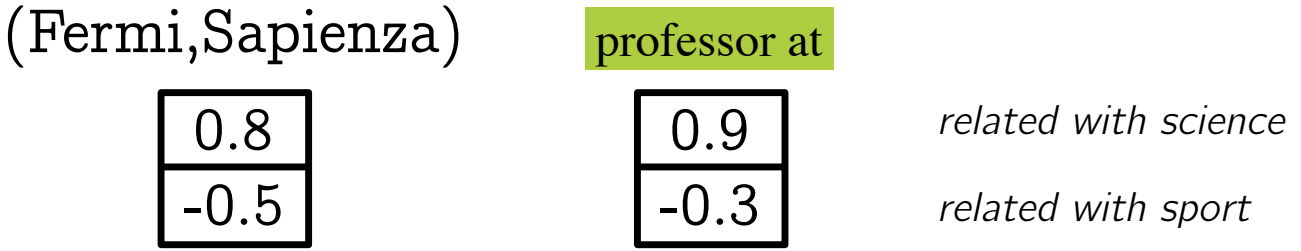


Matrix factorization

- learn latent semantic representations of tuples and relations



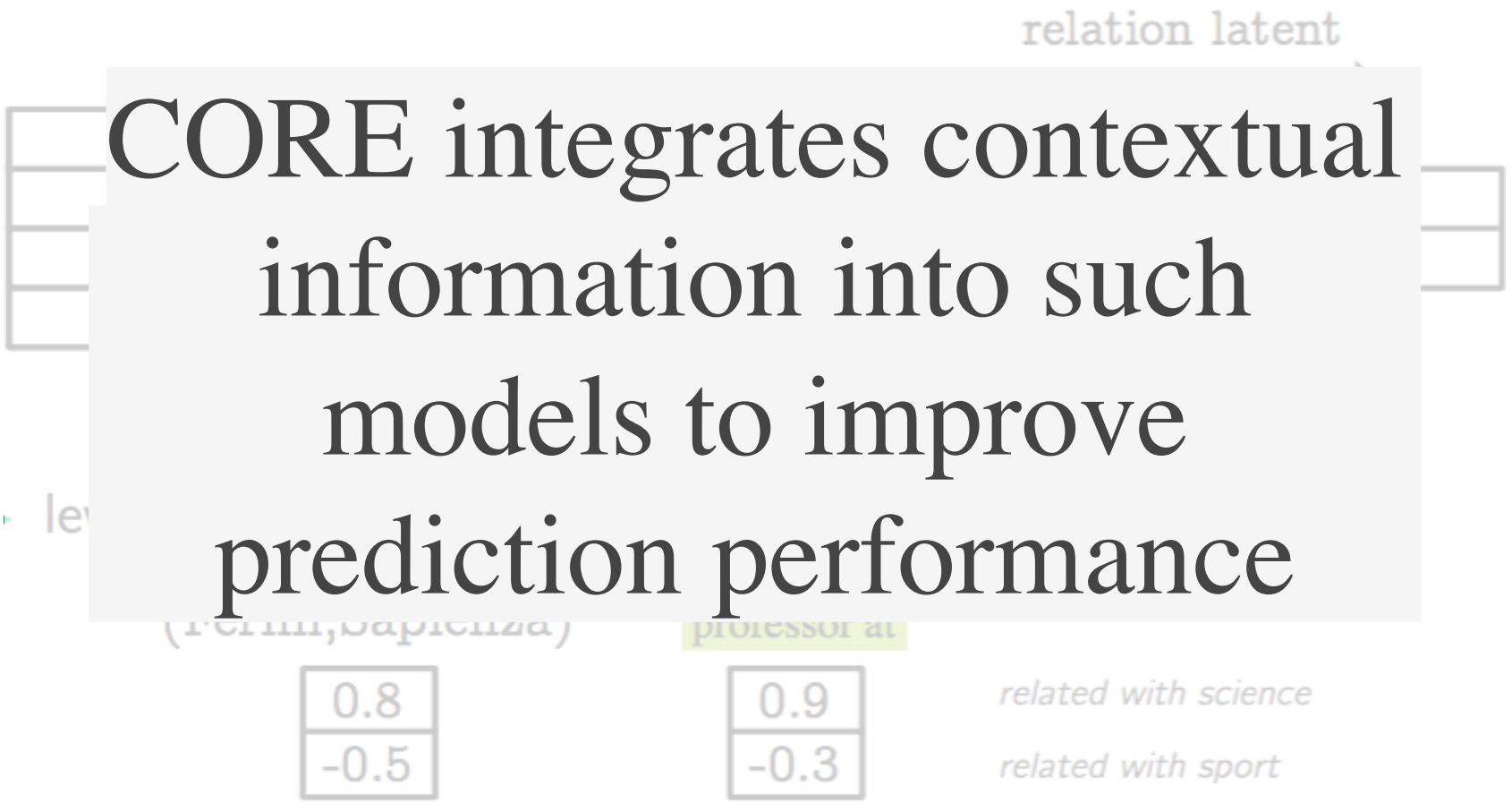
- leverage latent representations to predict new facts



