Relation Extraction with Matrix Factorization

Sebastian Riedel (University College London)

Computer Science *Department*
Statistics *Department*
Gatsby *Unit*

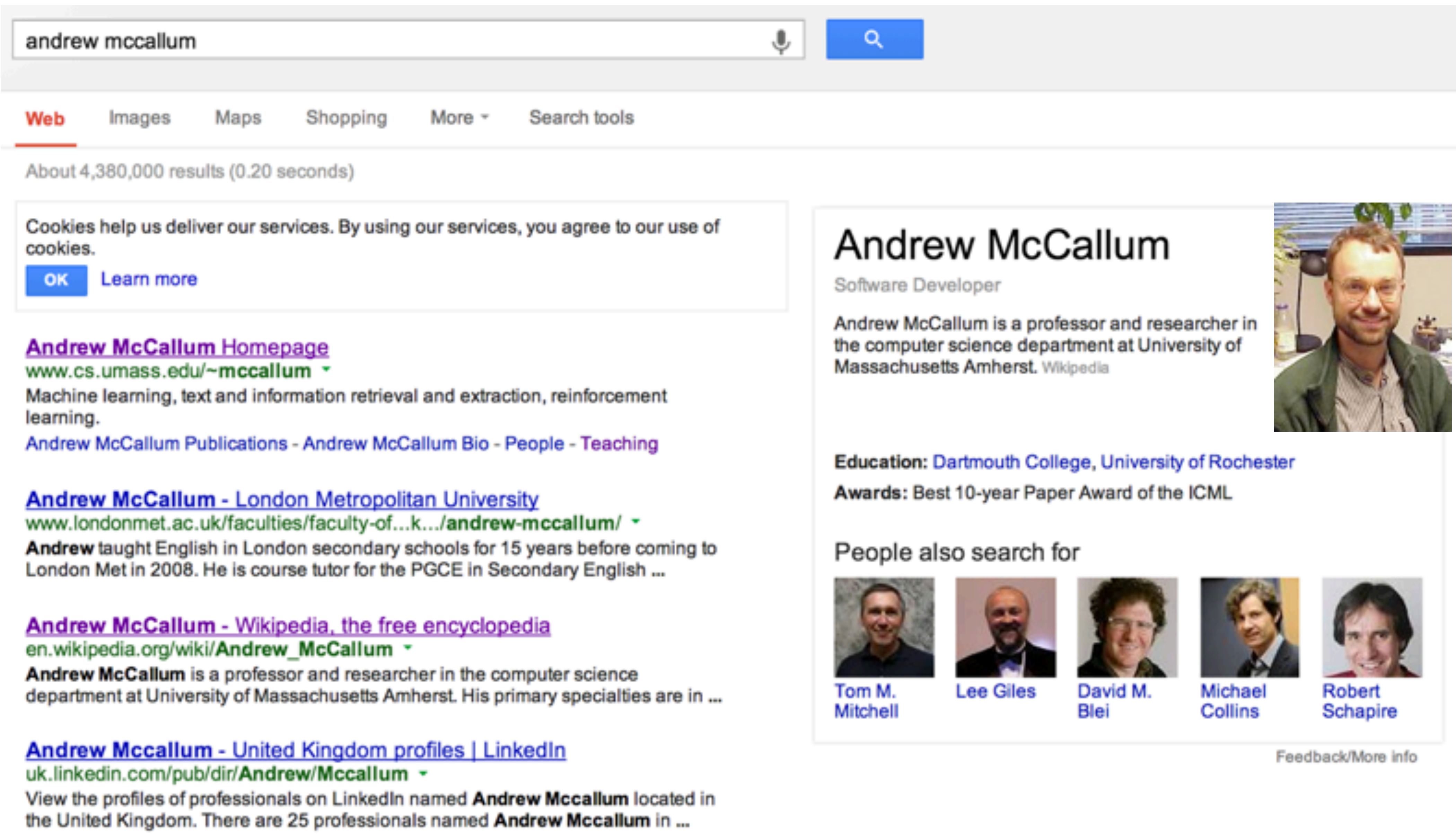
Contributors

Limin Yao
UMass Amherst

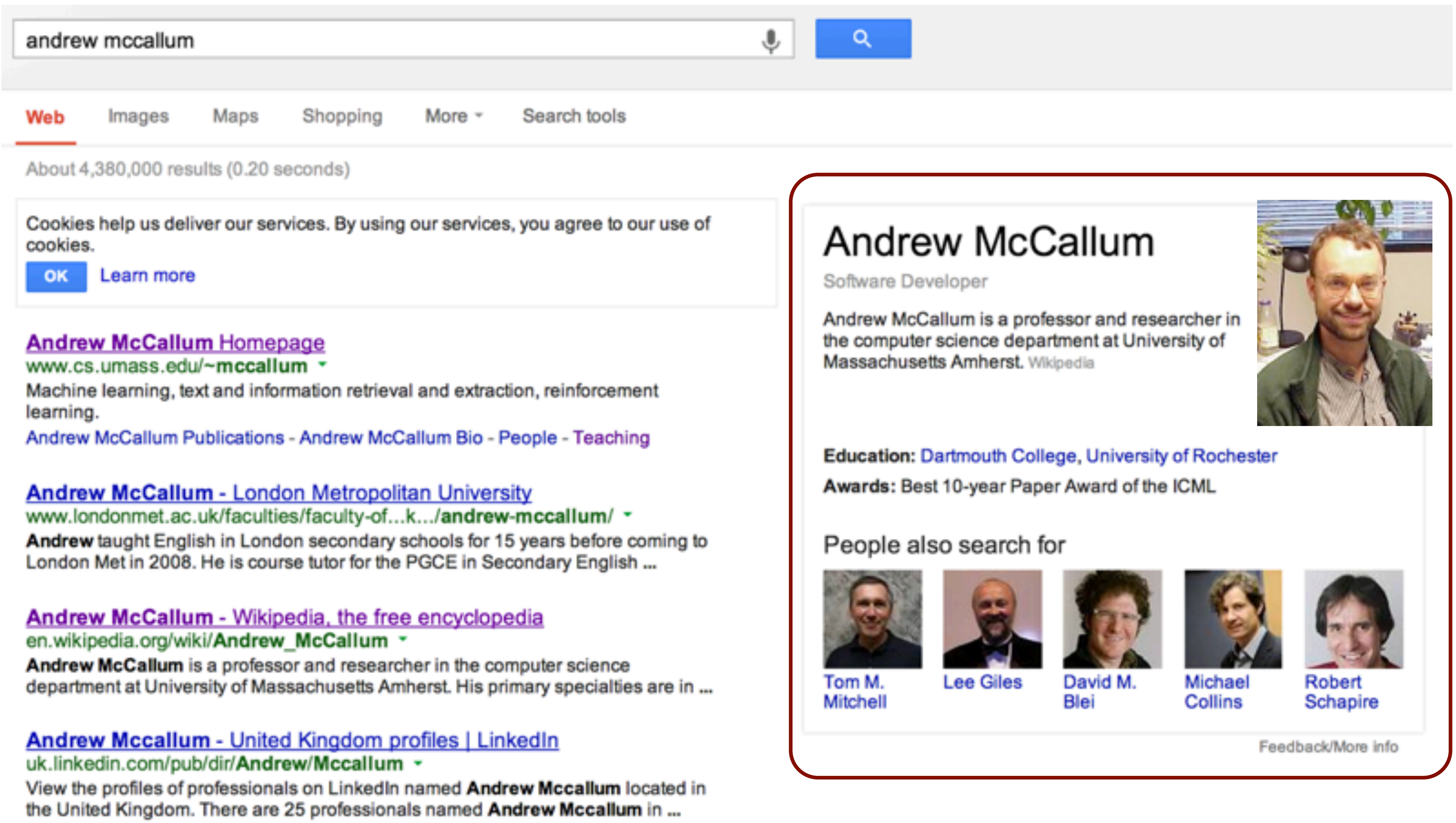
Andrew McCallum
UMass Amherst

Ben Marlin
UMass Amherst

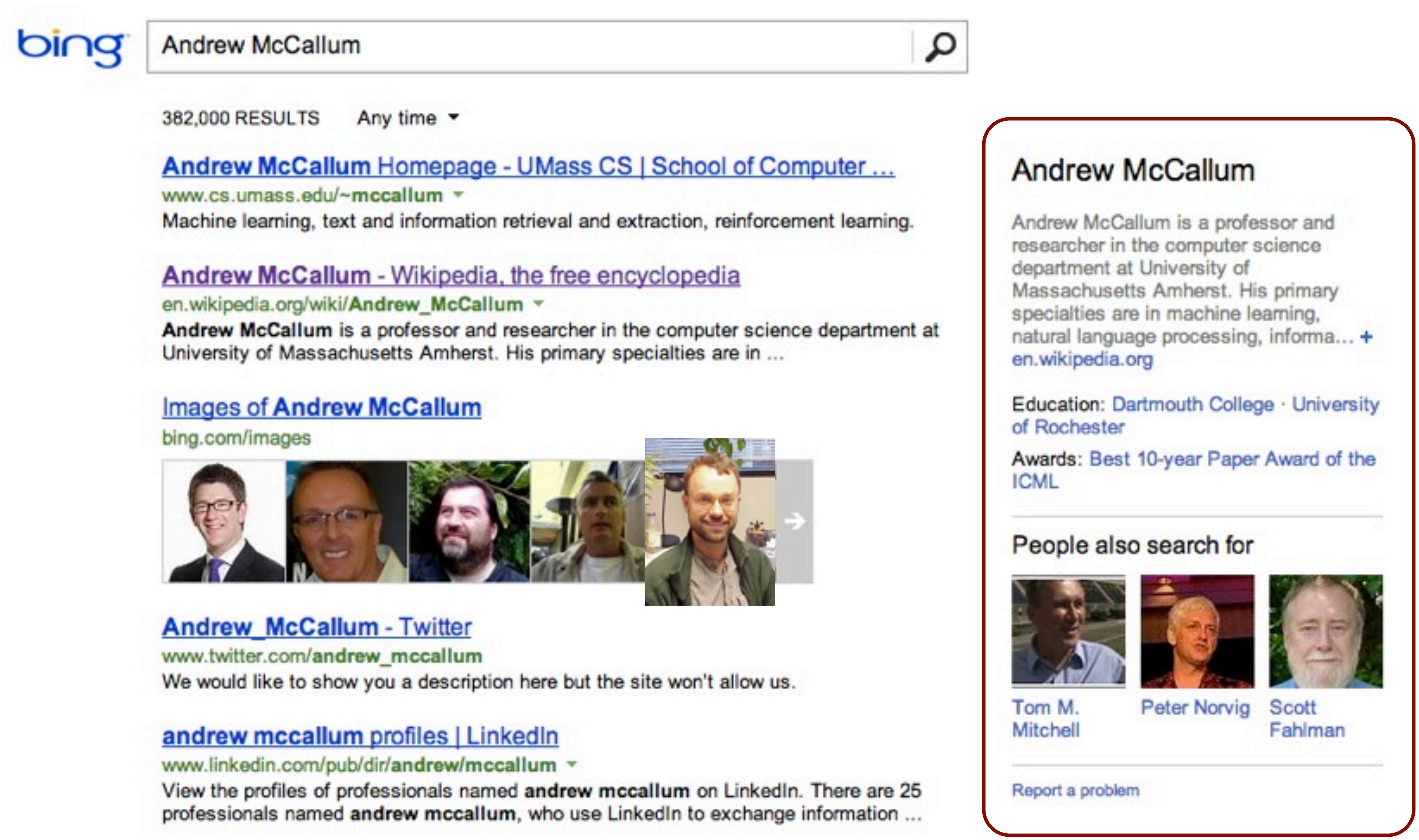
Motivation


andrew mccallum
Web
Images
Maps
Shopping
More
Search tools
About 4,380,000 results (0.20 seconds)
Cookies help us deliver our services. By using our services, you agree to our use of cookies.
OK
Learn more
Andrew McCallum Homepage
www.cs.umass.edu/~mccallum
Machine learning, text and information retrieval and extraction, reinforcement learning.
Andrew McCallum Publications - Andrew McCallum Bio - People - Teaching
Andrew McCallum - London Metropolitan University
www.londonmet.ac.uk/faculties/faculty-of...k.../andrew-mccallum/
Andrew taught English in London secondary schools for 15 years before coming to London Met in 2008. He is course tutor for the PGCE in Secondary English ...
Andrew McCallum - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Andrew_McCallum
Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. His primary specialties are in ...
Andrew Mccallum - United Kingdom profiles | LinkedIn
uk.linkedin.com/pub/dir/Andrew/Mccallum
View the profiles of professionals on LinkedIn named Andrew Mccallum located in the United Kingdom. There are 25 professionals named Andrew Mccallum in ...
Andrew McCallum
Software Developer
Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. Wikipedia
Education: Dartmouth College, University of Rochester
Awards: Best 10-year Paper Award of the ICML
People also search for
Tom M. Mitchell
Lee Giles
David M. Blei
Michael Collins
Robert Schapire
Feedback/More info

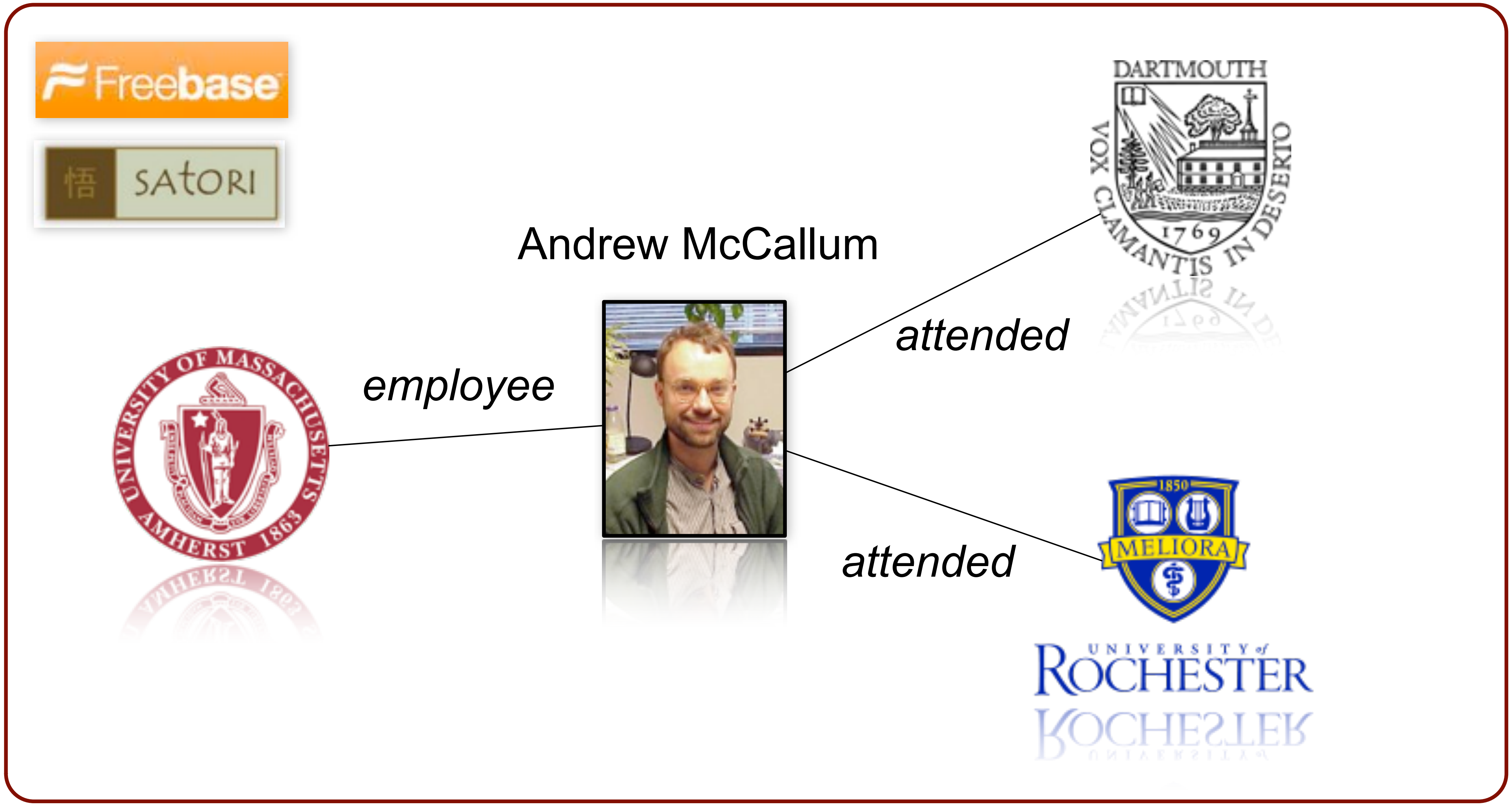
Motivation


andrew mccallum
Web
Images
Maps
Shopping
More
Search tools
About 4,380,000 results (0.20 seconds)
Cookies help us deliver our services. By using our services, you agree to our use of cookies.
OK
Learn more
Andrew McCallum Homepage
www.cs.umass.edu/~mccallum
Machine learning, text and information retrieval and extraction, reinforcement learning.
Andrew McCallum Publications - Andrew McCallum Bio - People - Teaching
Andrew McCallum - London Metropolitan University
www.londonmet.ac.uk/faculties/faculty-of...k.../andrew-mccallum/
Andrew taught English in London secondary schools for 15 years before coming to London Met in 2008. He is course tutor for the PGCE in Secondary English ...
Andrew McCallum - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Andrew_McCallum
Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. His primary specialties are in ...
Andrew Mccallum - United Kingdom profiles | LinkedIn
uk.linkedin.com/pub/dir/Andrew/Mccallum
View the profiles of professionals on LinkedIn named Andrew Mccallum located in the United Kingdom. There are 25 professionals named Andrew Mccallum in ...
Andrew McCallum
Software Developer
Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. Wikipedia
Education: Dartmouth College, University of Rochester
Awards: Best 10-year Paper Award of the ICML
People also search for
Tom M. Mitchell
Lee Giles
David M. Blei
Michael Collins
Robert Schapire
Feedback/More info

Motivation


bing
Andrew McCallum
382,000 RESULTS
Any time
Andrew McCallum Homepage - UMass CS | School of Computer ...
www.cs.umass.edu/~mccallum
Machine learning, text and information retrieval and extraction, reinforcement learning.
Andrew McCallum - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Andrew_McCallum
Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. His primary specialties are in ...
Images of Andrew McCallum
bing.com/images
Andrew_McCallum - Twitter
www.twitter.com/andrew_mccallum
We would like to show you a description here but the site won't allow us.
andrew mccallum profiles | LinkedIn
www.linkedin.com/pub/dir/andrew/mccallum
View the profiles of professionals named andrew mccallum on LinkedIn. There are 25 professionals named andrew mccallum, who use LinkedIn to exchange information ...
Andrew McCallum
Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. His primary specialties are in machine learning, natural language processing, informa... +
en.wikipedia.org
Education: Dartmouth College · University of Rochester
Awards: Best 10-year Paper Award of the ICML
People also search for
Tom M. Mitchell
Peter Norvig
Scott Fahlman
Report a problem

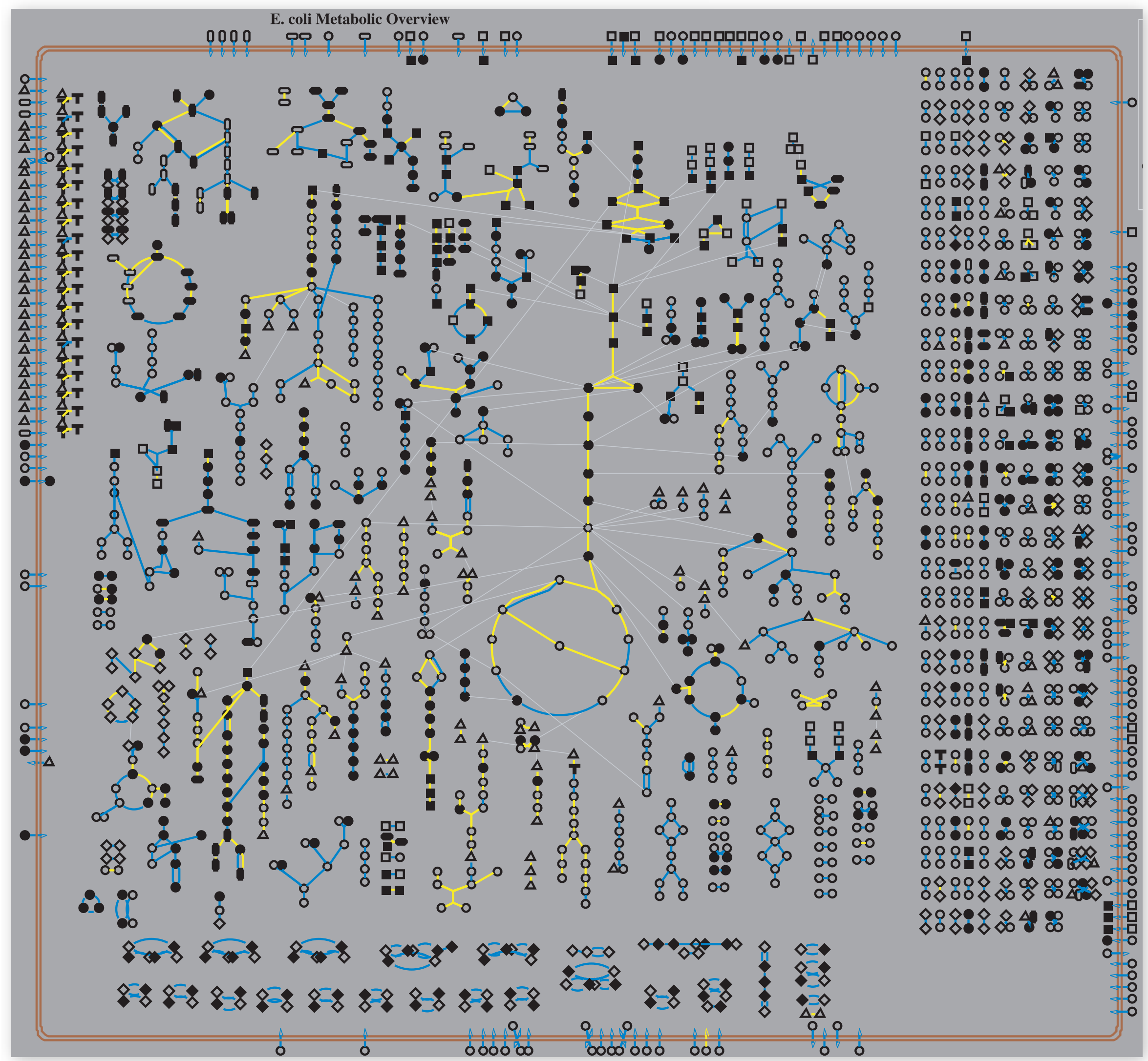
Knowledge Bases



Knowledge Bases

MetaCyc

EcoCyc



Knowledge Bases

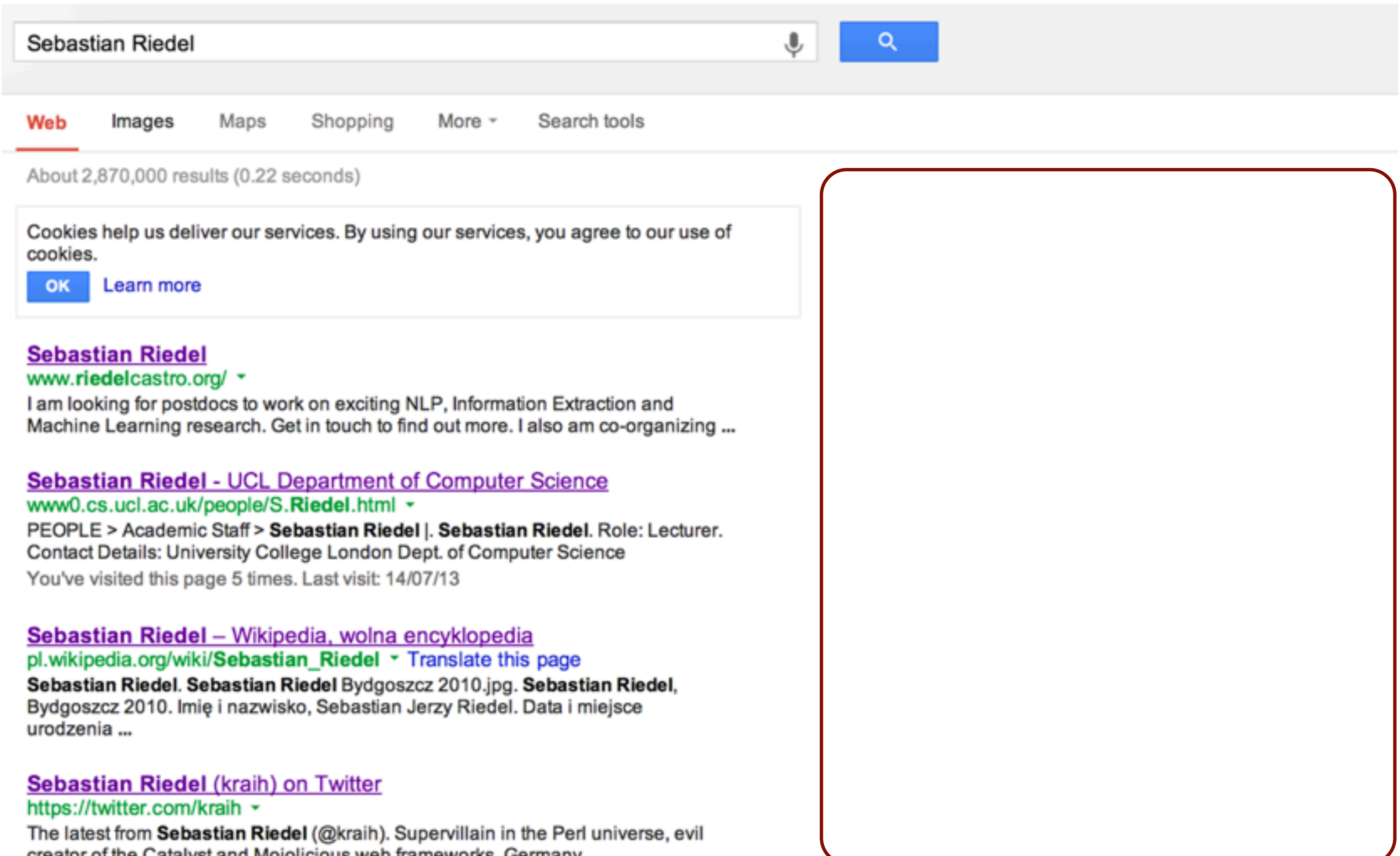
Knowledge Bases

- **Useful** for
  - Search
  - Data Mining
  - “Machines”/AI
  - Visualization

Knowledge Bases

- **Useful** for
  - Search
  - Data Mining
  - “Machines”/AI
  - Visualization
- Populated **Manually** (Freebase, Wikipedia,...)