- in real applications latent factors are uninterpretable

Matrix factorization

- learn latent semantic representations of tuples and relations



CORE integrates contextual information into such models to improve prediction performance

- in real applications latent factors are uninterpretable

Contextual information

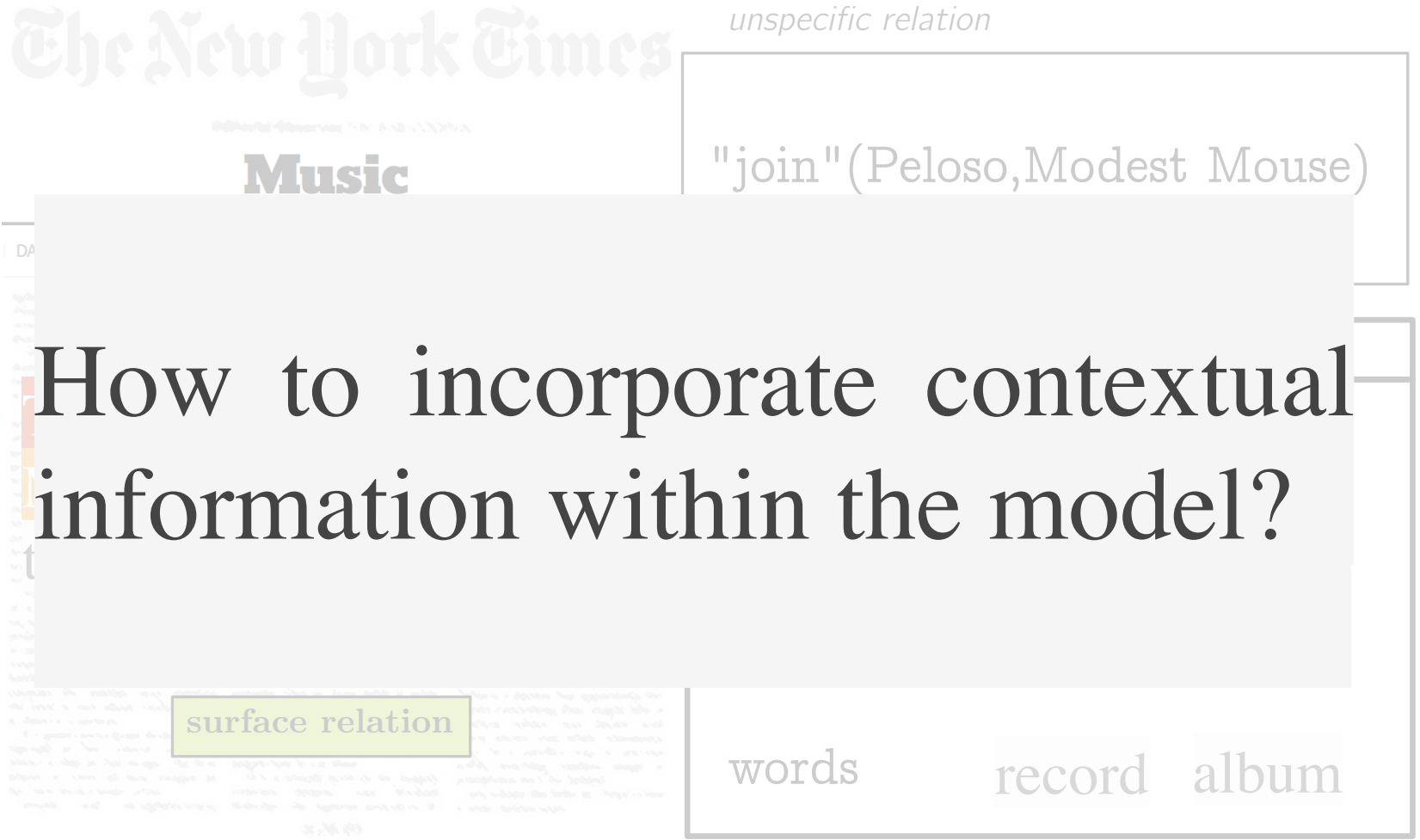
unspecific relation

"join"(Peloso,Modest Mouse)

Contextual information

entity types	person, organization
article topic	**Music**
words	record, album

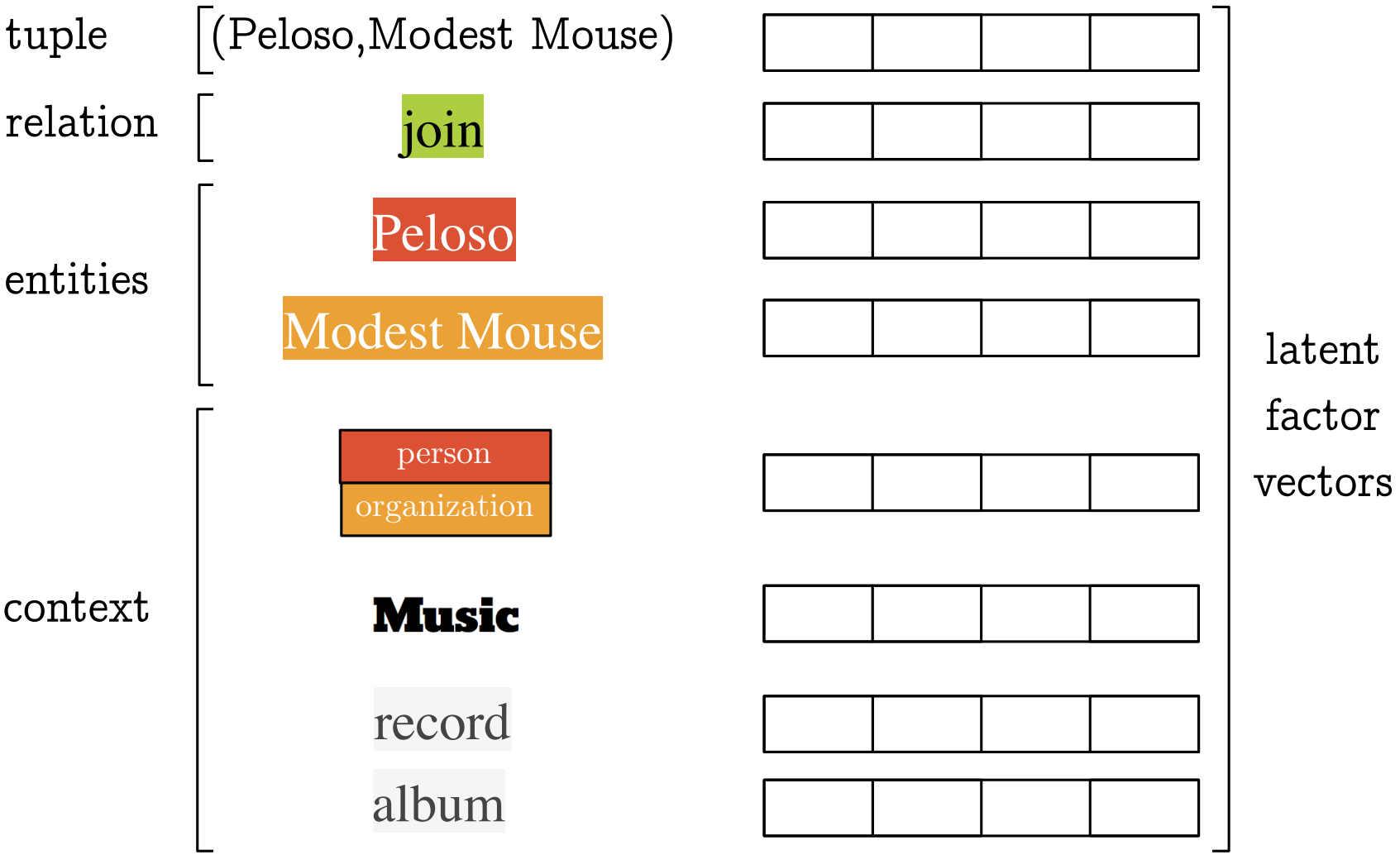