Coverage (Facts)



Coverage (Schema)

WEB IMAGES VIDEOS MAPS NEWS MORE

bing Entity Extraction Datasets

35,000 RESULTS Any time

[Creating New **Datasets** - PNNL: IN-SPIRE™ - Home]
in-spire.pnnl.gov/IN-SPIRE_Help/Creating_New_**Data**_**Sets**.htm
Creating New **Datasets** Accessing the **Dataset** Editor . ... **Entity extraction** provides text analysis technology that automatically identifies and extracts key **entities**.

[Tagged **datasets** for named **entity** recognition tasks]
www.cs.technion.ac.il/~gabr/resources/data/ne_**datasets**.html
Resources for named **entity** recognition ... Tagged **datasets** for named **entity** recognition tasks. 1999 Information **Extraction** – **Entity** Recognition Evaluation

[Scalable Adhoc **Entity Extraction** from Text Collections - Microsoft ...]
research.microsoft.com/apps/pubs/default.aspx?id=79067
Supporting **entity extraction** from large document collec- ... strate the efficiency of our techniques on real **datasets**. PDF file. In: VLDB Conference. Details. Type:

[Linked Data Support: **Entity Extraction** | AlchemyAPI]
www.alchemyapi.com/api/**entity**/ldata.html
Home » Documentation » **Entity Extraction** » Linked Data Support: **Entity Extraction**. Linked Data Support: **Entity Extraction** <aapi:EntityType ...

[Efficient Approximate **Entity Extraction** with Edit Distance ...]
www.cse.unsw.edu.au/~weiw/files/SIGMOD09-ApproxDictMatching-Final.pdf · PDF file
named **entity** recognition **datasets** in various domains. The ... Scalable ad-hoc **entity extraction** from text collections. PVLDB, 1(1):945–957, 2008.

[Tutorial: Combining Google News, **Entity Extraction**, and Linked ...]
www.developerfusion.com/event/...**entity-extraction**-and-linked-data
Entity Extraction is a technology to extract **entities** such as names, ... to demo some queries on this **dataset** that are currently impossible with regular search engines.

RELATED SEARCHES

SharePoint Entity Extraction
C# Entity Extraction
SharePoint 2013 Custom Entity Extraction
Key Entity Extraction **IV Lyrics**
Entity **Name**
Semantic Entity
Name of Entity **Definition**
Stanford Ner

Overview

- Relation Extraction
- Universal Schemas
- 3 Relation Extraction Models
- Training
- Evaluation

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

Relation Extraction

[Cullota and Sorenson; 04, ...]

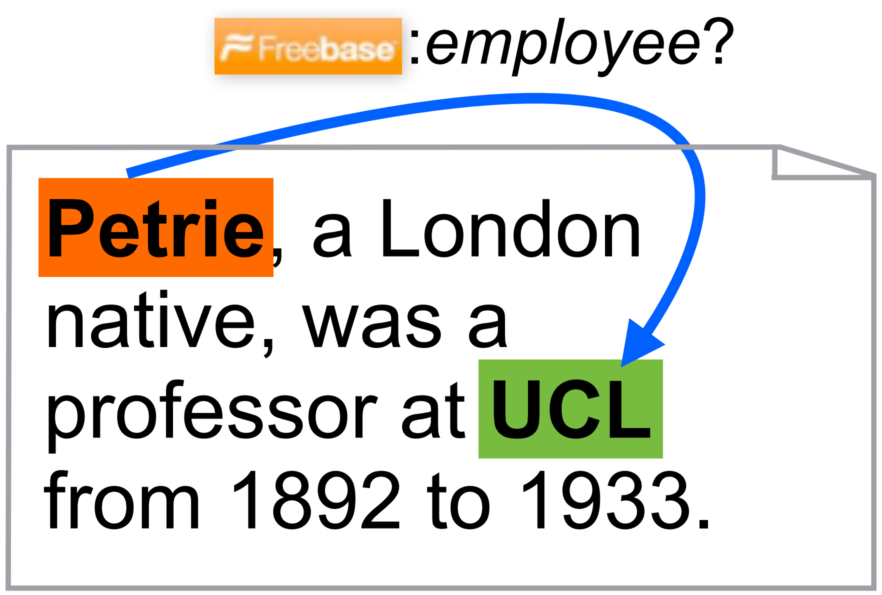
Predict **relations** between **entities** based on **mentions**

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**



Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Freebase:*employee*?

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Freebase:*employee*?

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Freebase:*employee*?

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \underline{\mathbf{f}^{x,y}_{\mathrm{emp}}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Freebase:*employee*?

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \underline{\mathbf{w}_{\mathrm{emp}}}) \propto \exp[< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Freebase:*employee*?

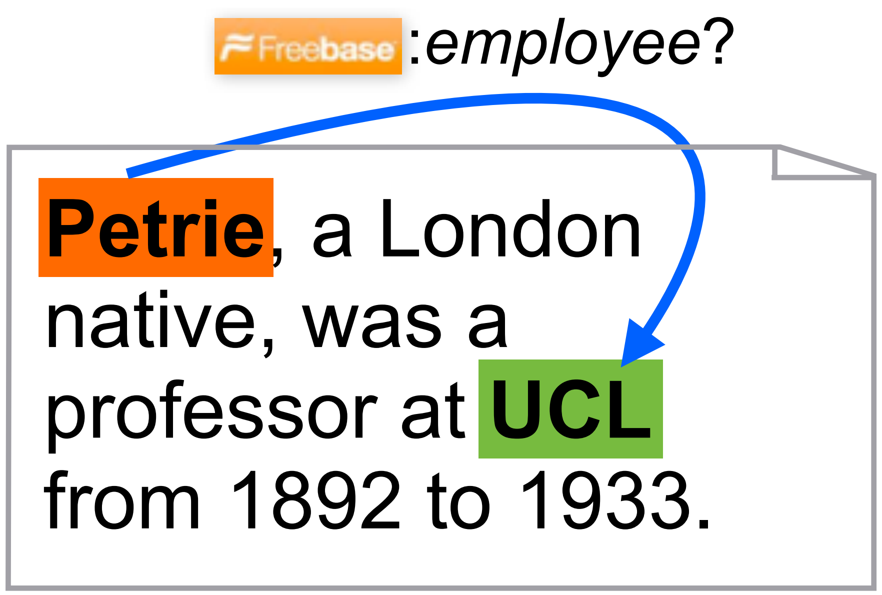
Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[\underline{< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >}]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**



$$
\begin{array}{ll}
 & \mathbf{f}^{x,y}_{\mathrm{emp}} \\
\begin{array}{l} \textit{arg1:PER} \\ \textit{arg1:ORG} \\ \textit{bw:native} \\ \textit{bw:professor} \\ \leftarrow\textit{was}\rightarrow\textit{professor}\rightarrow \\ \ldots \end{array} & \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}
\end{array}
$$

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Freebase :*employee*?

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$\begin{array}{ll} & \mathbf{f}^{x,y}_{\text{emp}} \\ \textit{arg1:PER} & \multirow{6}{*}{$\begin{bmatrix}1\\0\\1\\1\\1\\0\end{bmatrix}$} \\ \textit{arg1:ORG} & \\ \textit{bw:native} & \\ \textit{bw:professor} & \\ \leftarrow\textit{was}\rightarrow\textit{professor}\rightarrow & \\ \ldots & \end{array}$$

$$p(y^{x,y}_{\text{emp}} = 1 | \mathbf{f}^{x,y}_{\text{emp}}, \mathbf{w}_{\text{emp}}) \propto \exp[< \mathbf{f}^{x,y}_{\text{emp}}, \mathbf{w}_{\text{emp}} >]$$

Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >]$$

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Humans label text

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Humans label text

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Train

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Humans label text

Petrie, a London native, was a professor at **UCL** from 1892 to 1933. ✓

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects. ✗

Train

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism. ✗

Ferguson is a professor at **Harvard**. ✓

Andrew is a professor at **Cambridge** ✓

Andrew teaches history at **Cambridge** ✗

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**. ✗

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Predict **relations** between **entities** based on **mentions**

Humans label text

Freebase

employee(**X**,**Y**)

Train

Petrie, a London native, was a professor at **UCL** from 1892 to 1933. ✓

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects. ✗

Petrie, UCL: 1

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism. ✗

Ferguson is a professor at **Harvard**. ✓

Ferguson, Harvard: 1

Andrew is a professor at **Cambridge** ✓

Andrew teaches history at **Cambridge** ✗

Andrew, Cambridge: 1

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**. ✗

Trevelyan, Cambridge: 0

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Rows ordered by **entity pairs**

Humans label text

Freebase

				employee(**X**,**Y**)	
	Petrie, a London native, was a professor at **UCL** from 1892 to 1933.	The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.	Train	1	Petrie, UCL
Ferguson is a historian at **Harvard** focusing on British and American Imperialism.	**Ferguson** is a professor at **Harvard**.		Test	1	Ferguson, Harvard
	Andrew is a professor at **Cambridge**		**Andrew** teaches history at **Cambridge**	1	Andrew, Cambridge
His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.				0	Trevelyan, Cambridge

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Columns ordered by **patterns**

Humans label text

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Train

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Freebase

employee(**X**,**Y**)	
1	Petrie, UCL
1	Ferguson, Harvard
1	Andrew, Cambridge
0	Trevelyan, Cambridge

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Fails for **unseen** patterns

Humans label text

Freebase

employee(**X**,**Y**)

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

1 — Petrie, UCL

Train

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

1 — Ferguson, Harvard

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

1 — Andrew, Cambridge

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

0 — Trevelyan, Cambridge

Supervised Relation Extraction

[Cullota and Sorenson; 04, ...]

Annotation costly...

Humans label text

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Train

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Freebase

employee(**X**,**Y**)	
1	Petrie, UCL
1	Ferguson, Harvard
1	Andrew, Cambridge
0	Trevelyan, Cambridge

Distant Supervision

[Craven & Kumlien,99; Mintz et al.,09, Riedel et al., 10]

Let's leverage **pre-existing databases**

Freebase

employee(**X**,**Y**)

Database labels text

Train

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

1

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

1

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

1

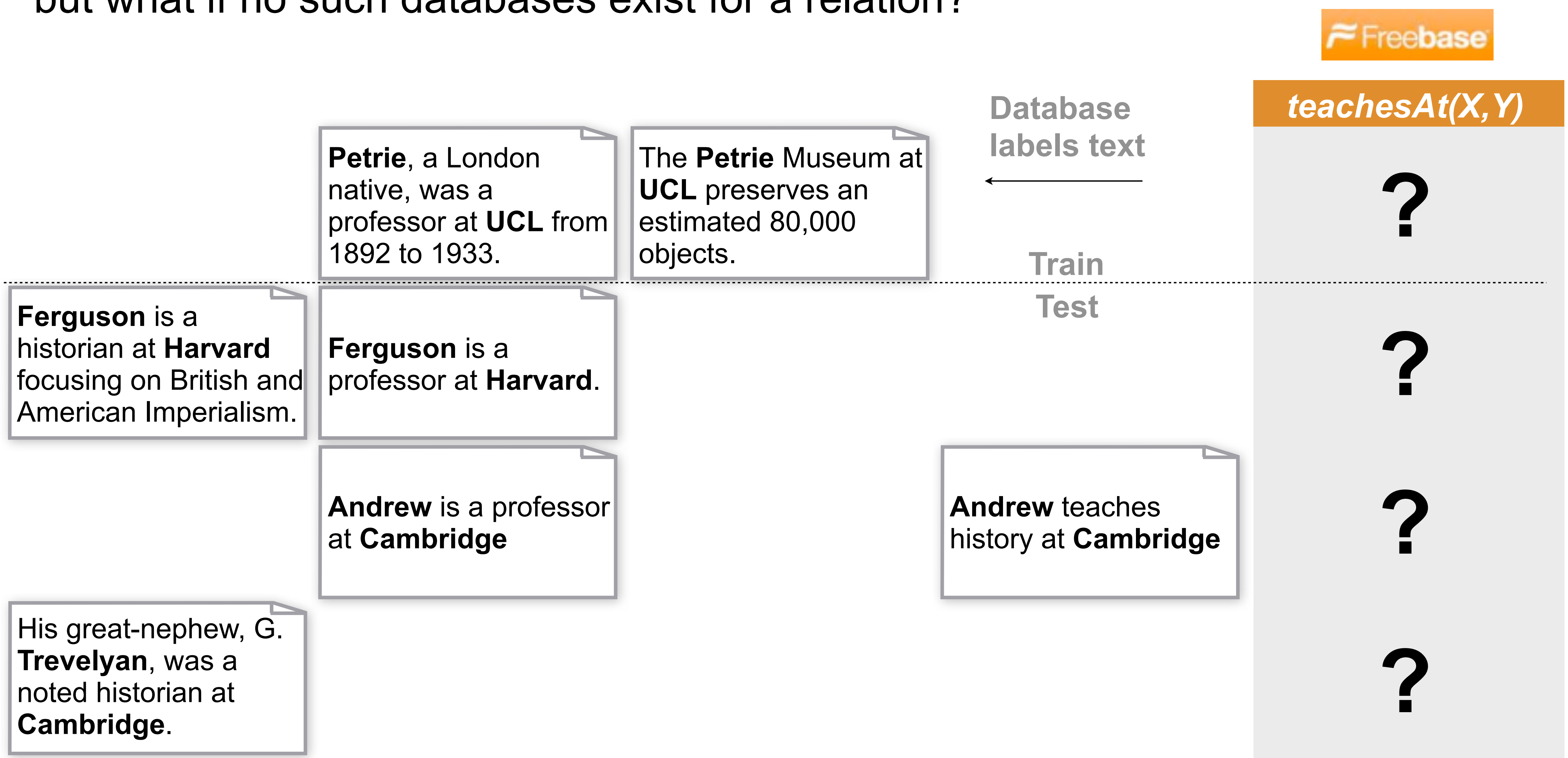
His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

0

Distant Supervision

[Riedel et al.,10; Hoffmann et al., 11, Surdeanu et al., 12]

but what if no such databases exist for a relation?



Distant Supervision

[Riedel et al.,10; Hoffmann et al., 11, Surdeanu et al., 12]

but what if no such databases exist for a relation?

Freebase

historianAt(X,Y)

Database labels text

Train

Test

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

?

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

?

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

?

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

?

Distant Supervision

[Riedel et al.,10; Hoffmann et al., 11, Surdeanu et al., 12]

but what if no such databases exist for a relation?

Freebase

Database labels text

Train

Test

criticizes(X,Y)
?
?
?
?

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Universal Schemas

Matrix Representation

Convert problem into a **matrix formulation**

Database labels text

employee(**X**,**Y**)

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

1

Ferguson *is a historian at* **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

1

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

1

His great-nephew, G. **Trevelyan**, *was* a noted *historian at* **Cambridge**.

0

Matrix Representation

Columns correspond to **patterns** between mentions

Freebase

Database labels text

X-*is-historian-at*-**Y**				*employee*(**X**,**Y**)
	Petrie, a London native, was a professor at **UCL** from 1892 to 1933.	The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.		1
1	**Ferguson** is a professor at **Harvard**.			1
	Andrew is a professor at **Cambridge**		**Andrew** teaches history at **Cambridge**	1
1				0

Matrix Representation

Columns correspond to **patterns** between mentions

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	*employee*(**X**,**Y**)
	1	1
1	1	1
	1	1
1		0

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Database labels text

Andrew teaches history at **Cambridge**

Matrix Representation

Columns correspond to **patterns** between mentions

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
1	**1**			**1**
	1		**1**	**1**
1				**0**

Extending the Schema

So what about relations with no pre-existing databases?

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	***historianAt(X,Y)***
	1	1		?
1	1			?
	1		1	?
1				?

Open Information Extraction

[Etzioni et al.,08]

Patterns become relations...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	***historianAt(X,Y)***
	1	1		?
1	1			?
	1		1	?
1				?

Open Information Extraction

[Etzioni et al.,08]

and often correspond to your target relations

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*historianAt(X,Y)*
	1	1		?
1	1			?
	1		1	?
1				?

Open Information Extraction

[Etzioni et al.,08]

...but no **reasoning** / generalization

X-*is-historian-at*-Y	X-*is-professor-at*-Y	X-*museum-at*-Y	X-*teaches-history-at*-Y	***historianAt(X,Y)***
	1	1		?
1	1			?
0	1		1	?
1				?

Pattern Clustering

[Lin & Pantel,01; Yates & Etzioni, 09, ...]

Find patterns with "*similar meaning*"

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*historianAt(X,Y)*
	1	1		?
1	1			?
	1		1	?
1				?

Pattern Clustering

[Lin & Pantel,01; Yates & Etzioni, 09, ...]

Clustering these into a **latent** relation...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	***historianAt(X,Y)***
	1	1		
1	1			1
	1		1	1
1				1

Pattern Clustering

[Lin & Pantel,01; Yates & Etzioni, 09, ...]

...assumes **symmetry**...

X-*is-historian-at*-**Y** ➔	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*historianAt(X,Y)*
	1	1		
1	1			1
	1		1	1
1				1

Pattern Clustering

[Lin & Pantel,01; Yates & Etzioni, 09, ...]

...assumes **symmetry**...

X-*is-historian-at*-**Y** ➔	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	***professorAt(X,Y)***
	1	**1**		**1**
1	**1**			**1**
	1		**1**	**1**
1				**1**

Pattern Clustering

[Lin & Pantel,01; Yates & Etzioni, 09, ...]

...ignores **Context**.

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*historianAt(X,Y)*
	1	1		
1	1			
	1		1	1
1				

Pattern Clustering

[Lin & Pantel,01; Yates & Etzioni, 09, ...]

...ignores **Context**.

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	***historianAt(X,Y)***
	1	1		
1	1			
	0		**1**	**0**
1				

Relation Extraction

Recall that relation extraction fills in cells

Freebase

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			?
	1		1	?
1				?

Universal Schema

Extend schema to the **universe** of *all input relations*

Freebase

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
1	**1**			**?**
	1		**1**	**?**
1				**?**

Reasoning with Universal Schema

Try to **fill in all cells**

Freebase

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
?	1	1	?	1
1	1	?	?	?
?	1	?	1	?
1	?	?	?	?

Mutually Supportive

Reasoning about patterns helps structured relations

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
	1		1	
1				0

Mutually Supportive

Reasoning about patterns helps structured relations

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1			1
1	1			
1				0

Mutually Supportive

Reasoning about patterns helps structured relations

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1			**1**
1	**1**			
1	**0.9**			**0.8**

Models

Model N: Classifier

[Mintz et al 2009,...]

Standard supervised relation extractor ...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		$y^{x,y}_{\mathrm{emp}}$

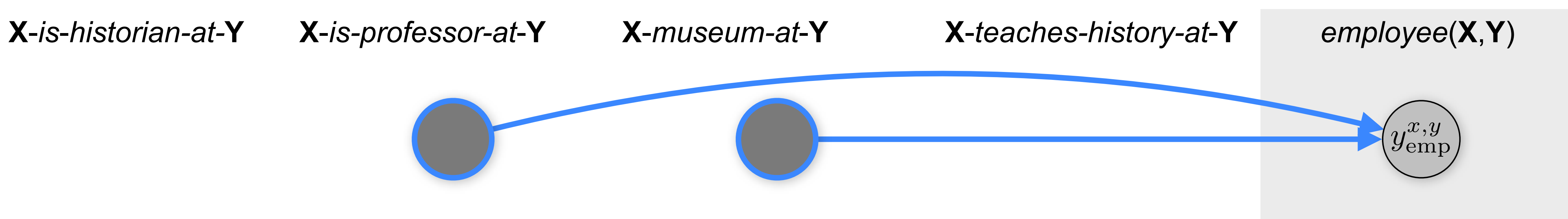
training data

$$p(y^{x,y}_{\mathrm{emp}} = 1 | \qquad\qquad\qquad)$$

Model N: Classifier

[Mintz et al 2009,...]