Contextual information



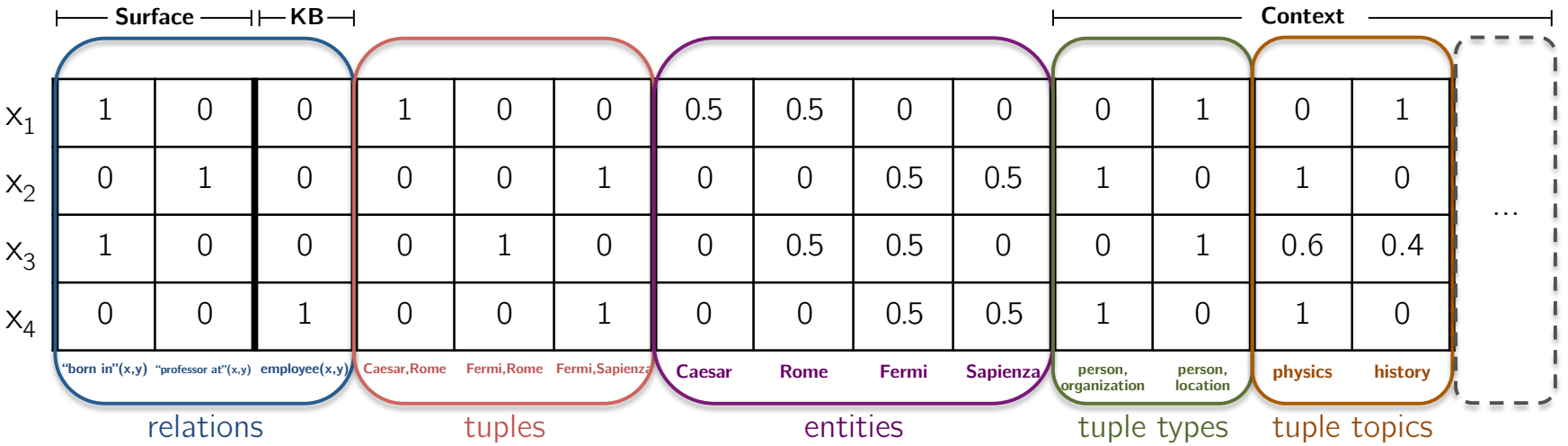
How to incorporate contextual information within the model?