Standard supervised relation extractor ...



observed

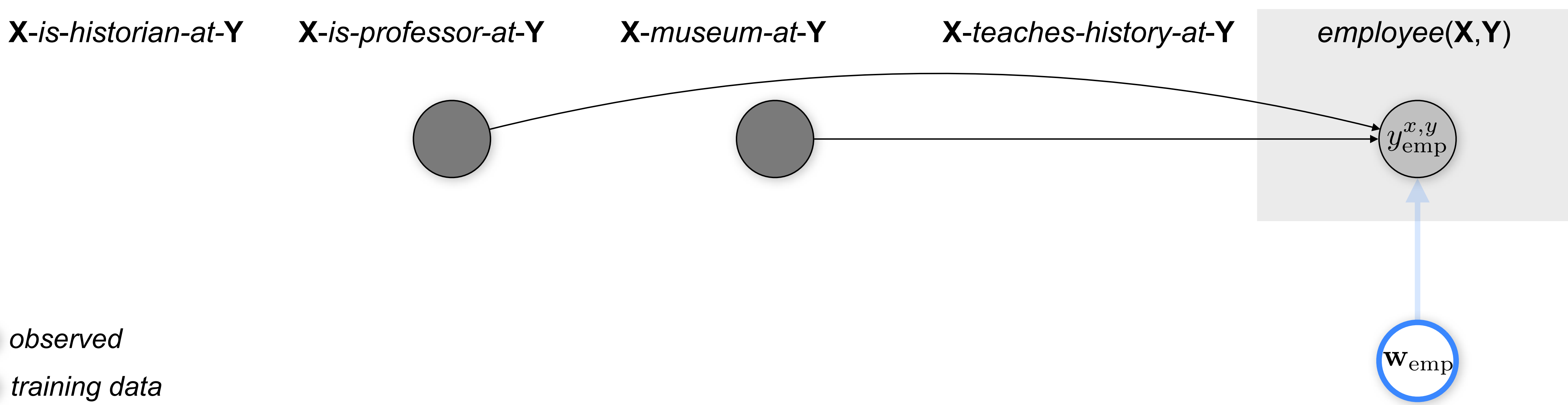
training data

$$p(y^{x,y}_{\text{emp}} = 1 | \underline{\mathbf{f}^{x,y}_{\text{emp}}} \qquad)$$

Model N: Classifier

[Mintz et al 2009,...]

Standard supervised relation extractor ...



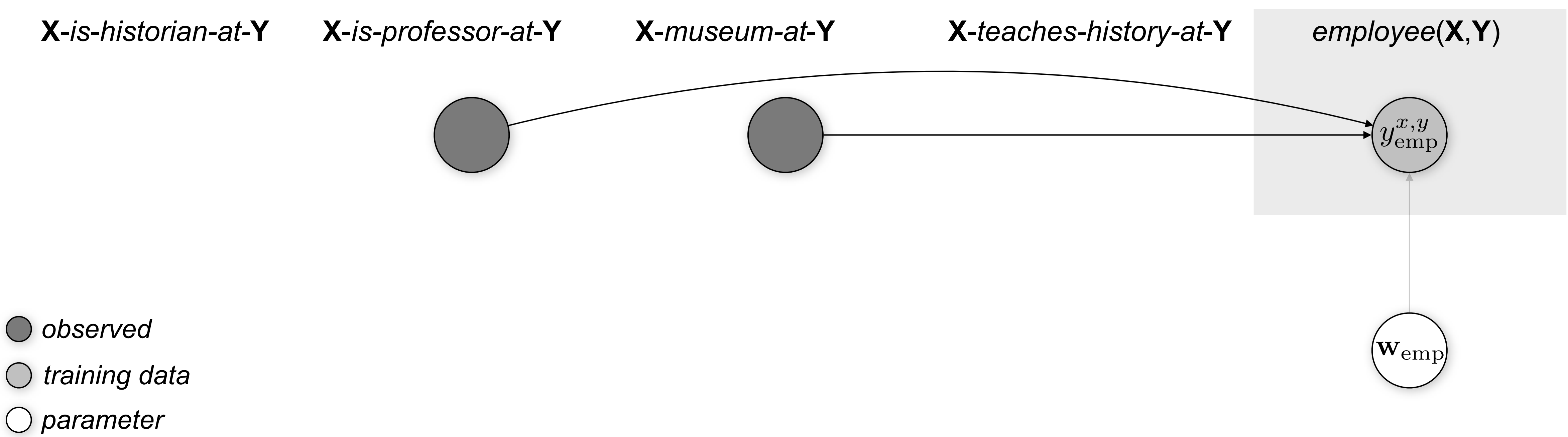
- *observed*
- *training data*
- *parameter*

$$p(y^{x,y}_{\text{emp}} = 1 | \mathbf{f}^{x,y}_{\text{emp}}, \underline{\mathbf{w}_{\text{emp}}})$$

Model N: Classifier

[Mintz et al 2009,...]

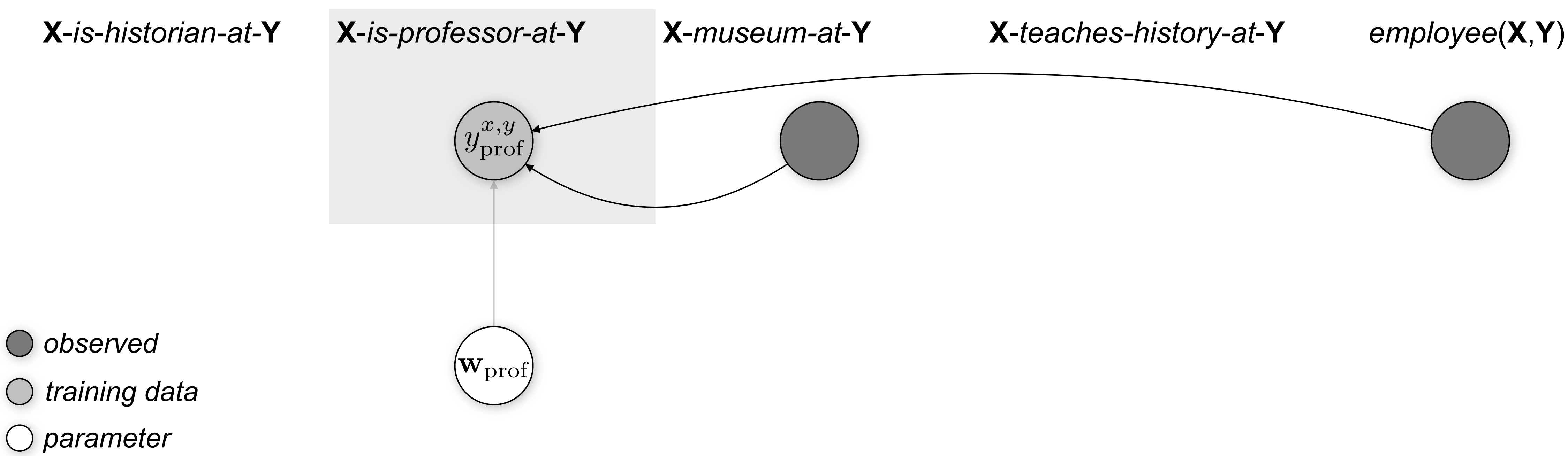
Standard supervised relation extractor ...



$$p(y^{x,y}_{\mathrm{emp}} = 1 | \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}}) \propto \exp[\underline{< \mathbf{f}^{x,y}_{\mathrm{emp}}, \mathbf{w}_{\mathrm{emp}} >}]$$

Model N: Classifier

... for each pattern

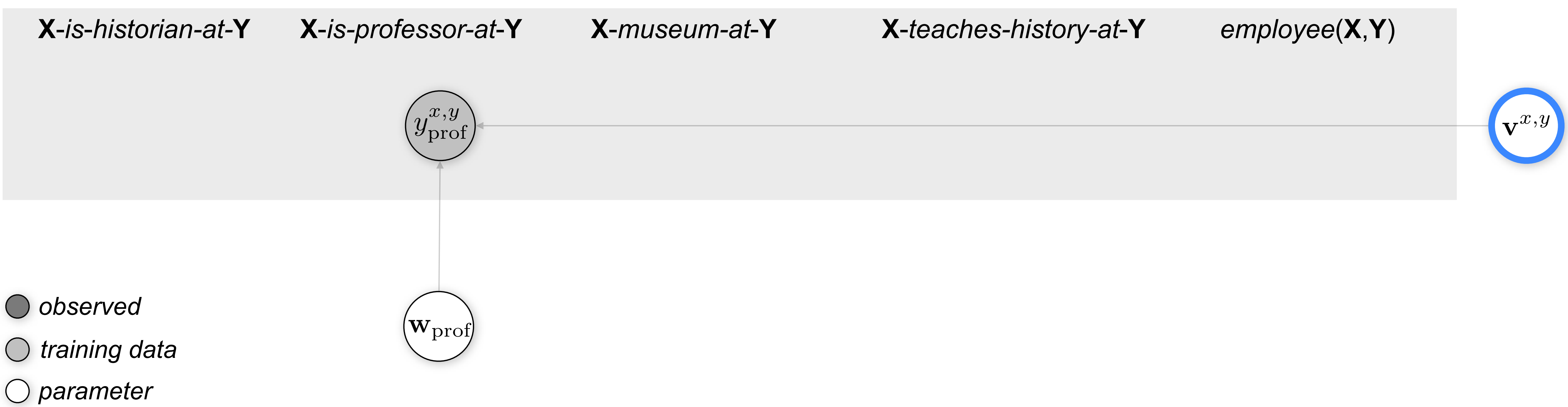


$$p(y^{x,y}_{\text{prof}} = 1 | \mathbf{f}^{x,y}_{\text{prof}}, \mathbf{w}_{\text{prof}}) \propto \exp[< \mathbf{f}^{x,y}_{\text{prof}}, \mathbf{w}_{\text{prof}} >]$$

Model F: Latent Feature (Factorization)

[Collins et al, 2001]

Per tuple **latent feature** vector

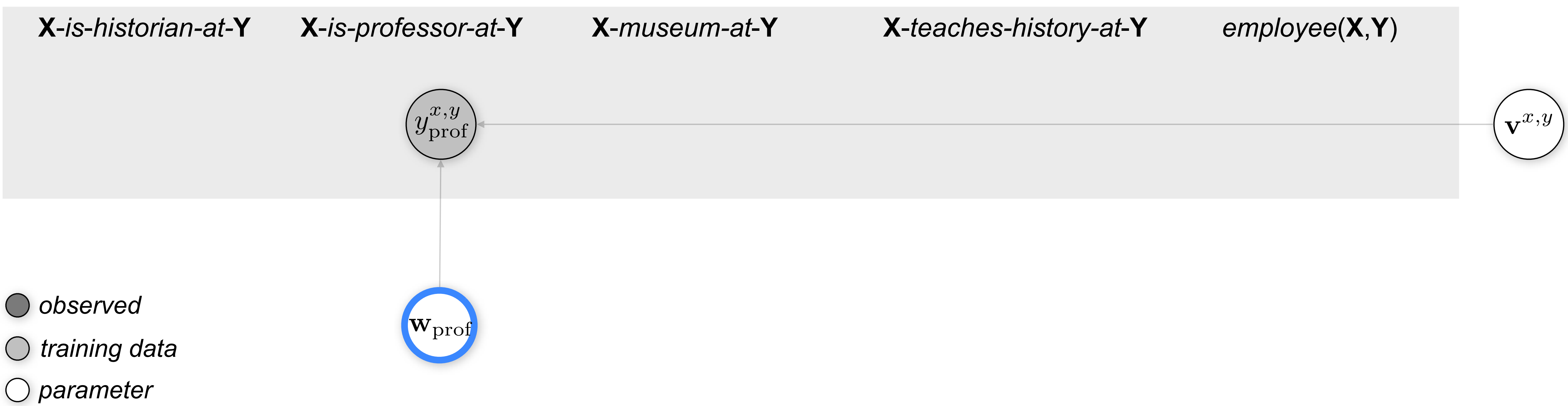


$$p(y^{x,y}_{\text{prof}} = 1 | \underline{\mathbf{v}^{x,y}}, \mathbf{w}_{\text{prof}}) \propto \exp[< \underline{\mathbf{v}^{x,y}}, \mathbf{w}_{\text{prof}} >]$$

Model F: Latent Feature (Factorization)

[Collins et al, 2001]

Per tuple **latent feature** vector



$$p(y^{x,y}_{\text{prof}} = 1 | \mathbf{v}^{x,y}, \underline{\mathbf{w}_{\text{prof}}}) \propto \exp[< \mathbf{v}^{x,y}, \underline{\mathbf{w}_{\text{prof}}} >]$$

Model F: Latent Feature (Factorization)

[Collins et al, 2001]

Per tuple **latent feature** vector

$$p(y^{x,y}_{\text{prof}} = 1 | \mathbf{v}^{x,y}, \mathbf{w}_{\text{prof}}) \propto \exp[< \mathbf{v}^{x,y}, \mathbf{w}_{\text{prof}} >]$$

Matrix Factorization: PCA

[Collins et al, 2001]

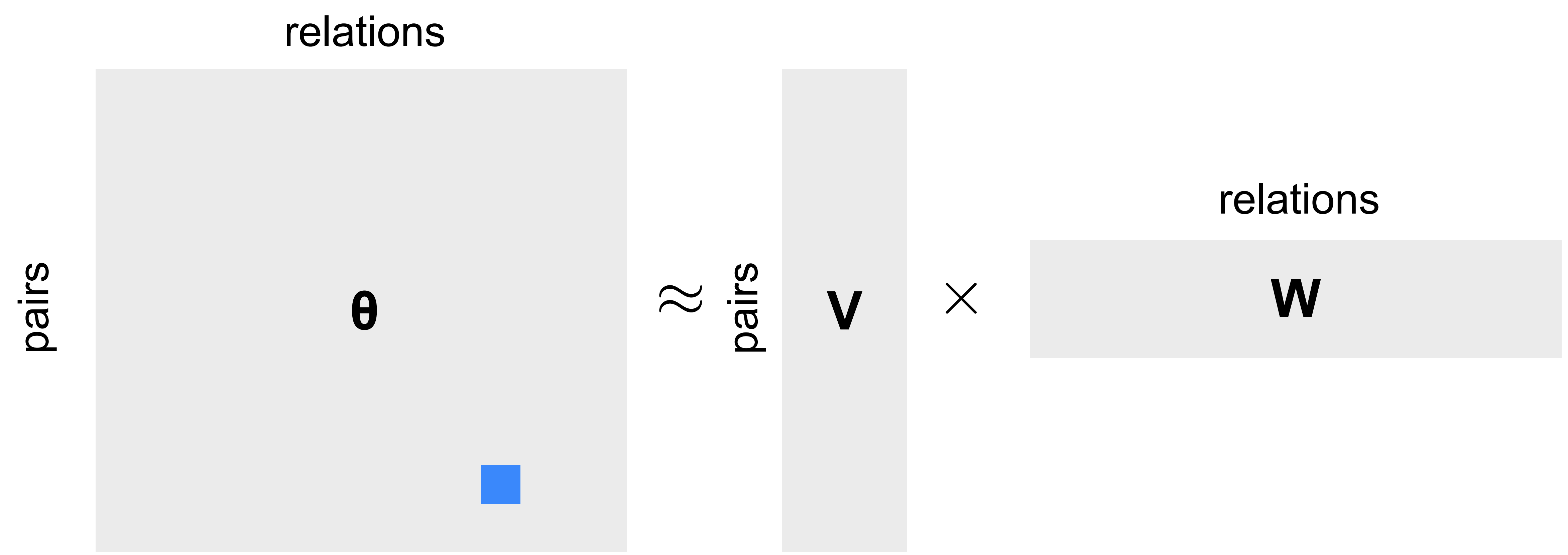
Natural parameters approximated by a low-rank matrix product

$$\theta^{x,y}_{\text{emp}} = < \mathbf{v}^{x,y}, \mathbf{w}_{\text{emp}} > \qquad p(y^{x,y}_{\text{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\text{emp}}$$

Matrix Factorization: PCA

[Collins et al, 2001]

Natural parameters approximated by a low-rank matrix product

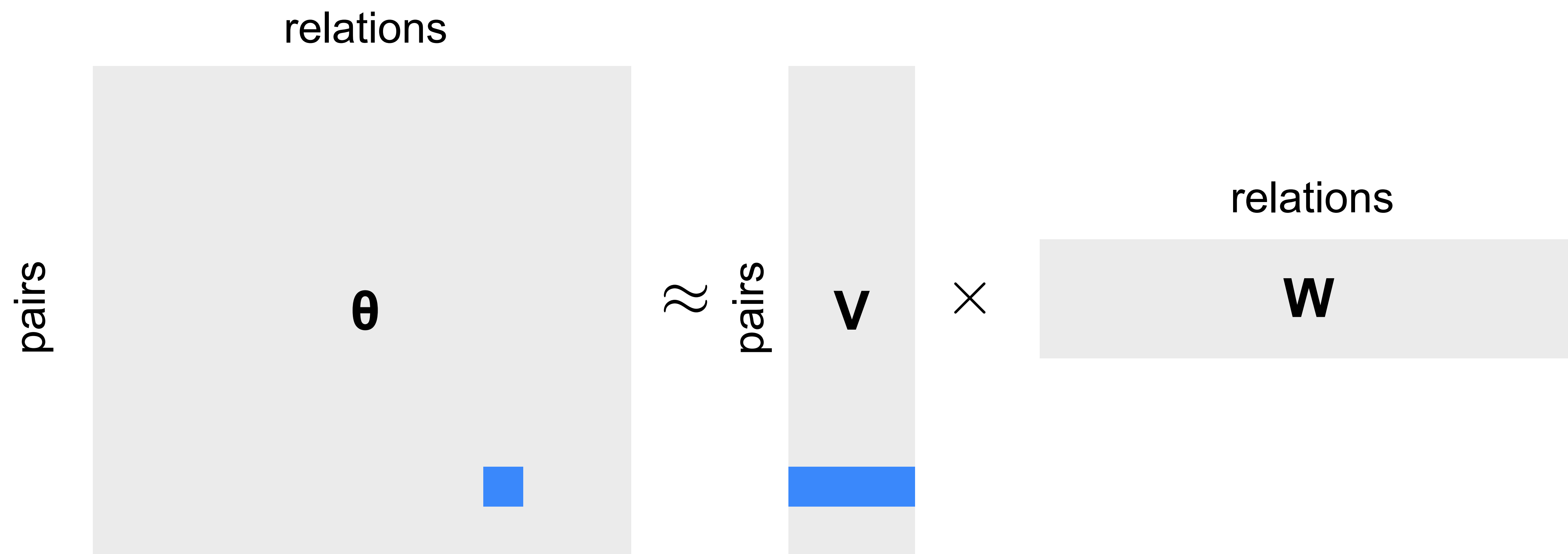


$$\theta^{x,y}_{\mathrm{emp}} = < \mathbf{v}^{x,y}, \mathbf{w}_{\mathrm{emp}} > \qquad p(y^{x,y}_{\mathrm{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\mathrm{emp}}$$

Matrix Factorization: PCA

[Collins et al, 2001]

Natural parameters approximated by a low-rank matrix product



$$\theta^{x,y}_{\mathrm{emp}} = < \underline{\mathbf{v}^{x,y}}, \mathbf{w}_{\mathrm{emp}} > \qquad p(y^{x,y}_{\mathrm{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\mathrm{emp}}$$

Matrix Factorization: PCA

[Collins et al, 2001]

Natural parameters approximated by a low-rank matrix product

$$\theta^{x,y}_{\mathrm{emp}} = < \mathbf{v}^{x,y}, \underline{\mathbf{w}_{\mathrm{emp}}} > \qquad p(y^{x,y}_{\mathrm{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\mathrm{emp}}$$

Matrix Factorization: PCA

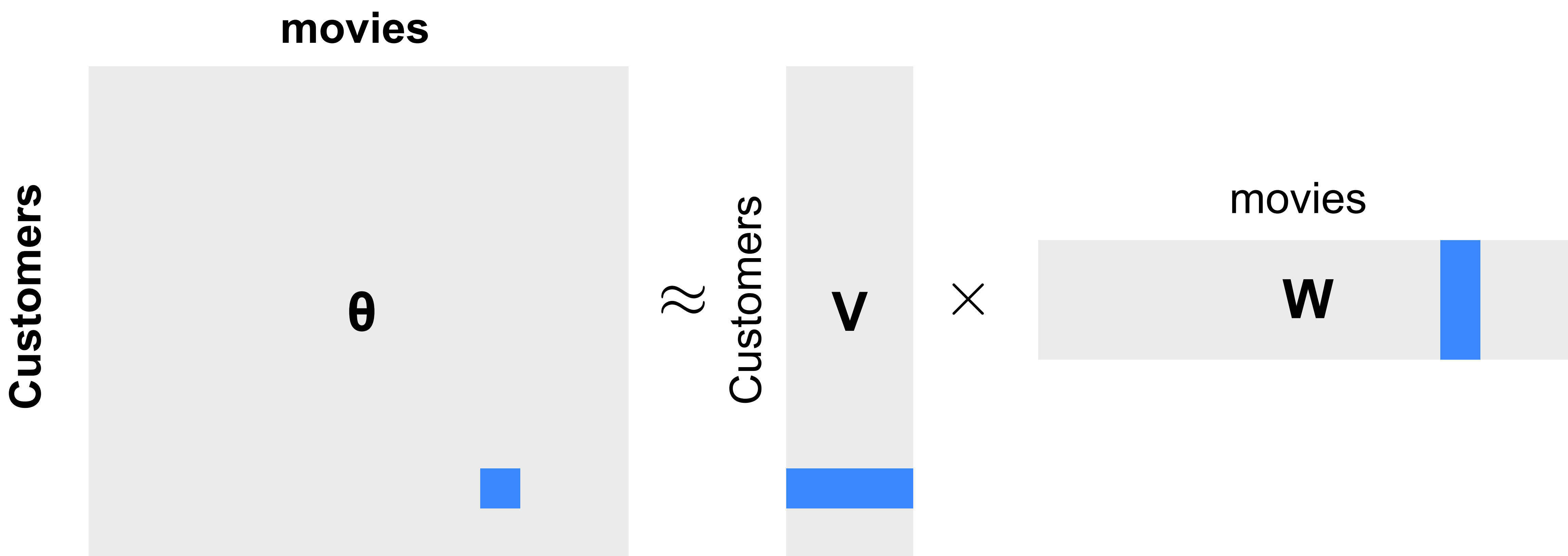
[Collins et al, 2001]

Natural parameters approximated by a low-rank matrix product

$$\theta^{x,y}_{\mathrm{emp}} = < \mathbf{v}^{x,y}, \mathbf{w}_{\mathrm{emp}} > \qquad p(y^{x,y}_{\mathrm{emp}} = 1 | \ldots) \propto \underline{\exp \theta^{x,y}_{\mathrm{emp}}}$$

Benefit

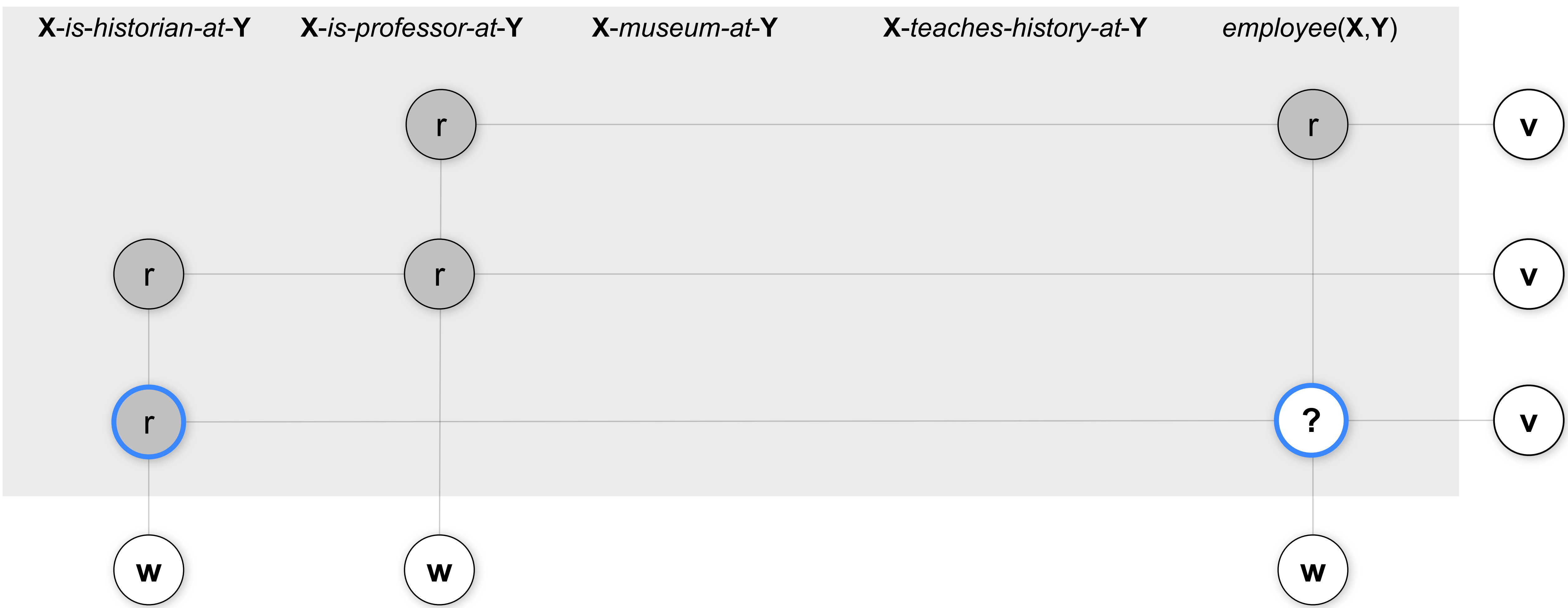
We can leverage large body of scalable methods in **collaborative filtering**



$$\theta^{x,y}_{\text{emp}} =< \mathbf{v}^{x,y}, \mathbf{w}_{\text{emp}} > \qquad p(y^{x,y}_{\text{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\text{emp}}$$

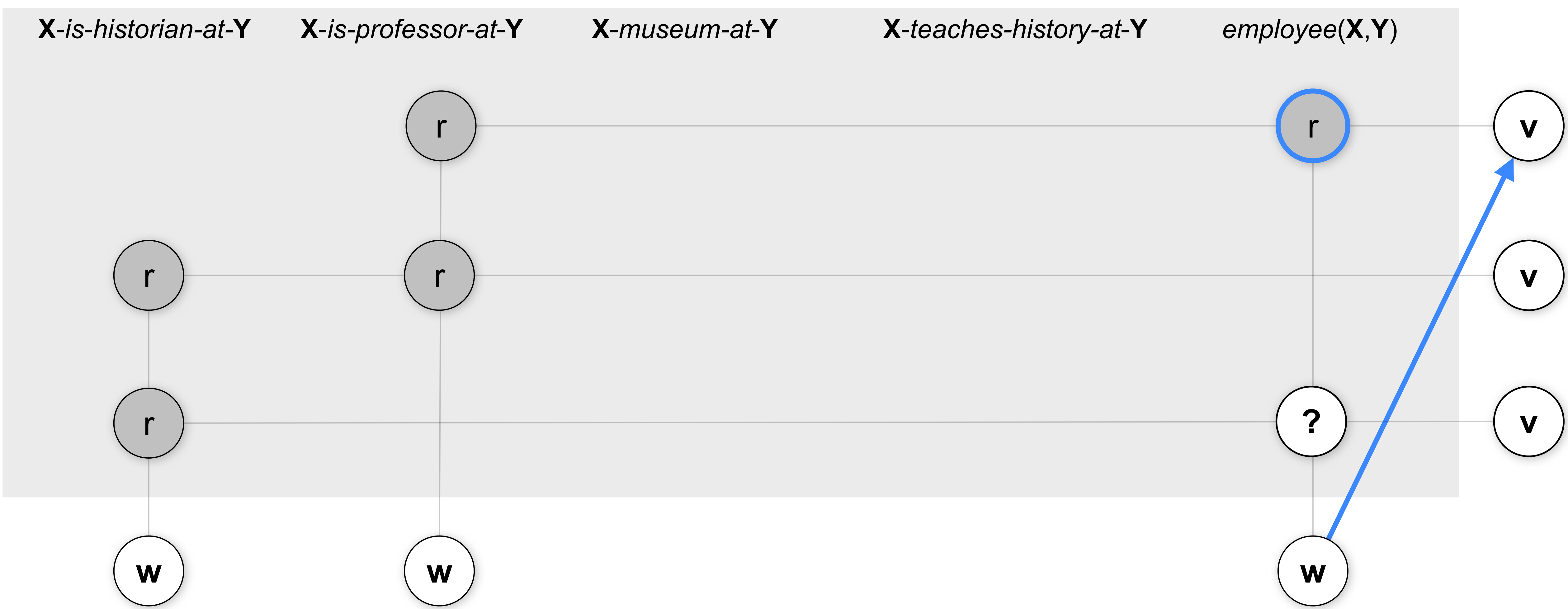
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



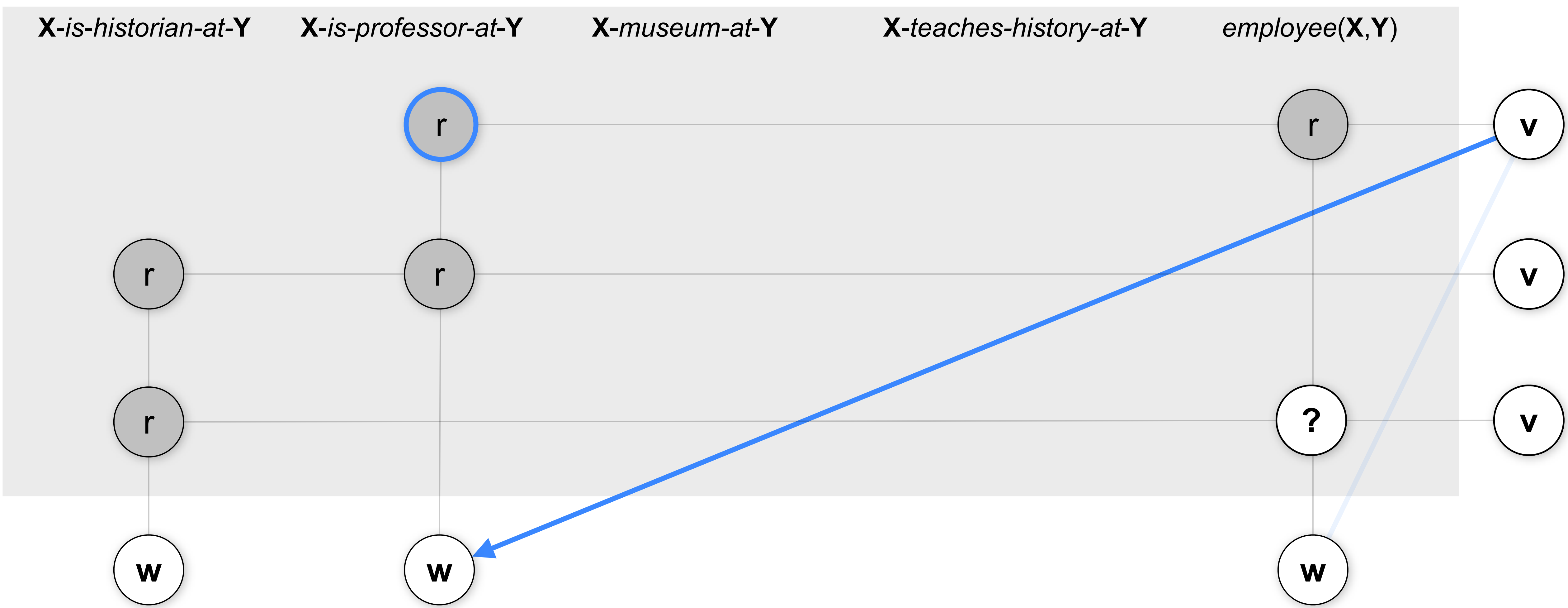
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



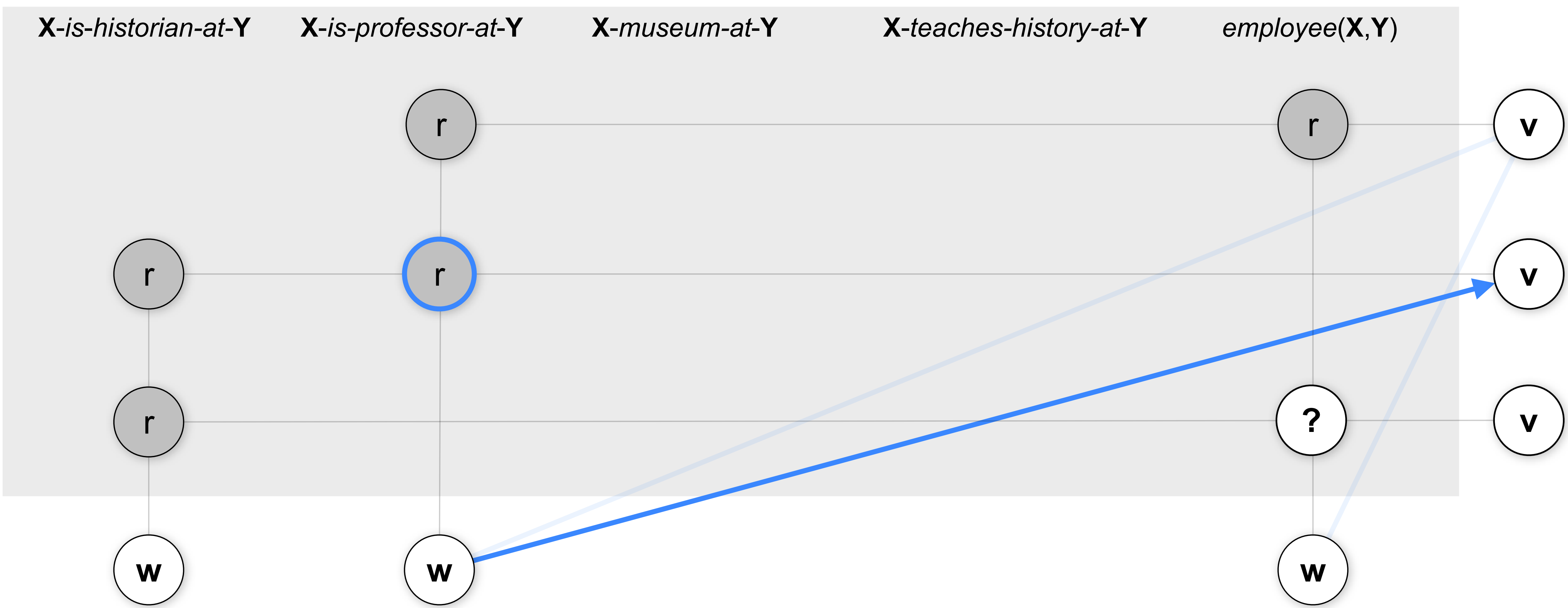
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



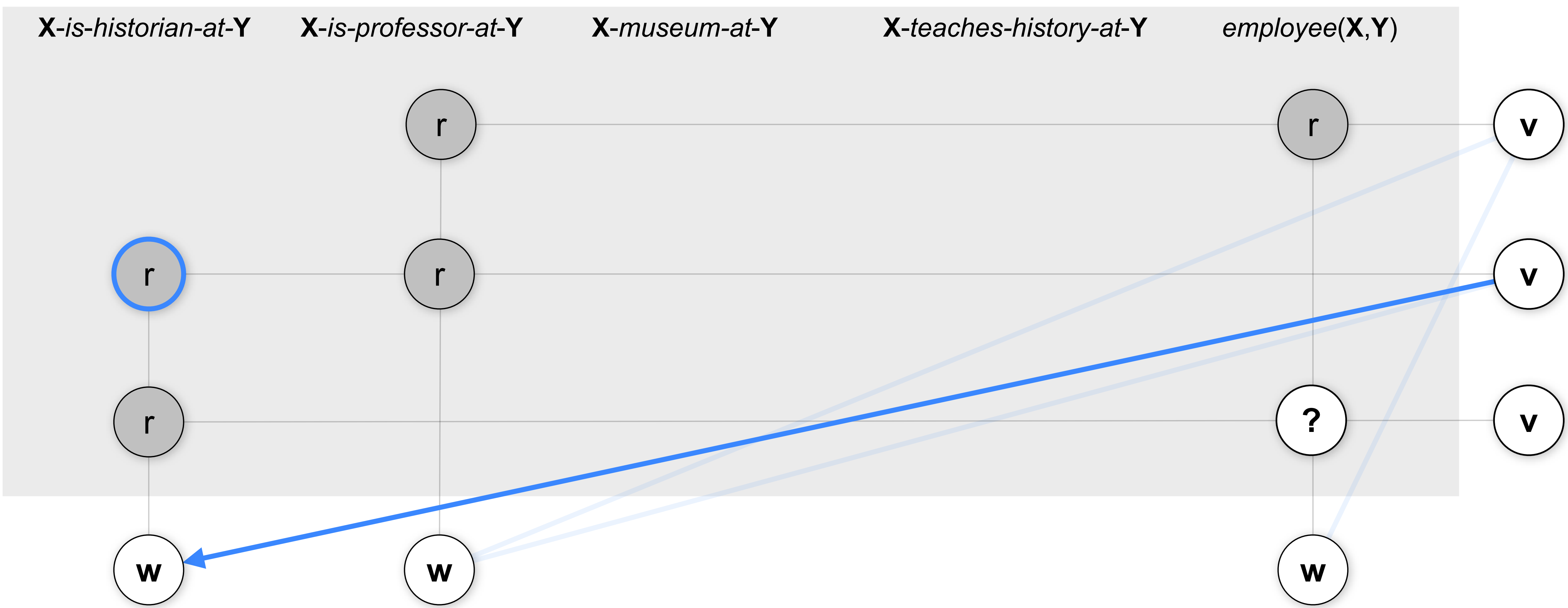
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



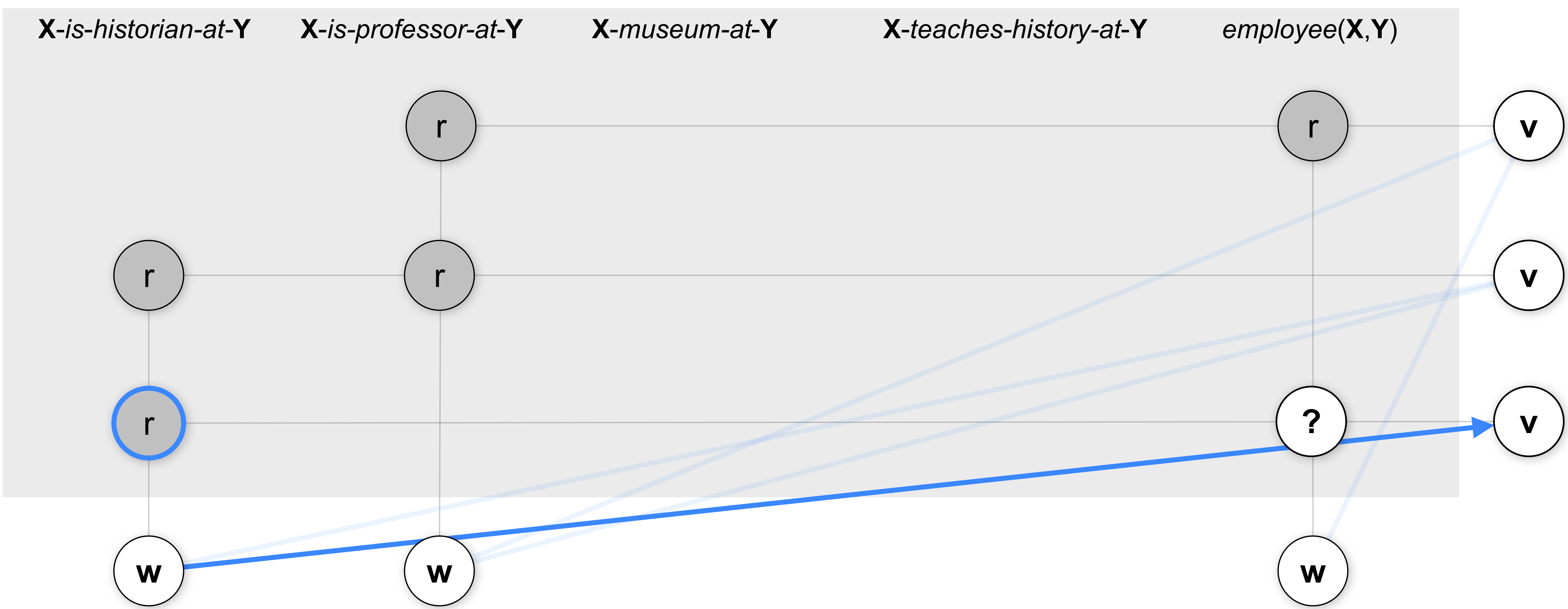
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



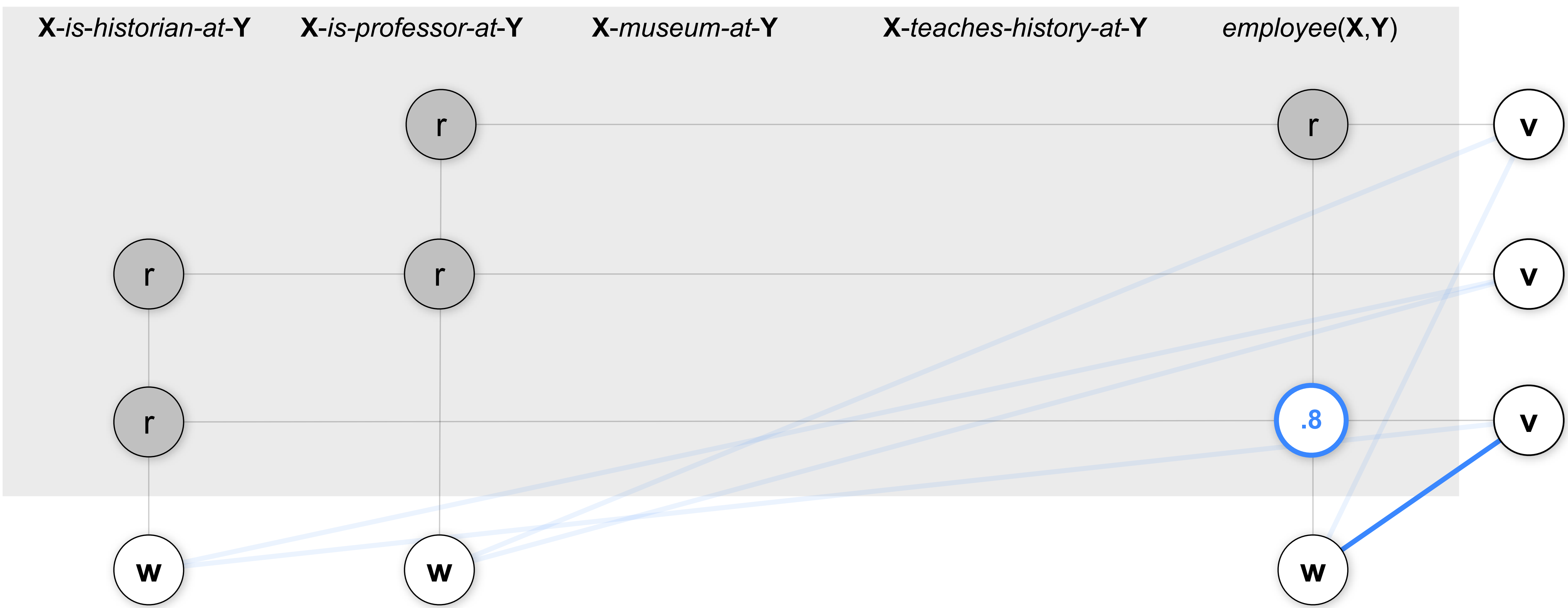
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



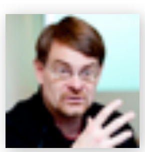
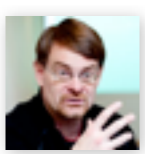
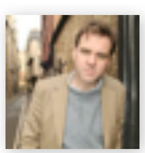
Model E: Selectional Preferences

Relations have **entity type restriction**

	X-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**
()			
()			
()			

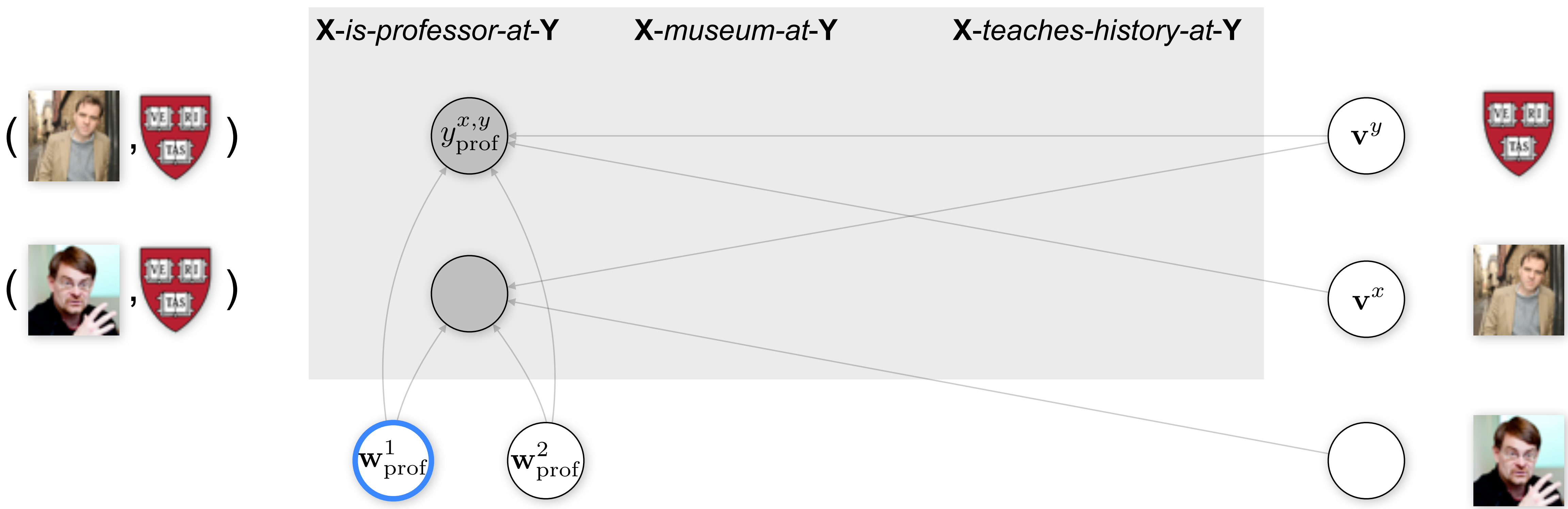
Model E: Selectional Preferences

Relations have **entity type restriction**

	X-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**
( , )			
( , )			
( , )	0	0	0