CORE - latent representations of variables

- associates latent representations f_v with each variable $v \in V$

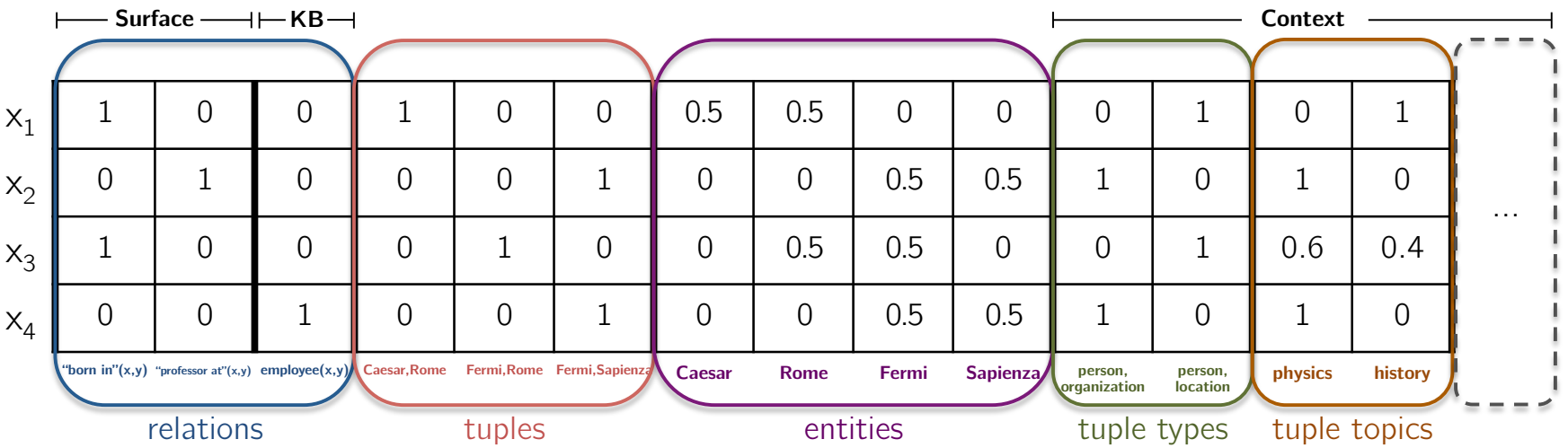


CORE - modeling facts



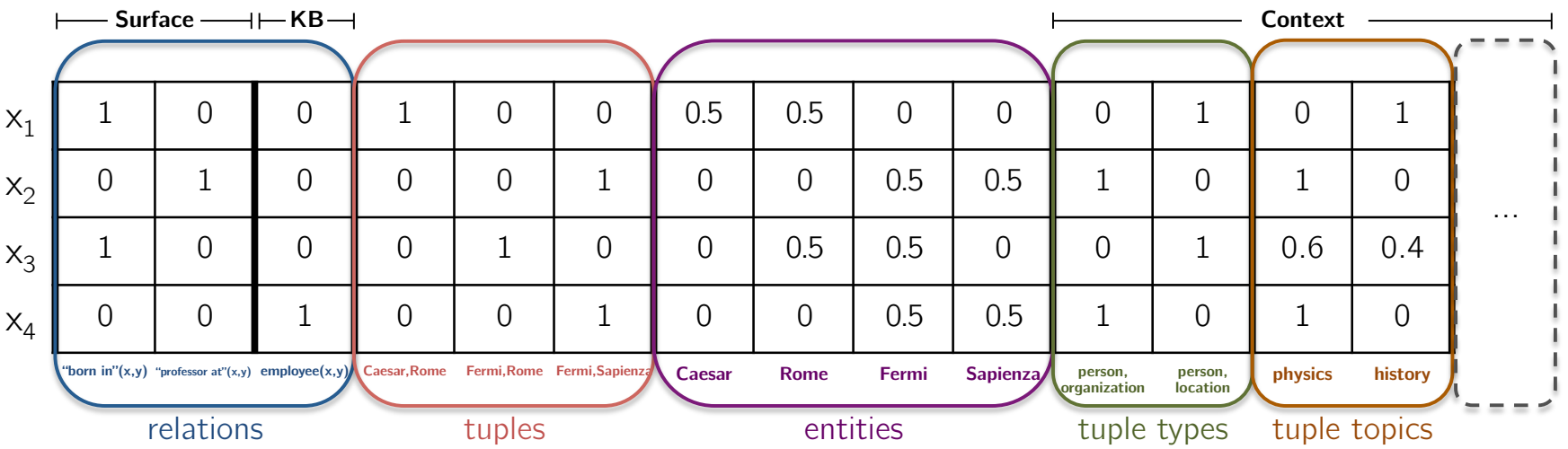
- models the input data in terms of a matrix in which each row corresponds to a fact x and each column to a variable v
- groups columns according to the type of the variables
- in each row the values of each column group sum up to unity

CORE - modeling context



- aggregates and normalizes contextual information by tuple
  - a fact can be observed multiple times with different context
  - there is no context for new facts (never observed in input)
- this approach allows us to provide comprehensive contextual information for both observed and unobserved facts

CORE - factorization model



- uses factorization machines as underlying framework
- associates a score $s(x)$ with each fact x

$$s(x) = \sum_{v_1 \in V} \sum_{v_2 \in V \setminus \{v_1\}} x_{v_1} x_{v_2} \boldsymbol{f}_{v_1}^T \boldsymbol{f}_{v_2}$$

- weighted pairwise interactions of latent factor vectors

CORE - prediction

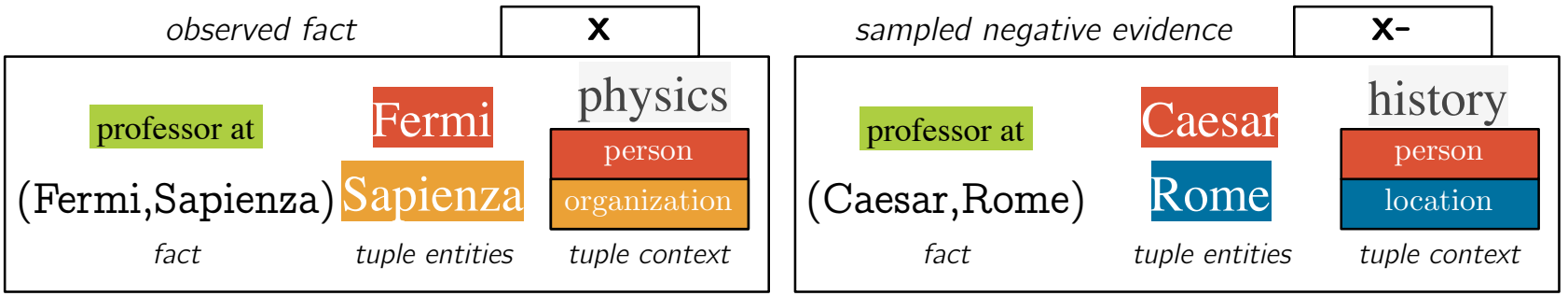
goal

produce a ranked list of tuples for each relation

- rank reflects the likelihood that the corresponding fact is true
- to generate this ranked list:
  - fix a relation **r**
  - retrieve all tuples **t**, s.t. the fact **r(t)** is not observed
  - add tuple context
  - rank unobserved facts by their scores

CORE - parameter estimation

- parameters: $\Theta = \{ b_v, \boldsymbol{f}_v \mid v \in V \}$
- all our observations are positive, no negative training data
- Bayesian personalized ranking, open-world assumption



- pairwise approach, **x** is more likely to be true than **x-**

$$\text{maximize} \sum_x f(s(\mathbf{x}) - s(\mathbf{x}\text{-}))$$

- stochastic gradient ascent

$$\Theta \leftarrow \Theta + \eta \nabla_\Theta \left(\quad \right)$$

Experiments - dataset

440k facts extracted from The New York Times corpus

entity mentions linked using string matching

15k facts from Freebase

- Contextual information

article metadata

news desk (e.g., foreign desk)
descriptors (e.g., finances)
online section (e.g., sports)
section (e.g., a, d)
publication year

m

entity type

person
organization
location
miscellaneous

t

bag-of-word

sentences where the fact has been extracted

w

- letters to indicate contextual information considered

Experiments - methodology

- we consider (to keep experiments feasible):

10k tuples	**19** Freebase relations	**10** surface relations

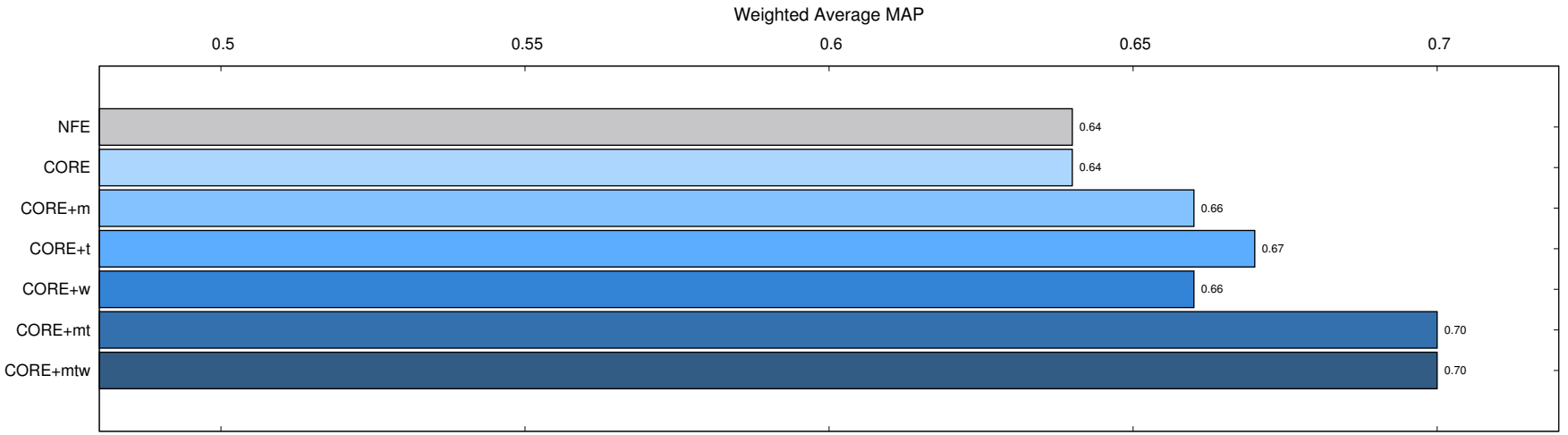
- for each relation and method:
  - we rank the tuples subsample
  - we consider the top-100 predictions and label them manually
- evaluation metrics:

number of true facts	**MAP** (quality of the ranking)