Model E: Selectional Preferences

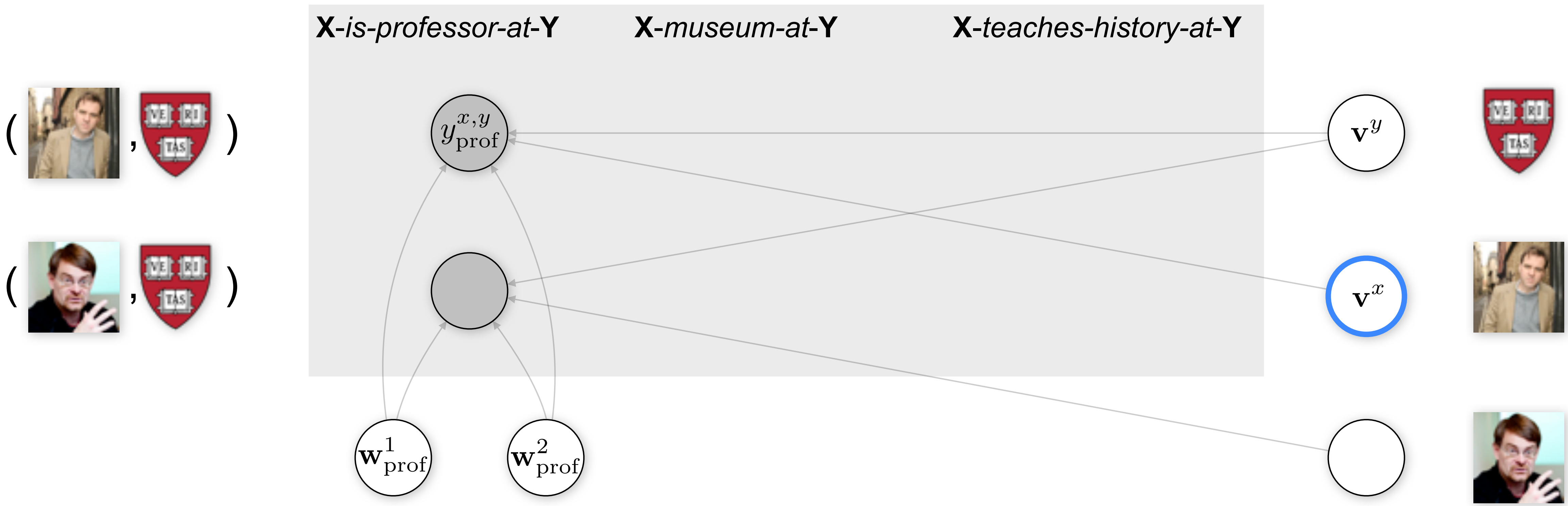
Argument Slot 1 weight vector ...



$$p(y^{x,y}_{\text{prof}} = 1 | \ldots) \propto \exp[< \mathbf{v}^x, \mathbf{w}^1_{\text{prof}} > + < \mathbf{v}^y, \mathbf{w}^2_{\text{prof}} >]$$

Model E: Selectional Preferences

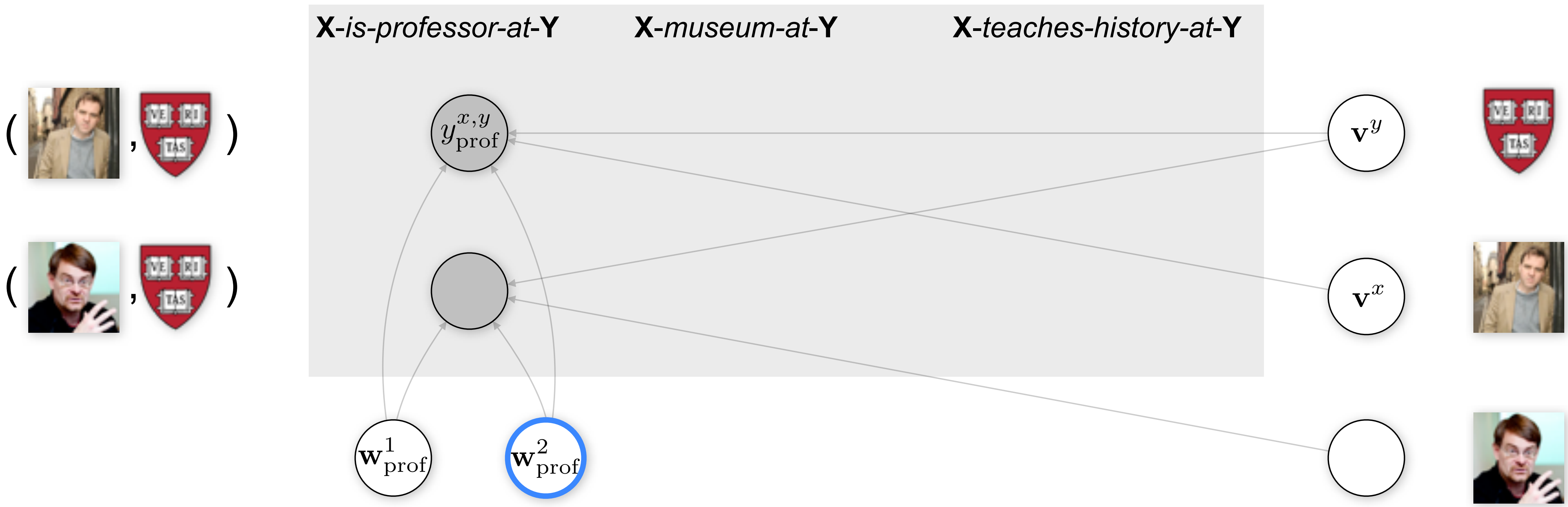
... dot-product with feature vector of **entity 1**



$$p(y^{x,y}_{\text{prof}} = 1 | \ldots) \propto \exp[< \underline{\mathbf{v}}^x, \mathbf{w}^1_{\text{prof}} > + < \mathbf{v}^y, \mathbf{w}^2_{\text{prof}} >]$$

Model E: Selectional Preferences

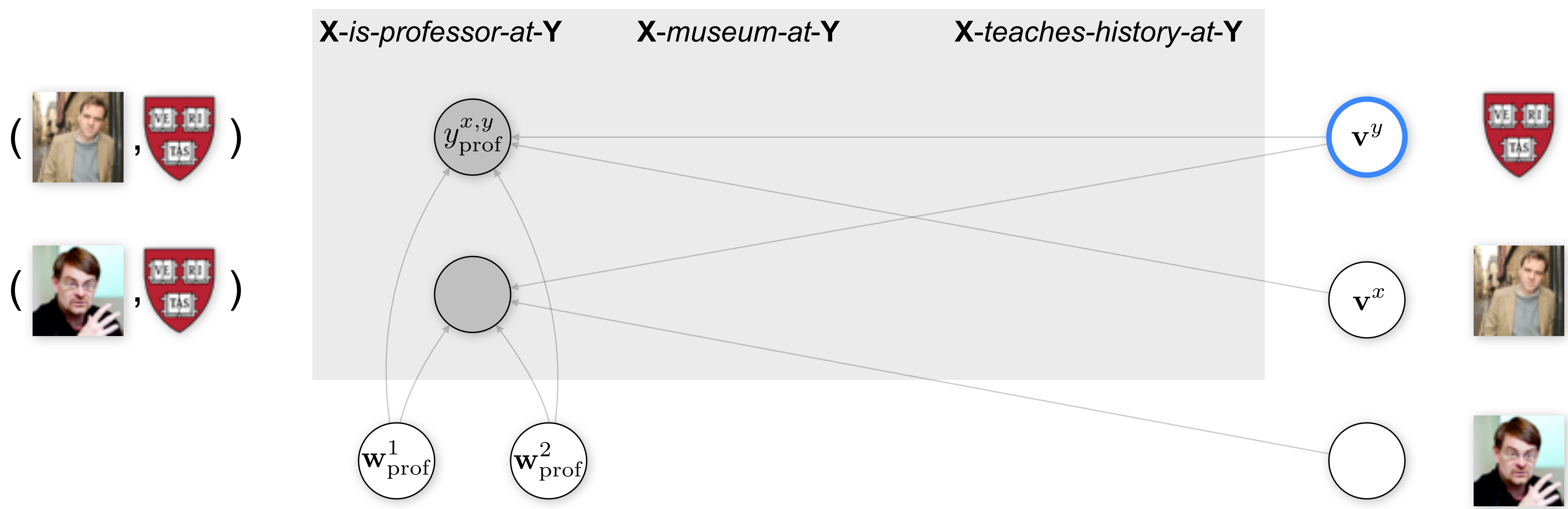
Argument Slot 2 weight vector ...



$$p(y^{x,y}_{\text{prof}} = 1 | \ldots) \propto \exp[< \mathbf{v}^x, \mathbf{w}^1_{\text{prof}} > + < \mathbf{v}^y, \mathbf{w}^2_{\text{prof}} >]$$

Model E: Selectional Preferences

... dot-product with feature vector of **entity 2**



$$p(y^{x,y}_{\mathrm{prof}} = 1 | \ldots) \propto \exp[< \mathbf{v}^x, \mathbf{w}^1_{\mathrm{prof}} > + < \underline{\mathbf{v}^y}, \mathbf{w}^2_{\mathrm{prof}} >]$$

Combinations

models capture different aspects of the data, combine them (e.g., NF)

$$p(y_{\mathrm{emp}}^{x,y} = 1 | \dots) \propto \exp[< \mathbf{f}_{\mathrm{emp}}^{x,y}, \mathbf{w}_{\mathrm{emp}}^{\mathrm{N}} > + < \mathbf{v}^{x,y}, \mathbf{w}_{\mathrm{emp}}^{\mathrm{F}} >]$$

Training

Negative Data

Usually **unavailable** or **sparse**, so...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
	1		1	
1				

Sample Unobserved Cells as Negative

Can work...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
1	**1**	**0**		
	1		**1**	
1				

Subsample

but often **does not**

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
1	**1**			
0	**1**		**1**	
1				

Subsample

and you need to sample a lot (wasting resources)

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
0	1	1	0	1
1	1	0		0
0	1	0	1	
1	0		0	

Implicit Feedback

Often users only **click/view/buy** items, or not, but **no rating**

User 1	User 2	User 3	User 4	User 5	
	1	1		1	Item 1
1	1				Item 2
	1		1		Item 3
1					Item 4

Ranking

[Rendle et al.,09]

for all (**o**bserved,**n**ot observed) pairs in column: prob(o) > prob(n)

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
0.9	**1**			
0.95	**1**		**1**	
1				

Ranking

[Rendle et al.,09]

for all (**o**bserved,**n**ot observed) pairs in a column: prob(o) > prob(n)

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
0.9	**1**			
0.85	**1**		**1**	
1				

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-**Y**

0.9

$\vee$

0.85

1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} = < \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-Y
0.9
$\vee$
0.85
1

$$\underline{\max_{\mathbf{v},\mathbf{w},\ldots}} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} = < \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-**Y**

0.9

0.85

1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} =< \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-Y
0.9
>
0.85
1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{\underline{r(x,y)}} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} = < \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-**Y**

0.9

$\vee$

0.85

1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} = < \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-**Y**

0.9

$\vee$

0.85

1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} =< \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-Y
0.9
∨
0.85
1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\underline{\theta_r^{x,y} - \theta_r^{x',y'}})]$$

for example: $\theta_r^{x,y} = < \mathbf{v}^{x,y}, \mathbf{w}_r >$

Ranking

[Rendle et al.,09]

Train by maximizing a LogLikelihood variant

X-*is-historian-at*-**Y**

0.9

$>$

0.85

1

$$\max_{\mathbf{v},\mathbf{w},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_r^{x,y} - \theta_r^{x',y'})]$$

for example: $\theta_r^{x,y} =< \mathbf{v}^{x,y}, \mathbf{w}_r >$

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Sample observed fact...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
	1		1	
1				

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Sample unobserved cell for same relation

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
	1		1	
1				

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Estimate current beliefs and gradient, **update** parameters accordingly

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	0.8	1		1
1	1			
	1		1	
1	0.9			

Training: Stochastic Gradient Descent

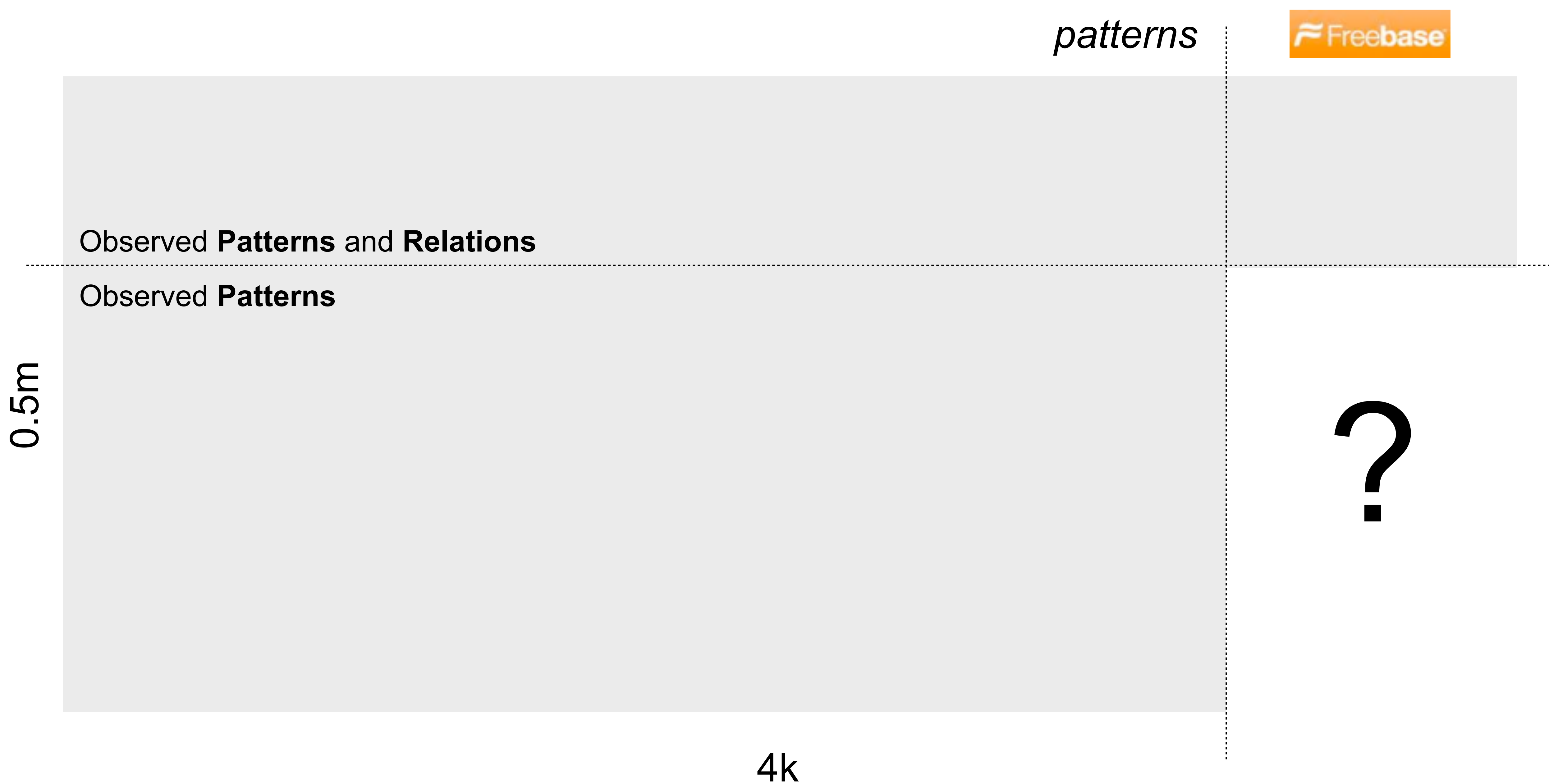
[Rendle et al.,09]

Estimate current beliefs and gradient, **update** parameters accordingly

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	0.85	1		1
1	1			
	1		1	
1	0.80			

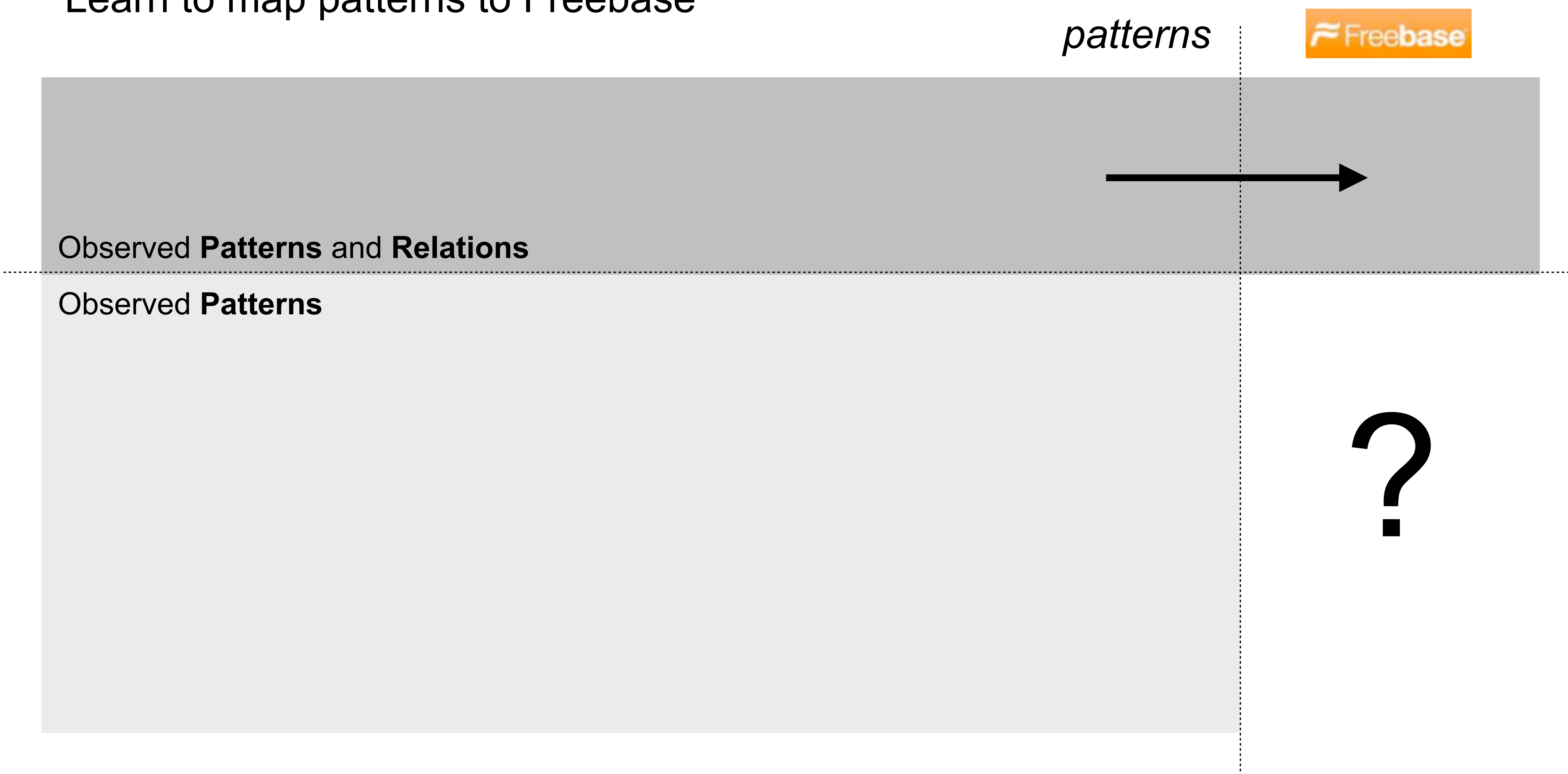
Evaluation

Setup


patterns
Freebase
Observed **Patterns** and **Relations**
Observed **Patterns**
0.5m
?
4k

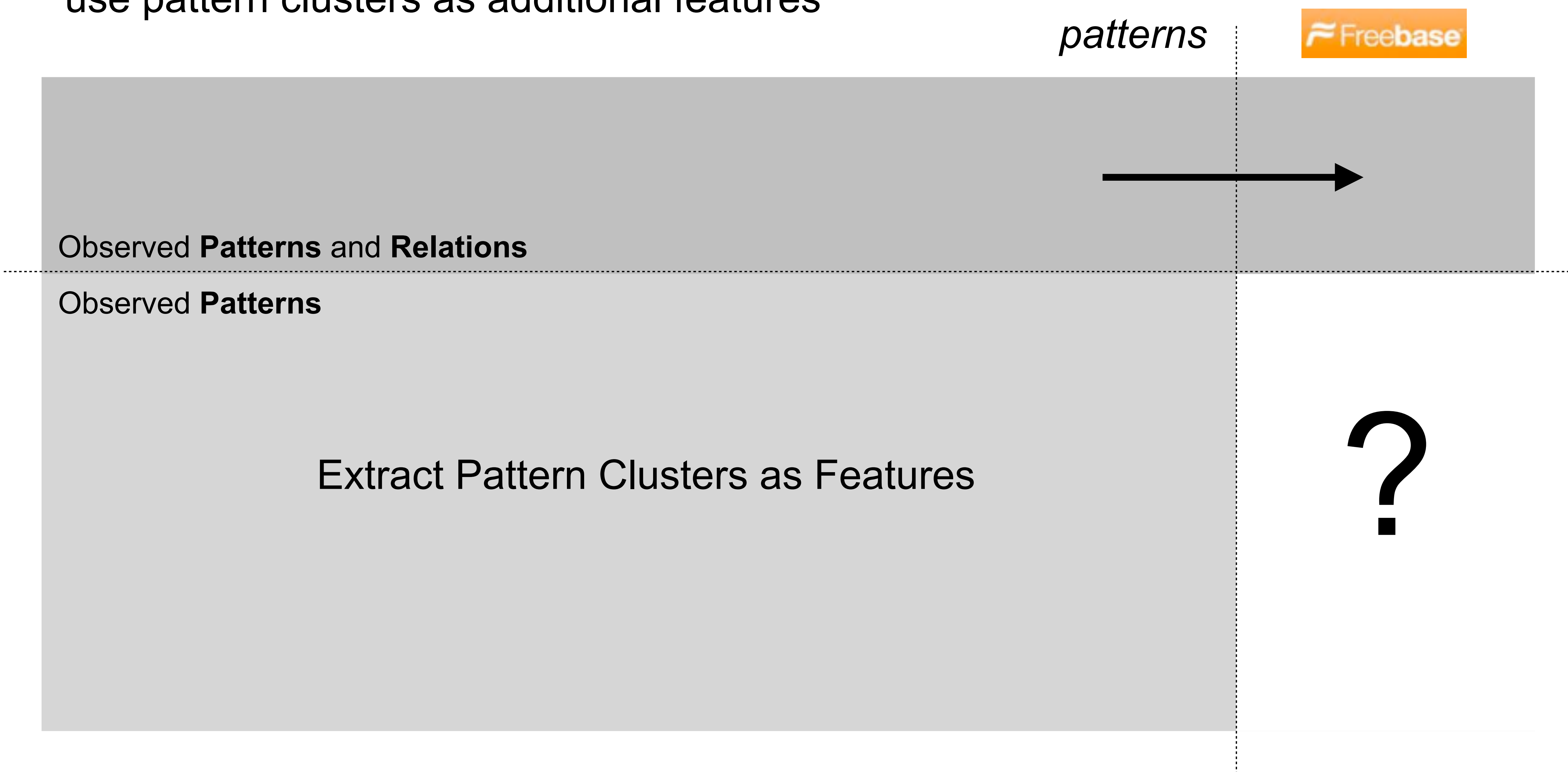
Baseline: Mintz 2009

Learn to map patterns to Freebase



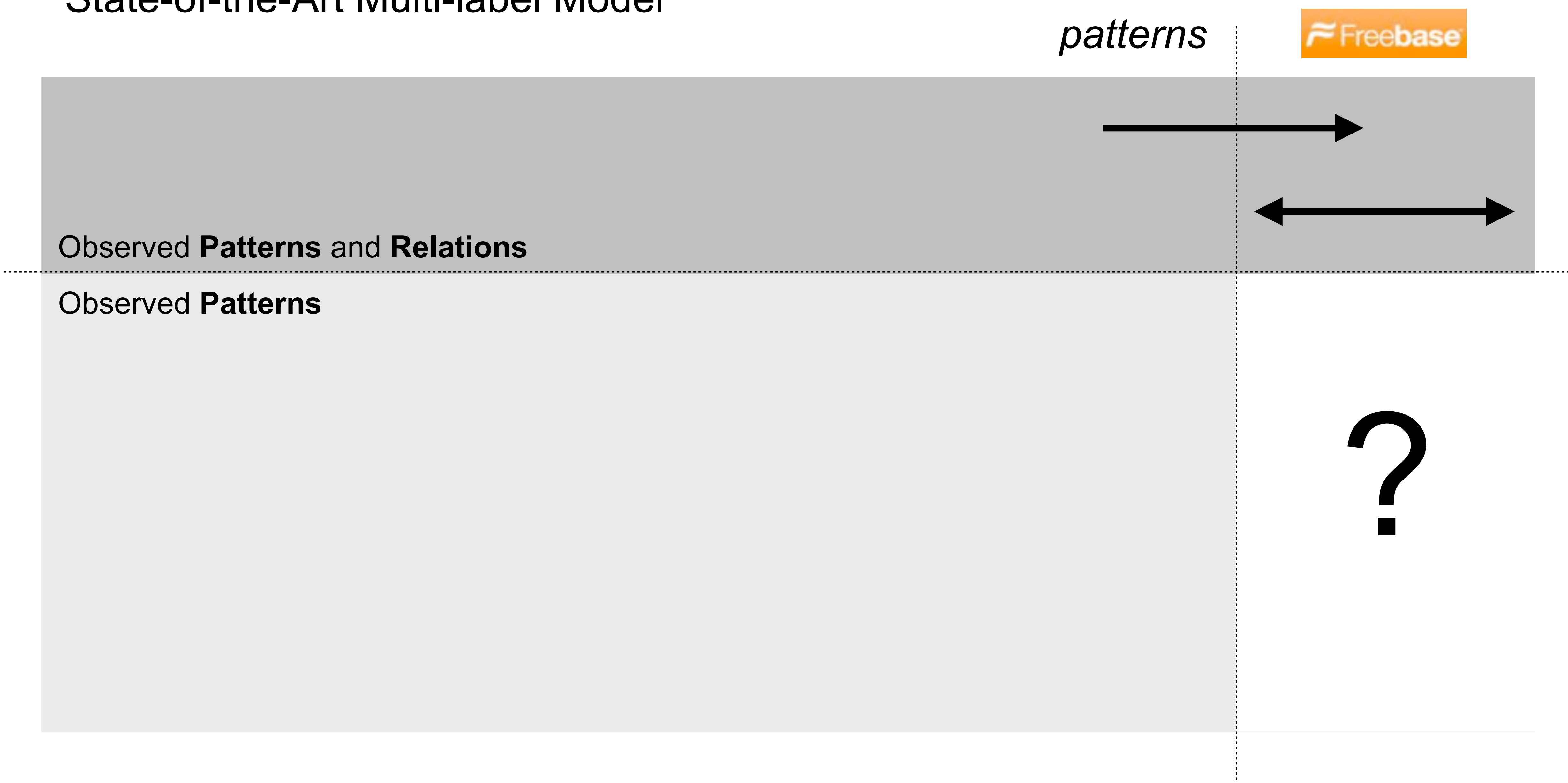
Baseline: Yao 2011

use pattern clusters as additional features



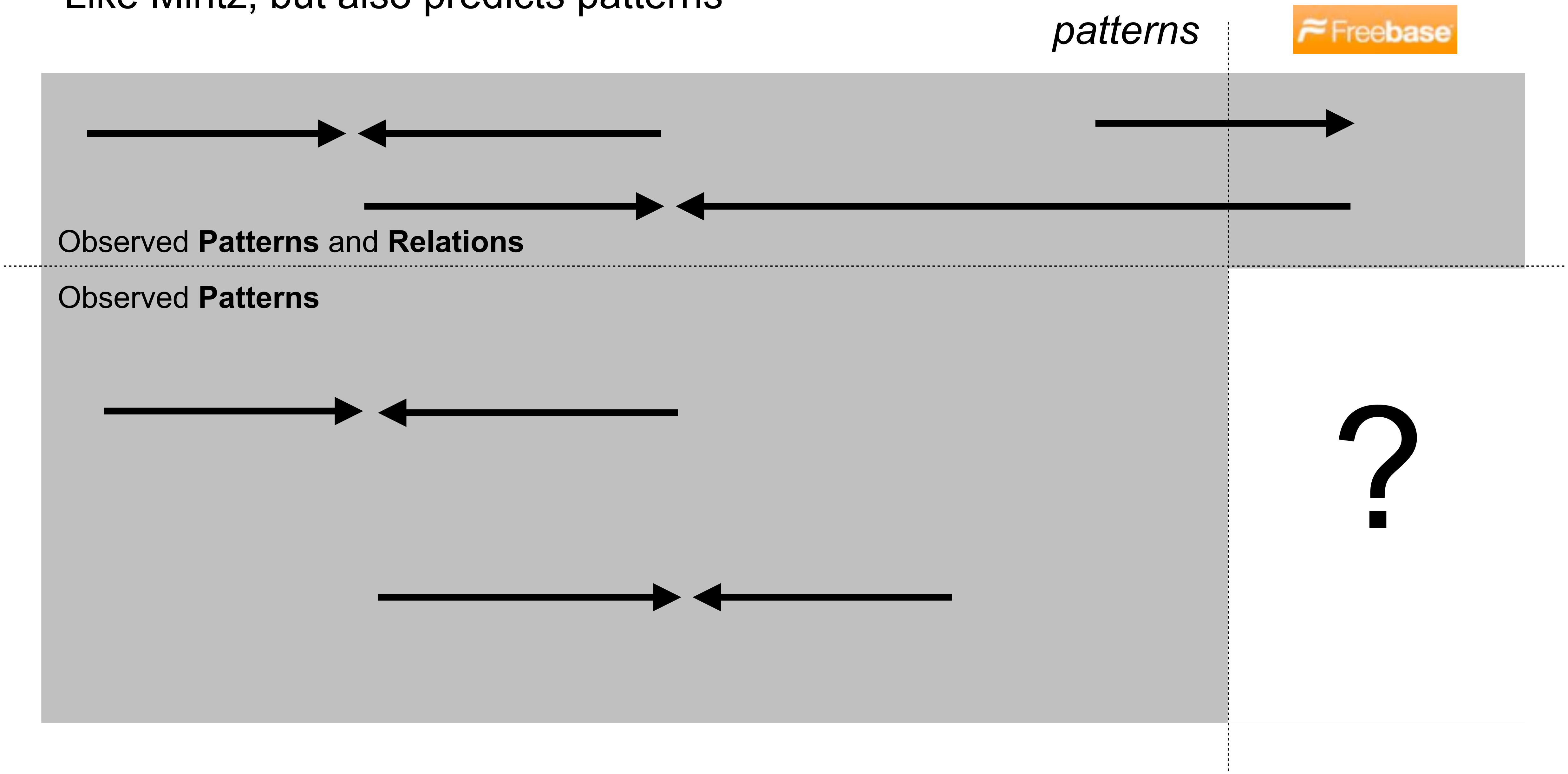
Baseline: Surdenau 2012

State-of-the-Art Multi-label Model



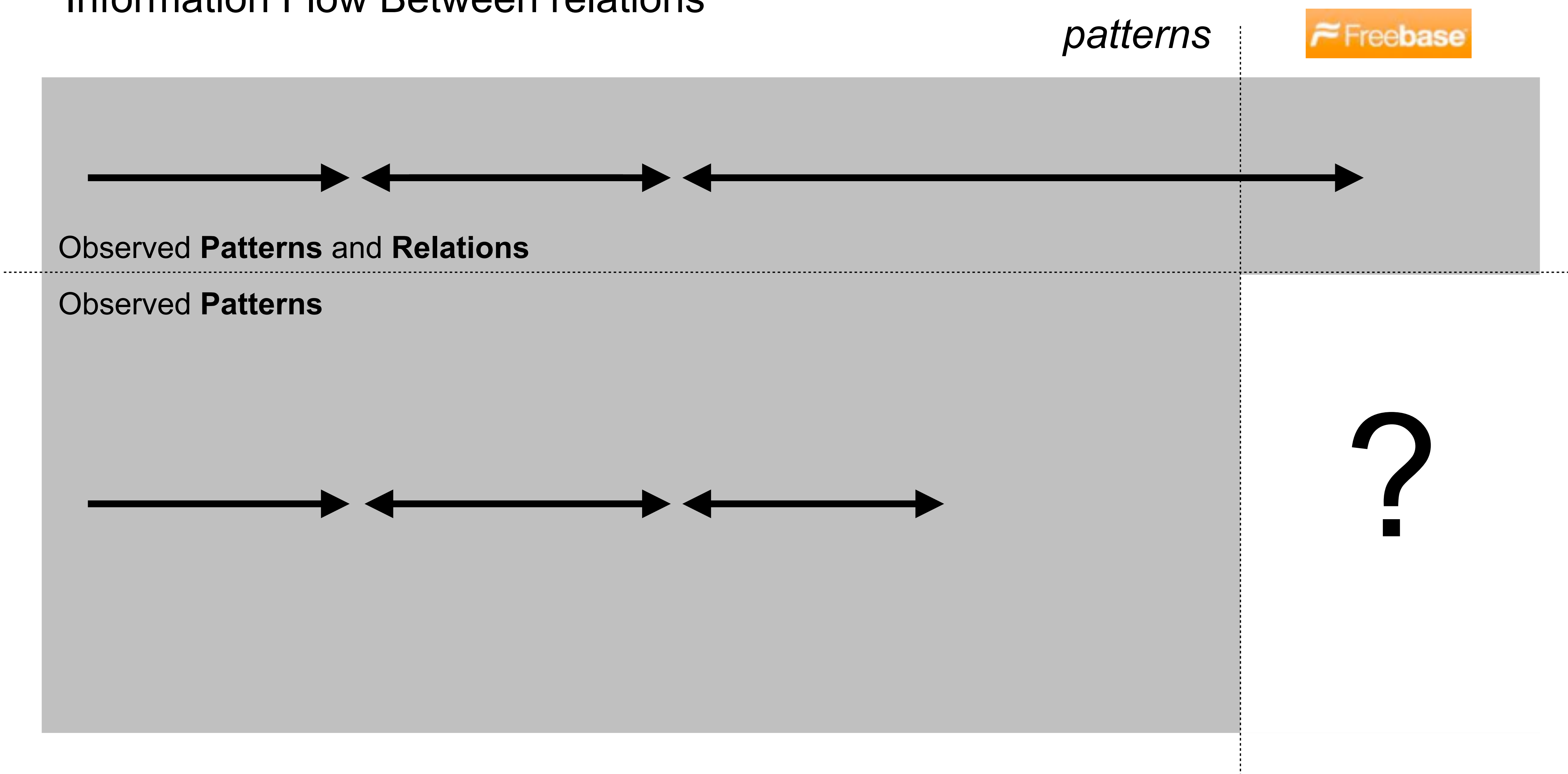
Model N

Like Mintz, but also predicts patterns



Model F, E, NF, NFE ...

Information Flow Between relations



Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

Evaluate **average precision** per **Freebase relation**.

Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

Evaluate **average precision** per **Freebase relation**.

Relation
employee
containedby
parents
...
Weighted MAP
MAP

Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

Evaluate **average precision** per **Freebase relation**.

Relation	MI09
employee	0.67
containedby	0.48
parents	0.24
...	...
Weighted MAP	0.48
MAP	0.32

Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

Evaluate **average precision** per **Freebase relation**.

Relation	MI09	YA11
employee	0.67	0.64
containedby	0.48	0.51
parents	0.24	0.27
...	...	...
Weighted MAP	0.48	0.52
MAP	0.32	0.42

Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

Evaluate **average precision** per **Freebase relation**.

Relation	MI09	YA11	SU12
employee	0.67	0.64	0.70
containedby	0.48	0.51	0.54
parents	0.24	0.27	**0.58**
...	...	...	...
Weighted MAP	0.48	0.52	0.57
MAP	0.32	0.42	0.56

Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

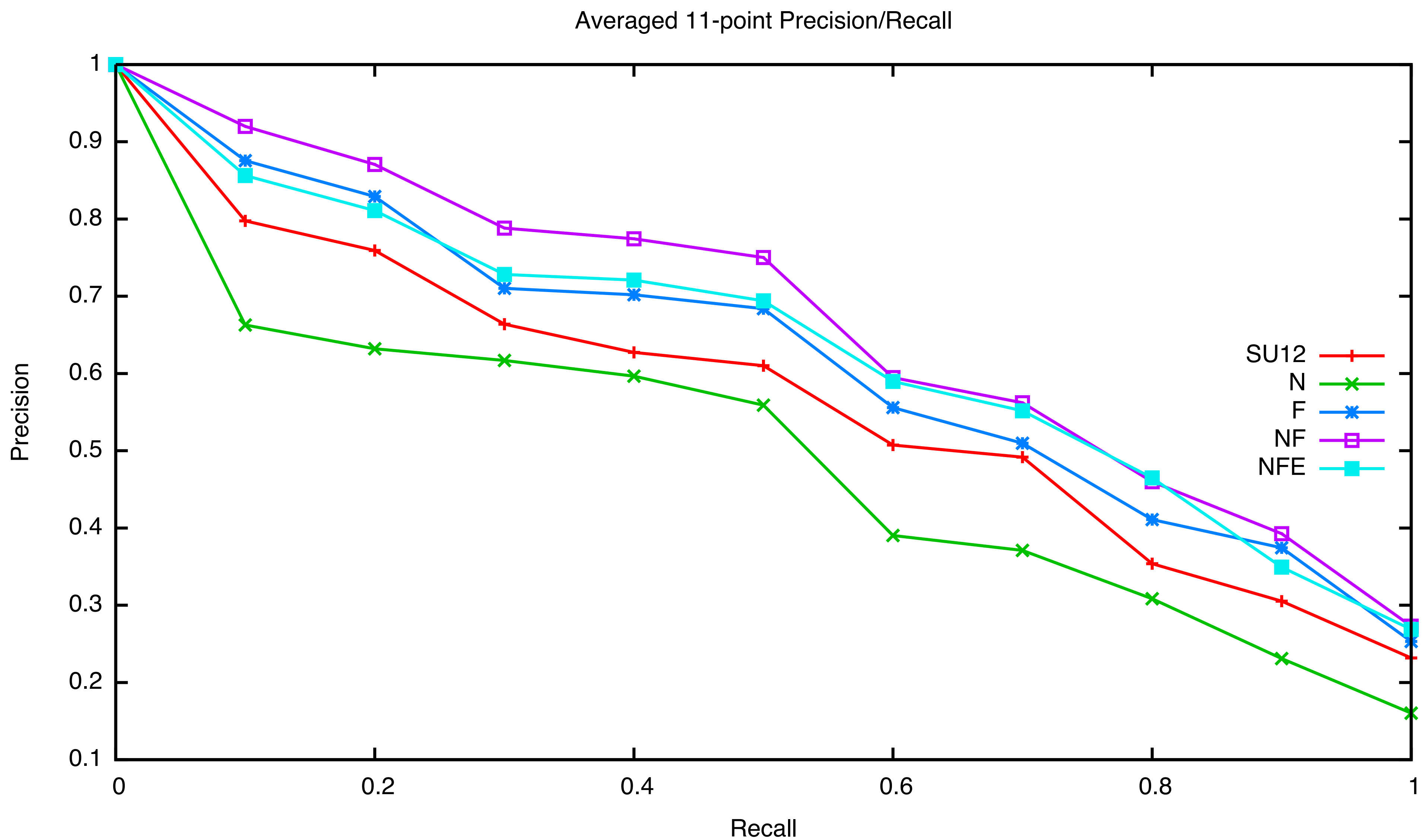
Evaluate **average precision** per **Freebase relation**.

Relation	MI09	YA11	SU12	N+F+E
employee	0.67	0.64	0.70	**0.79**
containedby	0.48	0.51	0.54	**0.69**
parents	0.24	0.27	**0.58**	0.39
...	...	...	...	...
Weighted MAP	0.48	0.52	0.57	**0.69**
MAP	0.32	0.42	0.56	**0.63**

Evaluation (Structured)

[Mintz 09; Yao 11; Surdenau 12]

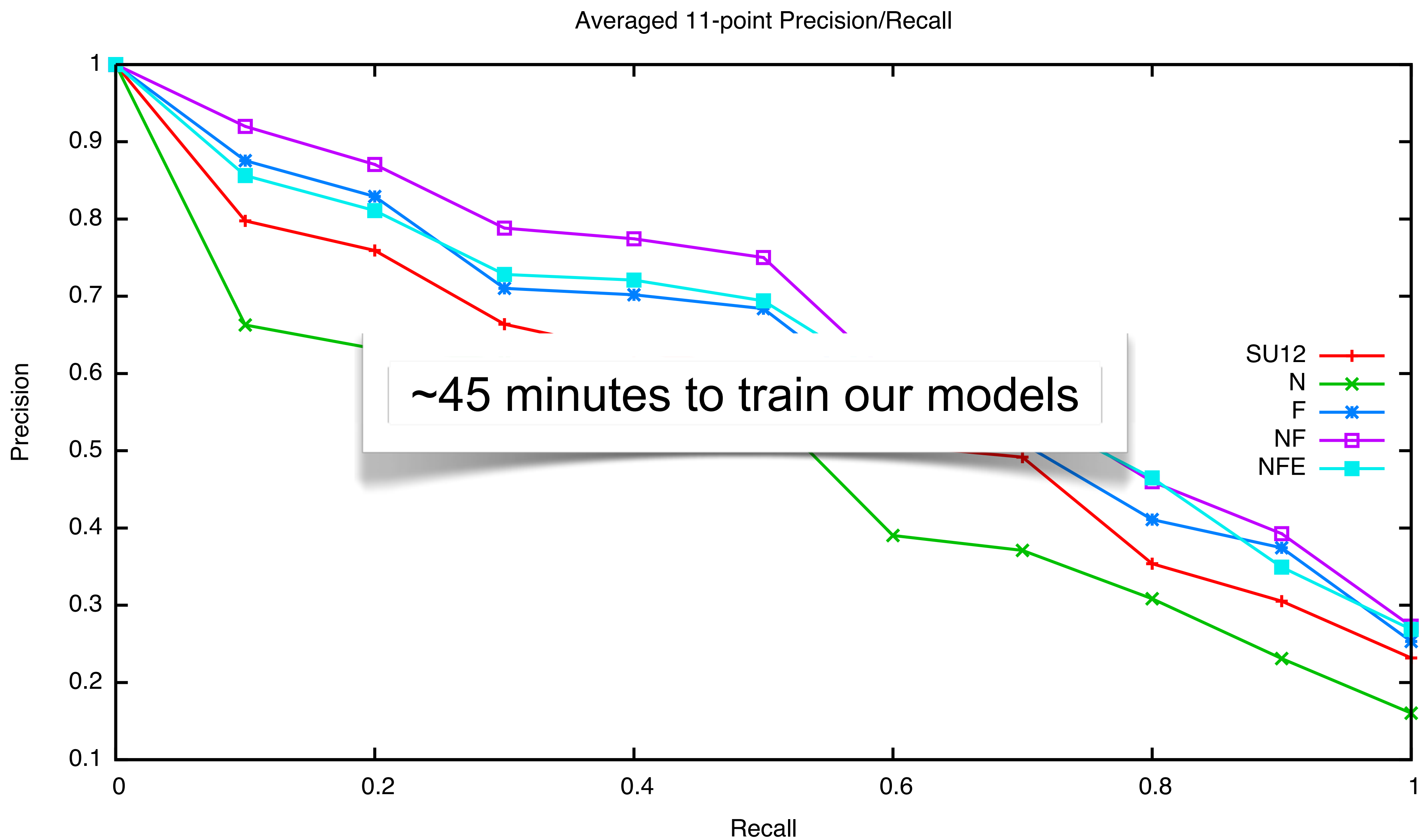
Averaged 11 point precision recall curve



Evaluation (Structured)

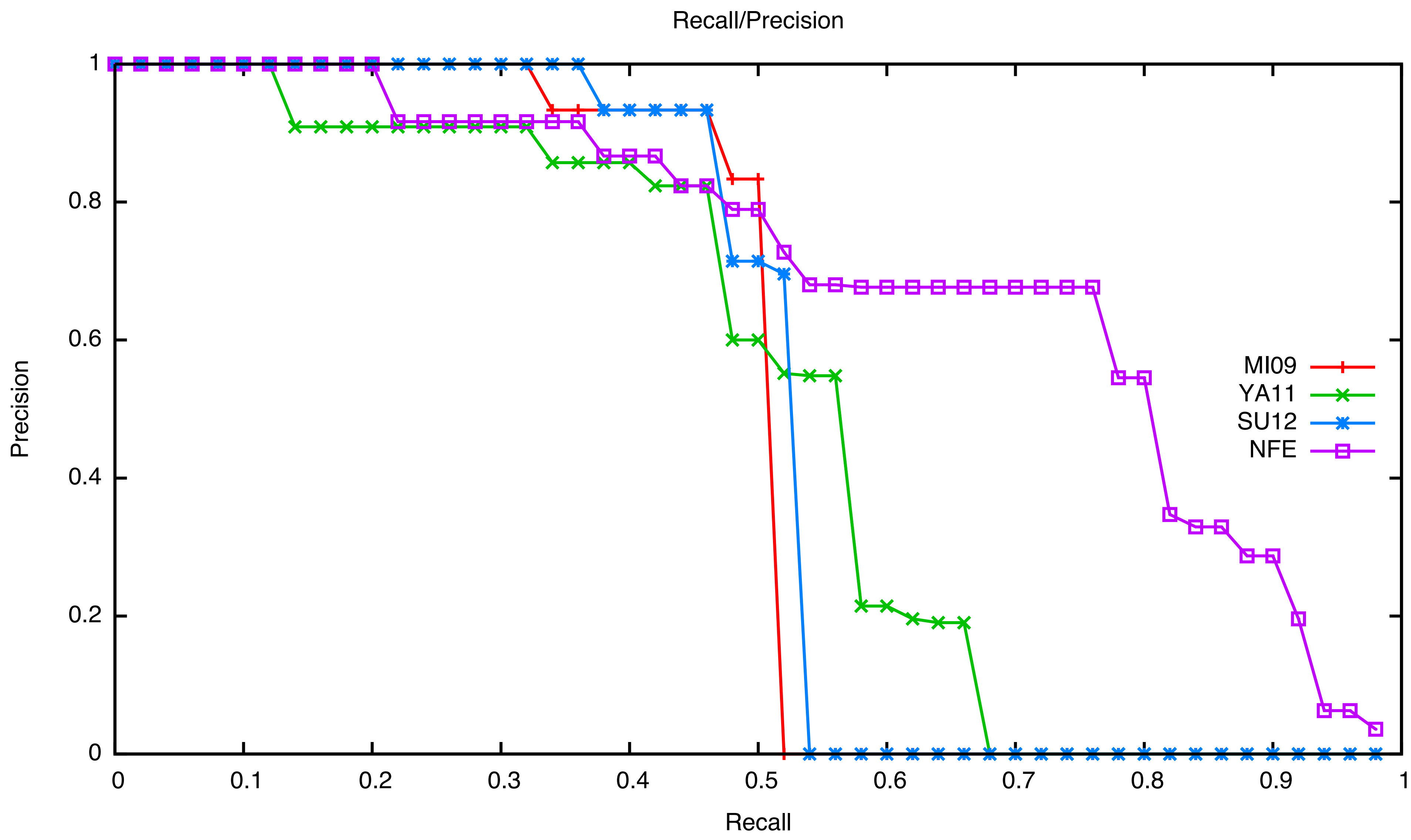
[Mintz 09; Yao 11; Surdenau 12]