- methods:
  - **PITF**, tensor factorization method
  - **NFE**, matrix completion method (context-agnostic)
  - **CORE**, uses relations, tuples and entities as variables
  - **CORE**+m, +t, +w, +mt, +mtw

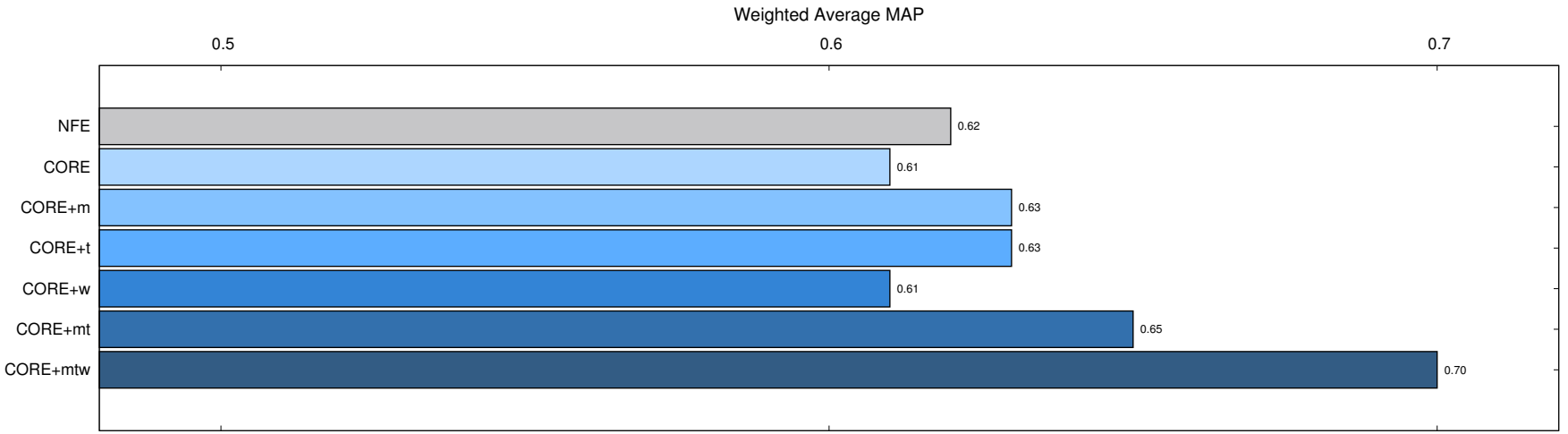
Results - Freebase relations

Relation	#	PITF	NFE	CORE	CORE+m	CORE+t	CORE+w	CORE+mt	CORE+mtw
person/company	208	70 (0.47)	92 (0.81)	91 (0.83)	90 (0.84)	91 (0.87)	92 (0.87)	95 (0.93)	**96** (**0.94**)
person/place_of_birth	117	1 (0.0)	92 (*0.9*)	90 (0.88)	92 (*0.9*)	92 (*0.9*)	89 (0.87)	**93** (*0.9*)	92 (*0.9*)
location/containedby	102	7 (0.0)	63 (0.47)	62 (0.47)	63 (0.46)	61 (0.47)	61 (0.44)	62 (0.49)	**68** (**0.55**)
parent/child	88	9 (0.01)	64 (0.6)	64 (0.56)	64 (0.59)	64 (0.62)	64 (0.57)	67 (**0.67**)	**68** (0.63)
person/place_of_death	71	1 (0.0)	67 (0.93)	67 (0.92)	*69* (*0.94*)	67 (0.93)	67 (0.92)	*69* (*0.94*)	67 (0.92)
person/parents	67	20 (0.1)	51 (0.64)	52 (0.62)	51 (0.61)	49 (0.64)	47 (0.6)	*53* (**0.67**)	*53* (0.65)
author/works_written	65	24 (0.08)	45 (0.59)	49 (0.62)	51 (0.69)	50 (0.68)	50 (0.68)	51 (**0.7**)	**52** (0.67)
person/nationality	61	21 (0.08)	25 (0.19)	27 (0.17)	28 (0.2)	26 (0.2)	**29** (0.19)	27 (0.18)	27 (**0.21**)
neighbor./neighborhood_of	39	3 (0.0)	24 (0.44)	23 (0.45)	26 (0.5)	27 (0.47)	27 (0.49)	*30* (0.51)	*30* (**0.52**)
...									
Average $MAP^{100}_{\#}$		0.09	0.46	0.47	0.49	0.47	0.49	0.49	**0.51**
Weighted Average $MAP^{100}_{\#}$		0.14	0.64	0.64	0.66	0.67	0.66	*0.70*	*0.70*



Results - surface relations

Relation	#	PITF	NFE	CORE	CORE+m	CORE+t	CORE+w	CORE+mt	CORE+mtw
head	162	34 (0.18)	80 (0.66)	83 (0.66)	82 (0.63)	76 (0.57)	77 (0.57)	83 (0.69)	**88** (**0.73**)
scientist	144	44 (0.17)	76 (0.6)	74 (0.55)	73 (0.56)	74 (0.6)	73 (0.59)	*78* (0.66)	*78* (**0.69**)
base	133	10 (0.01)	85 (0.71)	86 (0.71)	86 (0.78)	88 (0.79)	85 (0.75)	83 (0.76)	**89** (**0.8**)
visit	118	4 (0.0)	73 (0.6)	75 (0.61)	76 (0.64)	80 (0.68)	74 (0.64)	75 (0.66)	**82** (**0.74**)
attend	92	11 (0.02)	65 (0.58)	64 (0.59)	65 (0.63)	62 (0.6)	66 (0.63)	62 (0.58)	**69** (**0.64**)
adviser	56	2 (0.0)	42 (0.56)	**47** (0.58)	44 (0.58)	43 (0.59)	45 (*0.63*)	43 (0.53)	44 (*0.63*)
criticize	40	5 (0.0)	31 (0.66)	33 (0.62)	33 (**0.7**)	33 (0.67)	33 (0.61)	35 (0.69)	**37** (0.69)
support	33	3 (0.0)	19 (0.27)	22 (*0.28*)	18 (0.21)	19 (*0.28*)	22 (0.27)	**23** (0.27)	21 (0.27)
praise	5	0 (0.0)	2 (0.0)	2 (0.01)	4 (*0.03*)	3 (0.01)	3 (0.02)	**5** (*0.03*)	2 (0.01)
vote	3	2 (0.01)	*3* (0.63)	*3* (0.63)	*3* (0.32)	*3* (0.49)	*3* (0.51)	*3* (0.59)	*3* (**0.64**)
Average $MAP^{100}_{\#}$		0.04	0.53	0.53	0.51	0.53	0.53	0.55	**0.59**
Weighted Average $MAP^{100}_{\#}$		0.08	0.62	0.61	0.63	0.63	0.61	0.65	**0.70**



Anecdotal results

author(x,y)

ranked list of tuples
1 (Winston Groom, Forrest Gump)
2 (D. M. Thomas, White Hotel)
3 (Roger Rosenblatt, Life Itself)
4 (Edmund White, Skinned Alive)
5 (Peter Manso, Brando: The Biography)

similar relations
0.98 "reviews x by y"(x,y)
0.97 "book by"(x,y)
0.95 "author of"(x,y)
0.95 " 's novel"(x,y)
0.95 " 's book"(x,y)

"scientist at"(x,y)

ranked list of tuples
1 (Riordan Roett, Johns Hopkins University)
2 (Dr. R. M. Roberts, University of Missouri)
3 (Linda Mayes, Yale University)
4 (Daniel T. Jones, Cardiff Business School)
5 (Russell Ross, University of Iowa)

similar relations
0.87 "scientist"(x,y)
0.84 "scientist with"(x,y)
0.80 "professor at"(x,y)
0.79 "scientist for"(x,y)
0.78 "neuroscientist at"(x,y)

- semantic similarity of relations is one aspect of our model
- similar relations treated differently in different contexts

Conclusion

- CORE, a matrix factorization model for open relation extraction that incorporates contextual information
- based on factorization machines and open-world assumption
- extensible model, additional contextual information can be integrated when available
- experimental study suggests that exploiting context can significantly improve prediction performance
- Source code, datasets, and supporting material are available at https://github.com/fabiopetroni/CORE

Thank you!

Questions?

Fabio Petroni

Sapienza University of Rome, Italy

Current position:

PhD Student in Engineering in Computer Science

Research Interests:

data mining, machine learning, big data

petroni@dis.uniroma1.it