Averaged 11 point precision recall curve



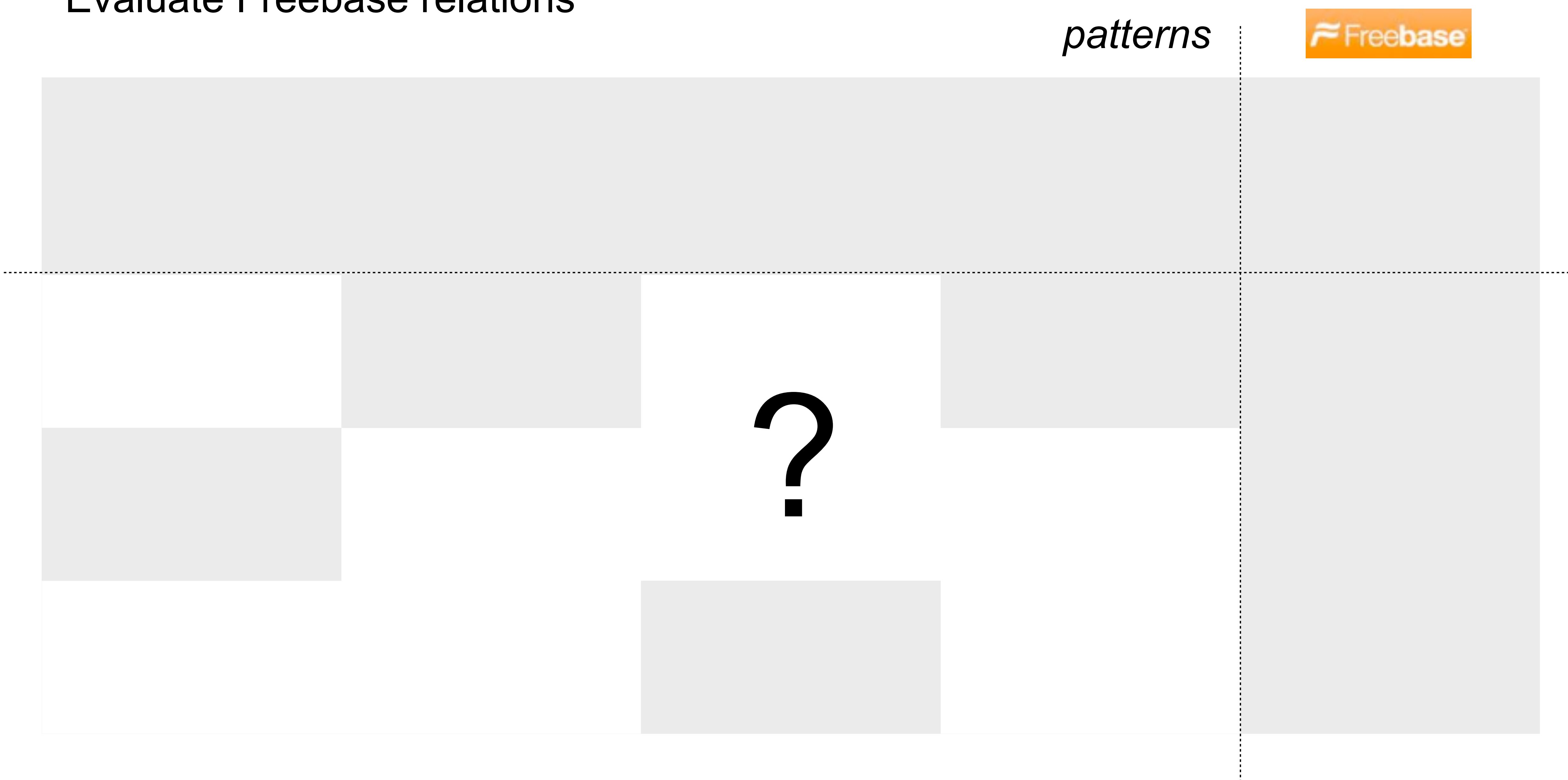
Evaluation (Structured)

Precision Recall curve for *works_written*



Setup

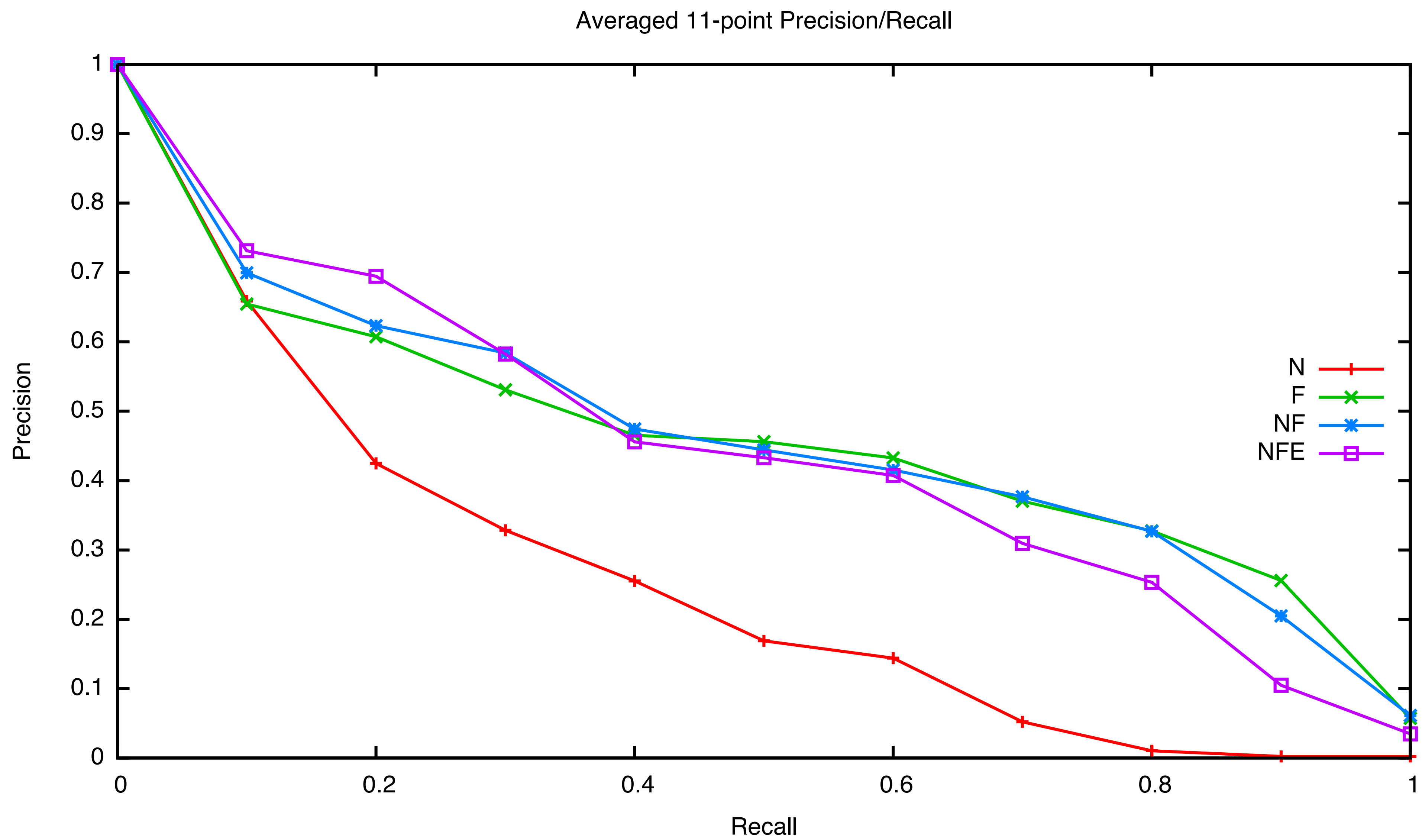
Evaluate Freebase relations



Evaluation (Patterns)

[Mintz 09; Yao 11; Surdenau 12]

Averaged 11 point precision recall curve



Conclusion

- Challenge: Relations **w/o preexisting databases**
- Solution: Extraction in **Universal Schemas** with ...
- ... **Patterns**-based + **Structured** Relations
- Latent Feature models support **information flow**...
- ... and **outperform** classifiers to get ...
- ... **State-of-the-art** results

Thanks!

Ranking

[Rendle et al.,09]

Train by maximizing a LL variant

X-*is-historian-at*-**Y**

0.9

$>$

0.85

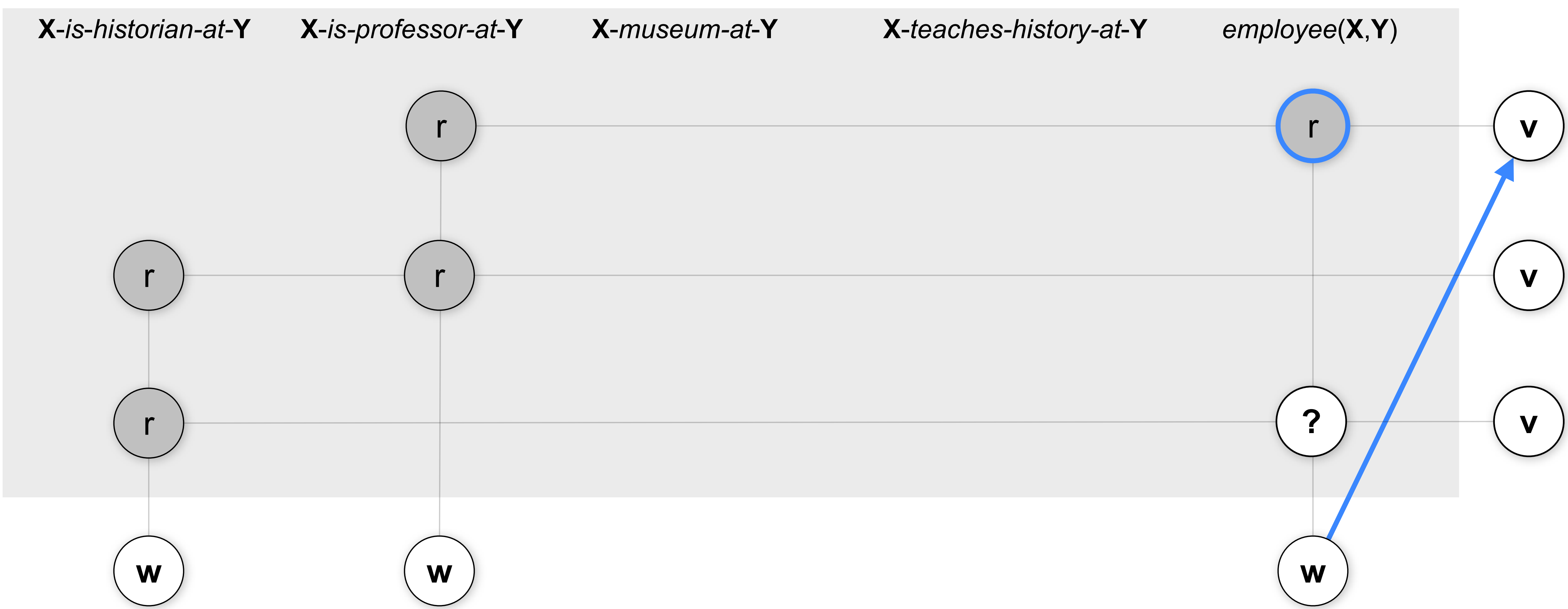
1

$$\max_{\mathbf{w},\mathbf{v},\ldots} \sum_{r} \sum_{r(x,y)} \sum_{\neg r(x',y')} \log[\sigma(\theta_{r,x,y} - \theta_{r,x',y'})]$$

for example: $\theta_{r,x,y} =< \mathbf{v}^{x,y}, \mathbf{w}_{\mathrm{r}} >$

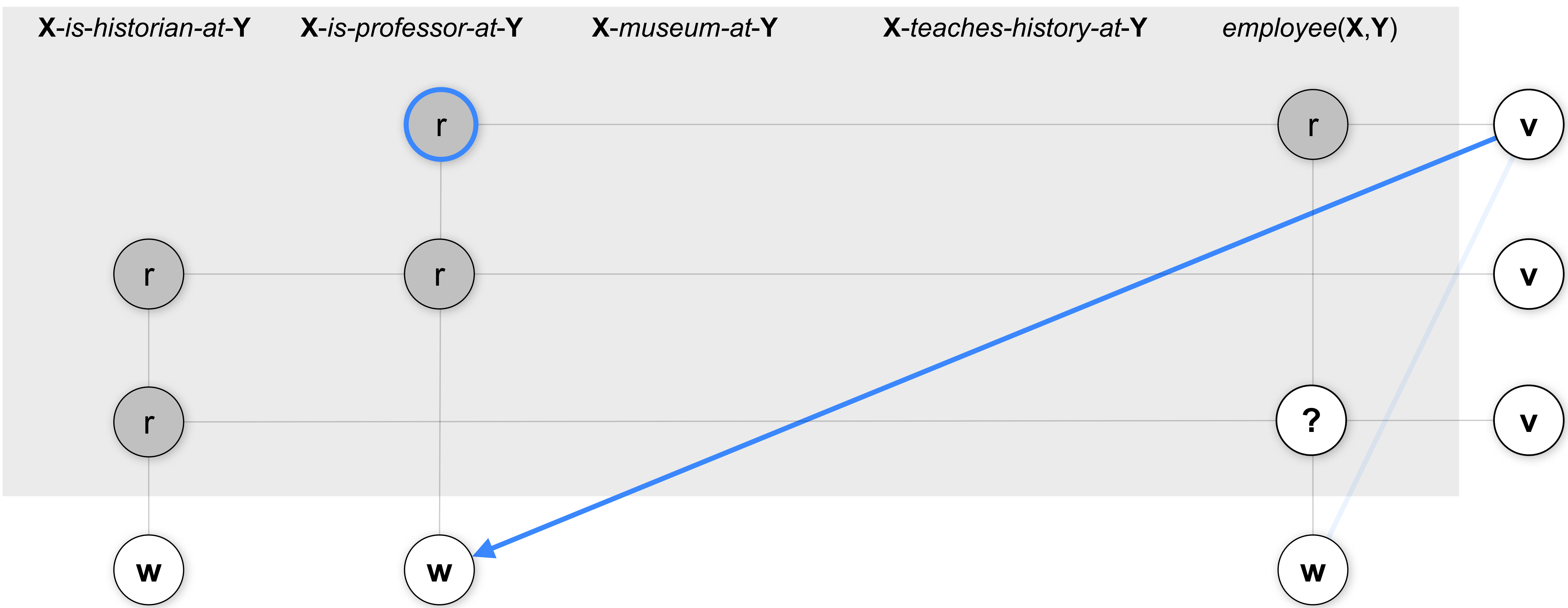
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



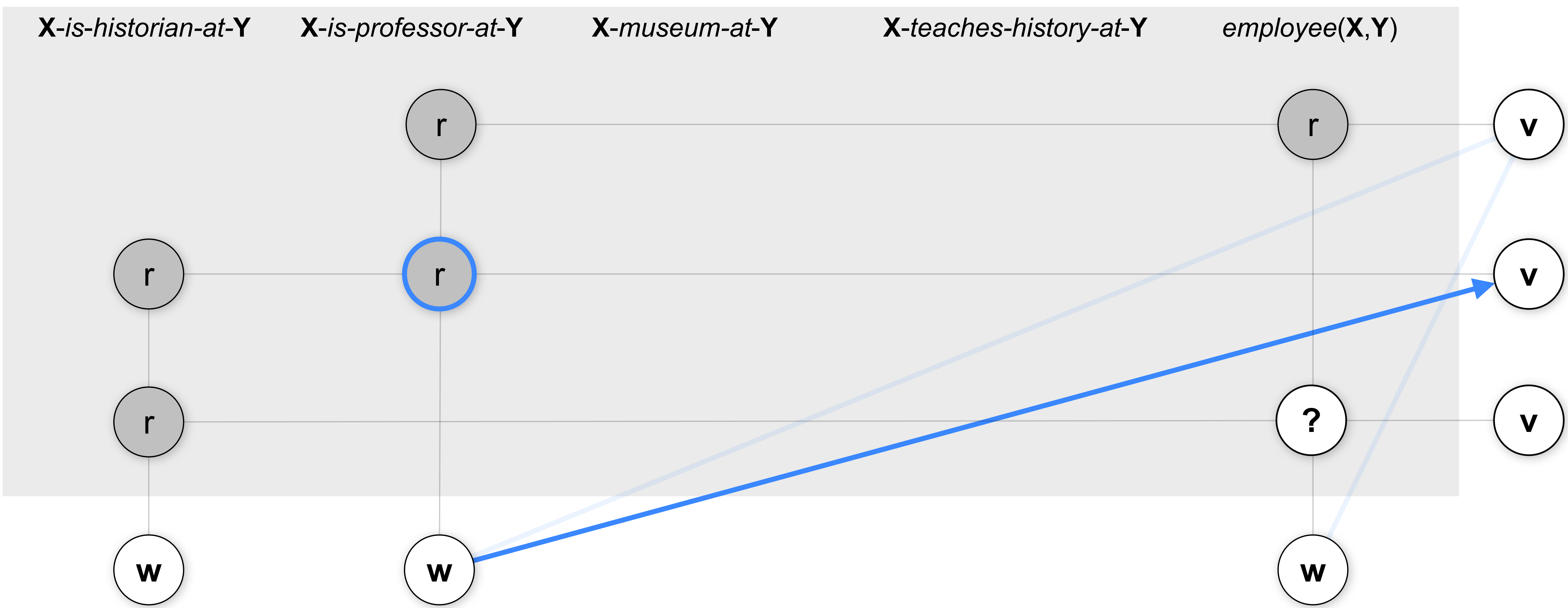
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



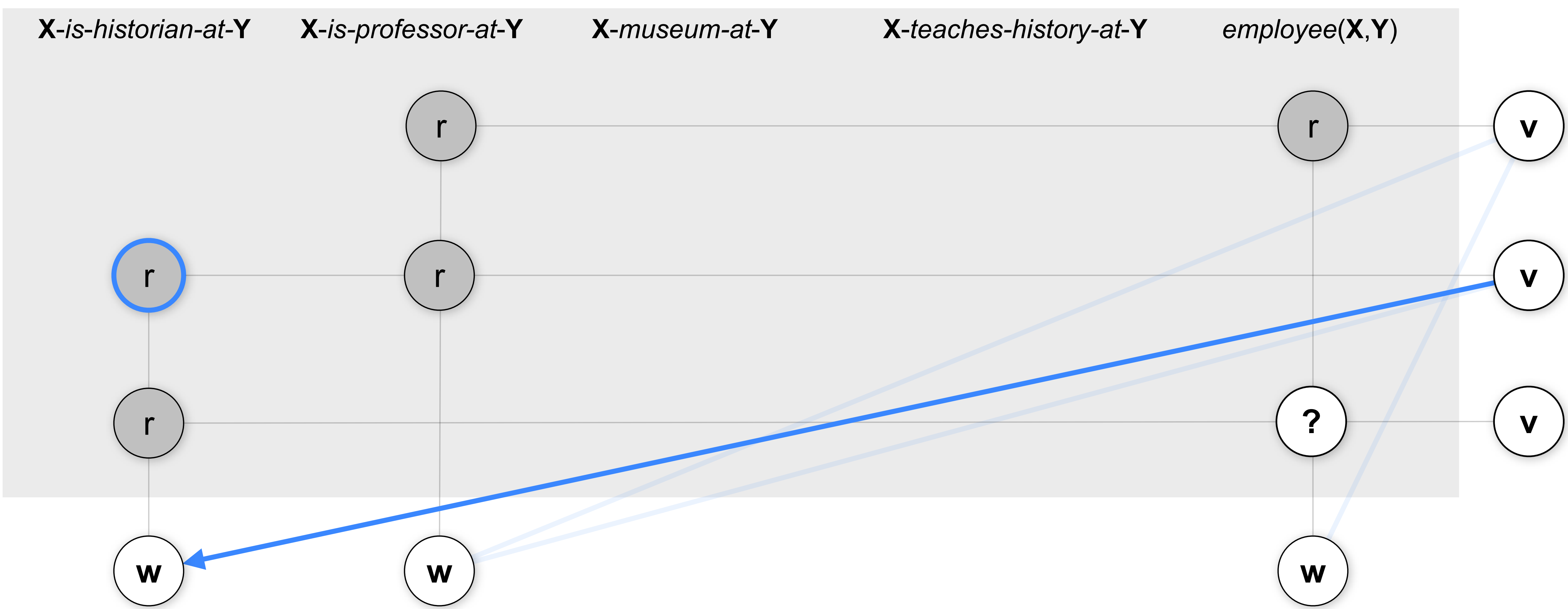
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



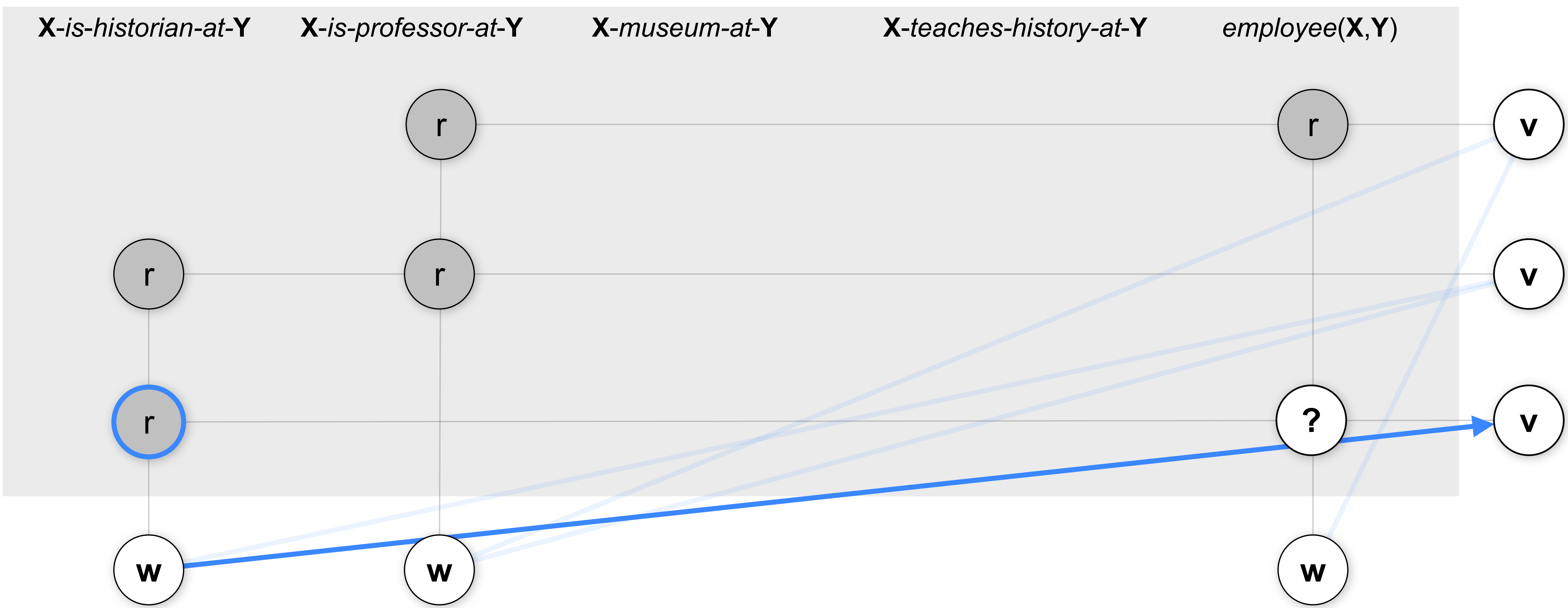
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



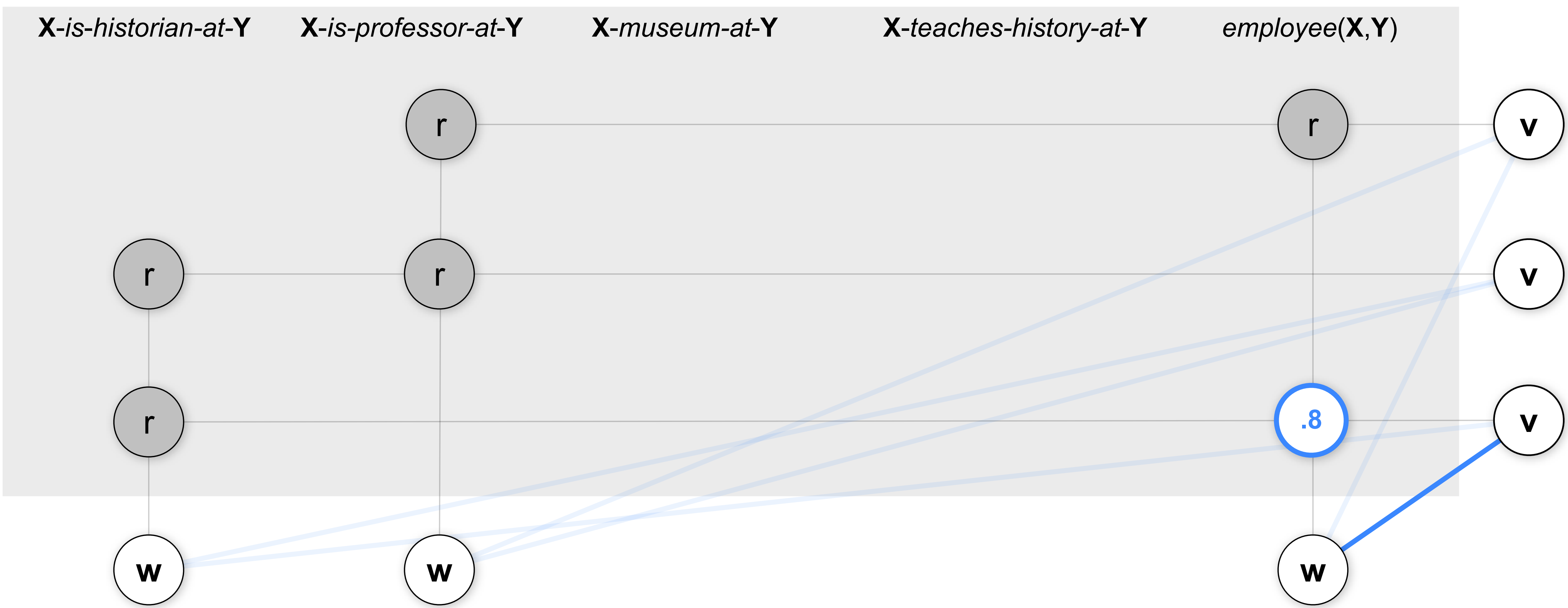
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



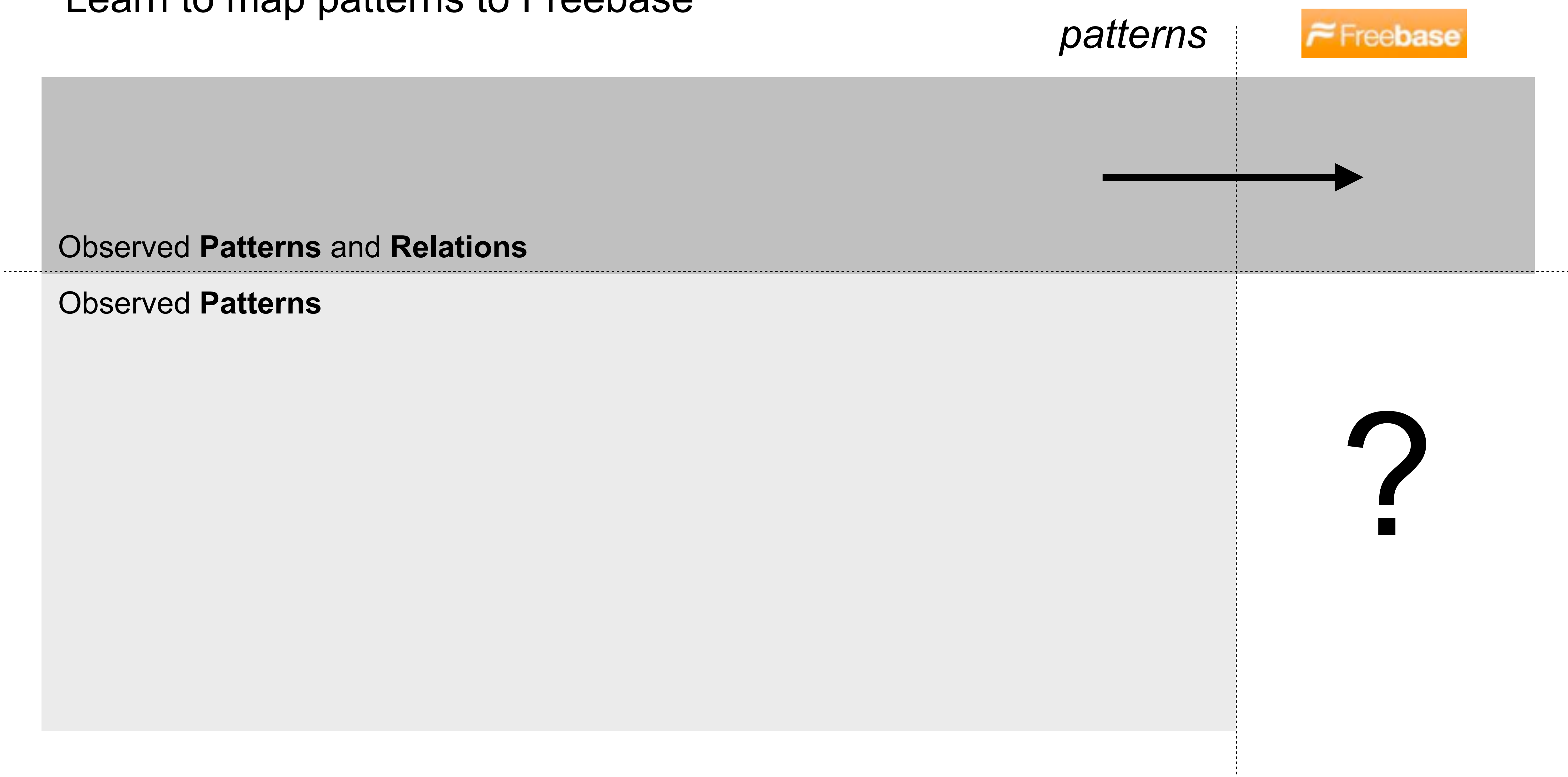
Model F: Latent Feature (Factorization)

Pattern reasoning helps structured relations



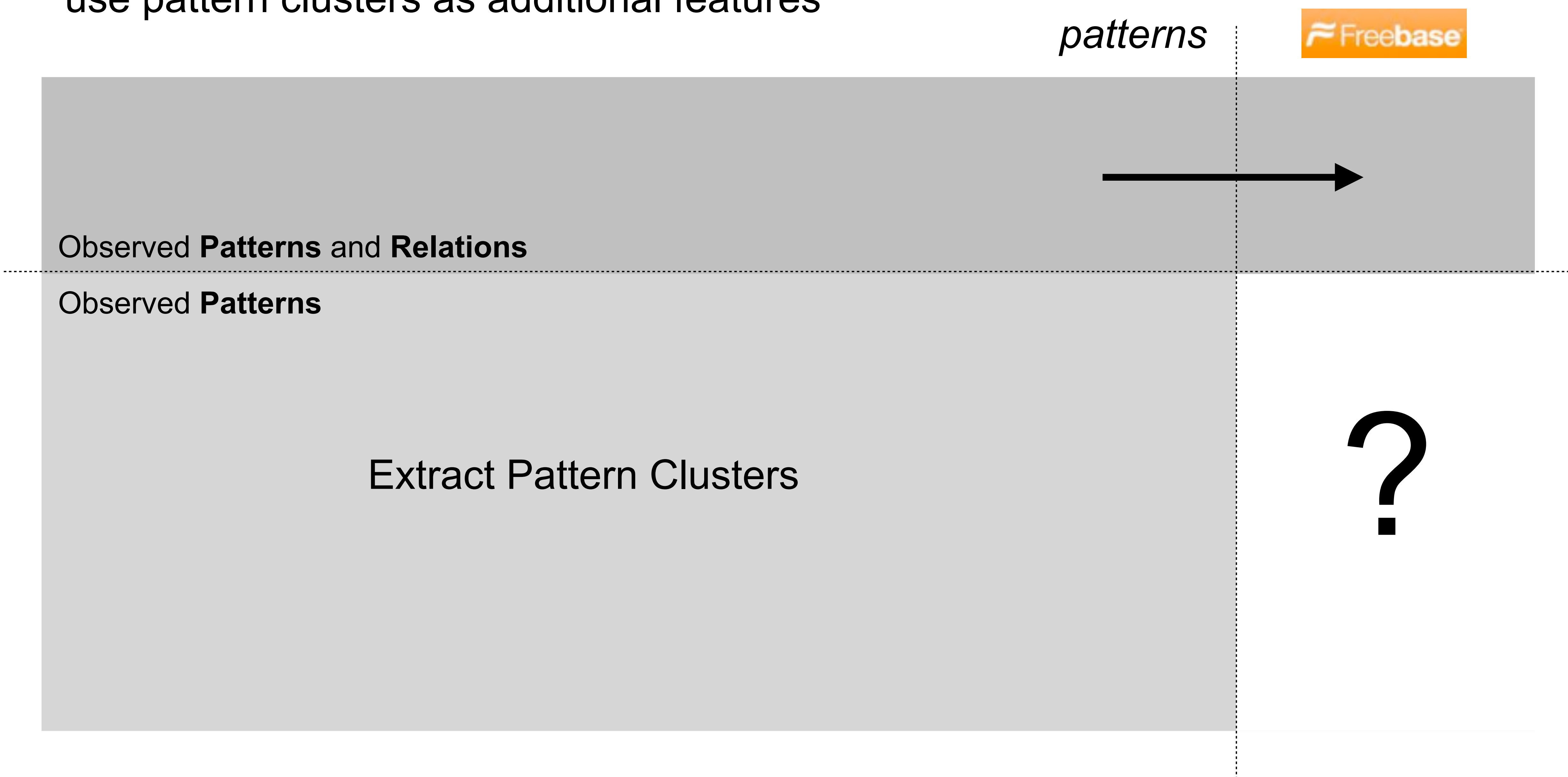
Baseline: Mintz 2009

Learn to map patterns to Freebase



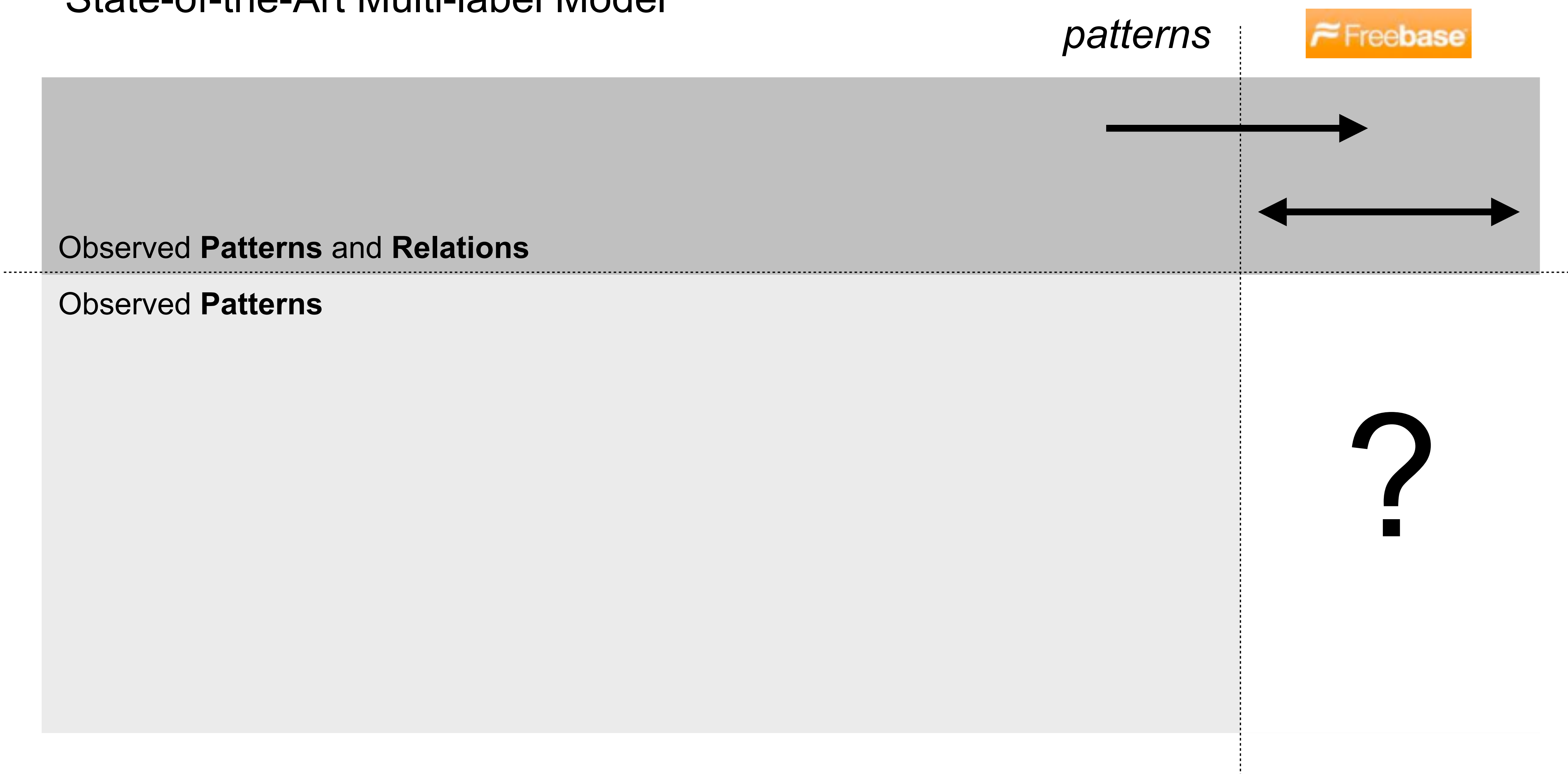
Baseline: Yao 2011

use pattern clusters as additional features



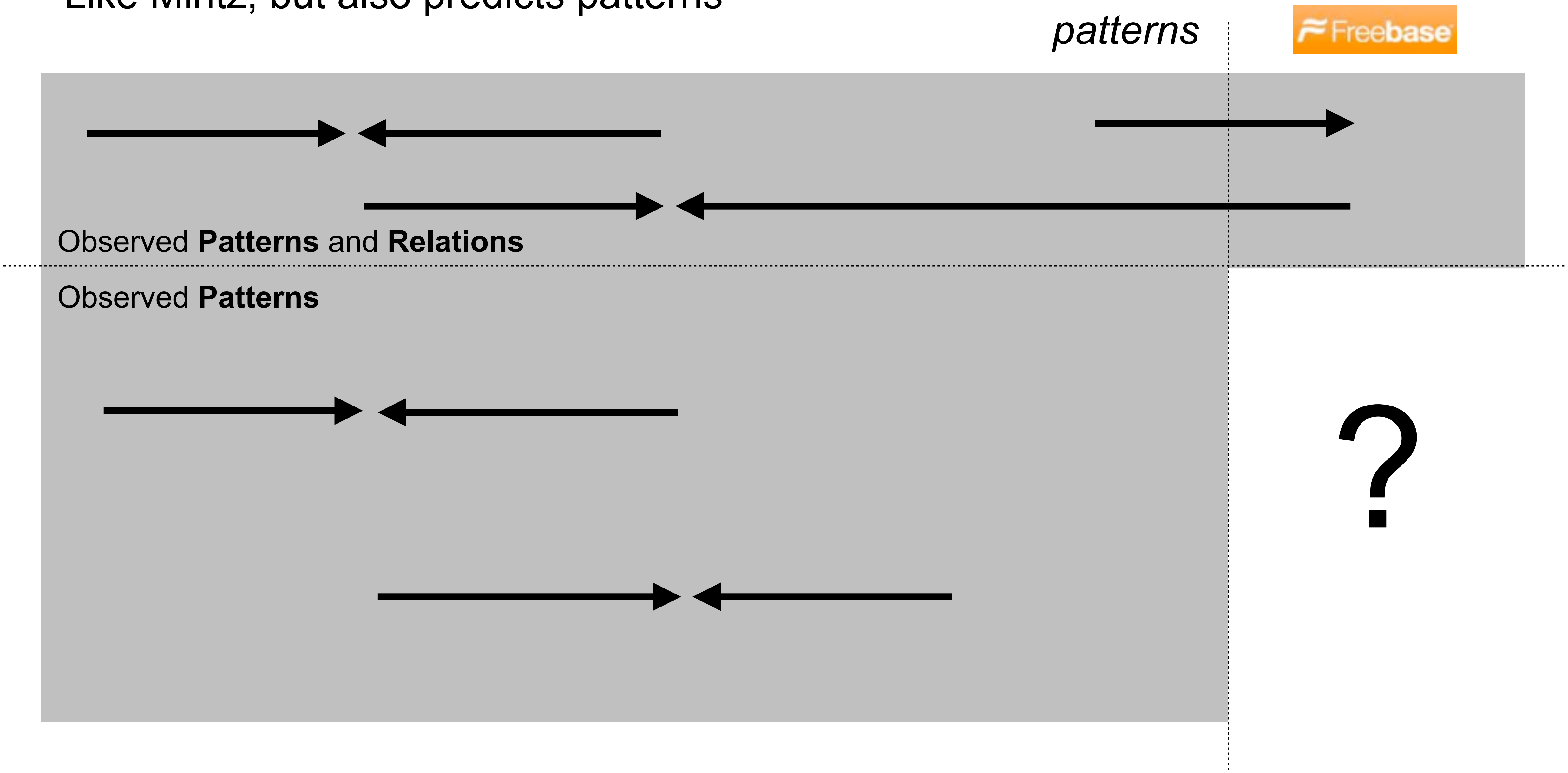
Baseline: Surdenau 2012

State-of-the-Art Multi-label Model



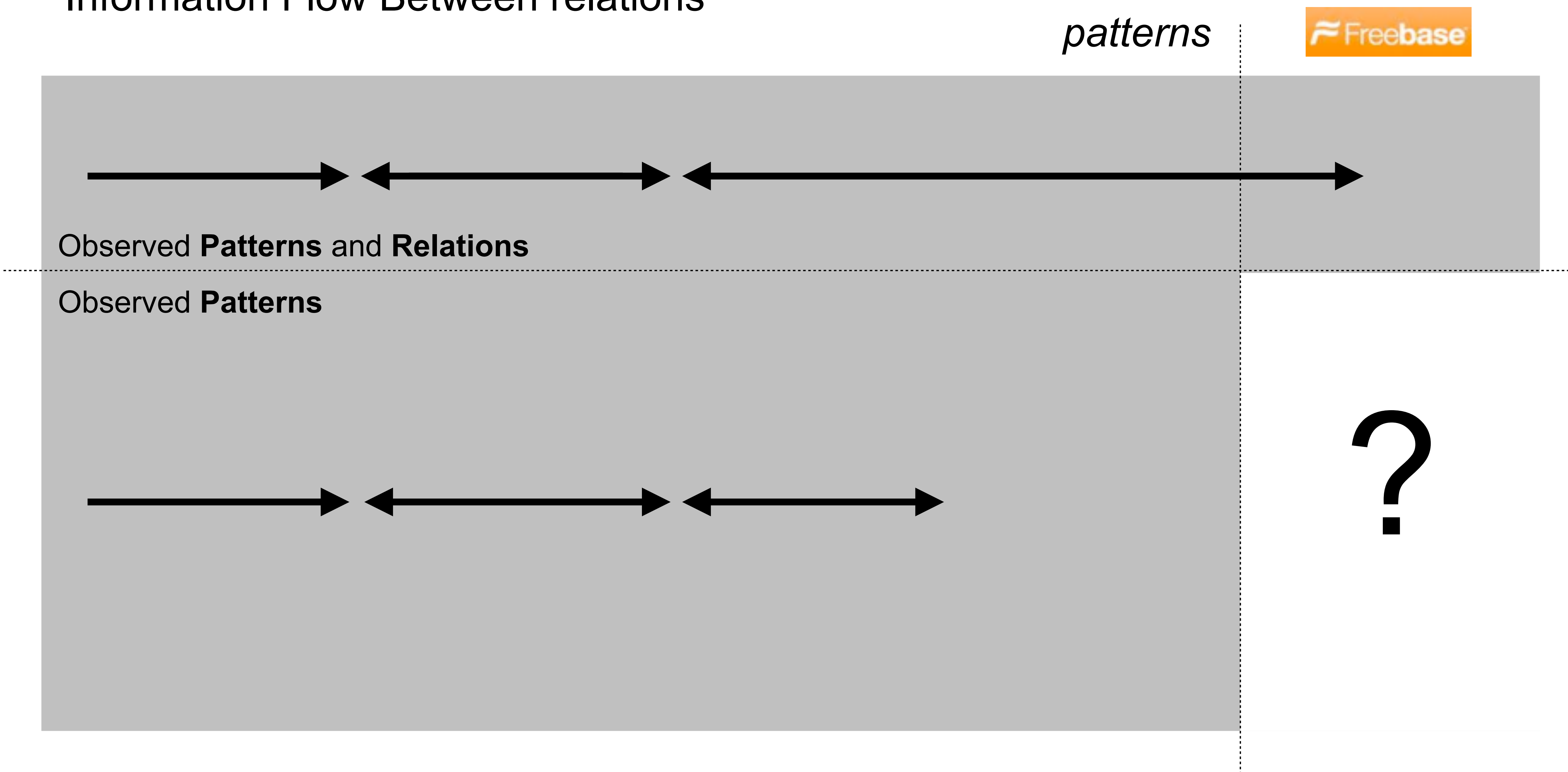
Model N

Like Mintz, but also predicts patterns



Model F, E, NF, NFE ...

Information Flow Between relations



Training

Negative Data

Usually **unavailable** or **sparse**, so...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
1	**1**			
	1		**1**	
1				

Sample Unobserved Cells as Negative

Can work...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1	0		
	1		1	
1				

Subsample

but often **does not** (and wastes resources)

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
0	1		1	
1				

Subsample

and you need to sample a lot (wasting resources)

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
0	1	1	0	1
1	1	0		0
0	1	0	1	
1	0		0	

Ranking

[Rendle et al.,09]

for all (**o**bserved,**n**ot observed) pairs in column: prob(o) > prob(n)

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**		**1**
0.9	**1**			
0.95	**1**		**1**	
1				

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Sample observed fact...

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
	1		1	
1				

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Sample unobserved cell for same relation

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			
	1		1	
1				

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Estimate current beliefs and gradient, **update** parameters accordingly

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	0.8	1		1
1	1			
	1		1	
1	0.9			

Training: Stochastic Gradient Descent

[Rendle et al.,09]

Estimate current beliefs and gradient, **update** parameters accordingly

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	0.85	1		1
1	1			
	1		1	
1	0.80			