

# Cross-lingual infobox alignment in Wikipedia using Entity-Attribute Factor Graph

Yan Zhang, Juanzi Li, Jing Zhang

Knowledge Engineering Group, Tsinghua University, Beijing China

[z-y14@mails.tsinghua.edu.cn](mailto:z-y14@mails.tsinghua.edu.cn)

[lijuanzi@tsinghua.edu.cn](mailto:lijuanzi@tsinghua.edu.cn)

[jing-zha15@mails.tsinghua.edu.cn](mailto:jing-zha15@mails.tsinghua.edu.cn)



# Outline

- Background: Wikipedia infobox alignment
- Our Approach: Entity-Attribute Factor Graph
- Experiments & Conclusion
- Future Work



# Outline

- Background: Wikipedia infobox alignment
- Our Approach: Entity-Attribute Factor Graph
- Experiments & Conclusion
- Future Work

# An Example

<b>Born</b>	Mark Elliot Zuckerberg May 14, 1984 (age 32) <sup>[1]</sup> White Plains, New York, U.S.	<b>出生</b>	马克·埃利奥特·扎克伯格 Mark Elliot Zuckerberg 1984年5月14日 (32岁) <sup>[1]</sup> 美国纽约州白原市	<b>Nom de naissance</b>	Mark Elliot Zuckerberg
<b>Residence</b>	Palo Alto, California, U.S.	<b>居住地</b>	美国加利福尼亚州帕罗奥图 <sup>[2]</sup>	<b>Naissance</b>	14 mai 1984 (32 ans) White Plains, État de New York (États-Unis)
<b>Alma mater</b>	Harvard University	<b>母校</b>	哈佛大学	<b>Nationalité</b>	 Américain
<b>Occupation</b>	Computer programmer, internet entrepreneur	<b>职业</b>	软件设计师、创业家	<b>Pays de résidence</b>	États-Unis
<b>Years active</b>	2004–present	<b>活跃时期</b>	2004年至今	<b>Profession</b>	Informaticien, entrepreneur, programmeur
<b>Known for</b>	Co-founder of Facebook	<b>知名于</b>	Facebook创始人	<b>Activité principale</b>	Fondateur et CEO de Facebook
<b>Home town</b>	Dobbs Ferry, New York, U.S.	<b>家乡</b>	美国纽约州多布斯费里	<b>Formation</b>	Phillips Exeter Academy Harvard College
<b>Salary</b>	One-dollar salary <sup>[2]</sup>	<b>薪金</b>	1美元 <sup>[3]</sup>	<b>Conjoint</b>	Priscilla Chan
<b>Net worth</b>	▲ US \$53.6 billion (2017 estimate) <sup>[3]</sup>	<b>净资产</b>	▲ US\$553亿 (2016年10月) <sup>[4]</sup>	<b>Descendants</b>	Maxima Zuckerberg
<b>Title</b>	Chairman and CEO of Facebook	<b>头衔</b>	Facebook董事长兼首席执行官		
<b>Spouse(s)</b>	Priscilla Chan (m. 2012)	<b>配偶</b>	普莉希拉·陈 (2012年结婚)		
<b>Children</b>	1	<b>儿女</b>	麦丝玛·陈·扎克伯格		
<b>Relatives</b>	Randi Zuckerberg (sister)	<b>网站</b>	facebook.com/zuck		
<b>Website</b>	facebook.com/zuck				

# Motivation

- In Wikipedia, each language version maintains its own set of infoboxes with their own set of attributes.
- English Wikipedia is obviously larger of higher quality than low resource language.
- The number of existing attributes mappings is limited.

# Challenges

- Short text
  
- Polysemy & Synonym
  - a given attribute can have different semantics
  - different attributes may have the same meaning
  
- Different measurement & Timeliness
  - population of Beijing is 21,700,000 in English and 2170 ten thousand in Chinese.
  - population of Beijing is 21,150,000 (in 2013) in French edition.

# Existing methods

Methods	Feature	Data
Bouma et al. 2009	Value based	English-Dutch Wiki
Rinser et al. 2013	Instance & graph based	Several Wikis
Adar et al. 2009	Learning-based	Several Wikis

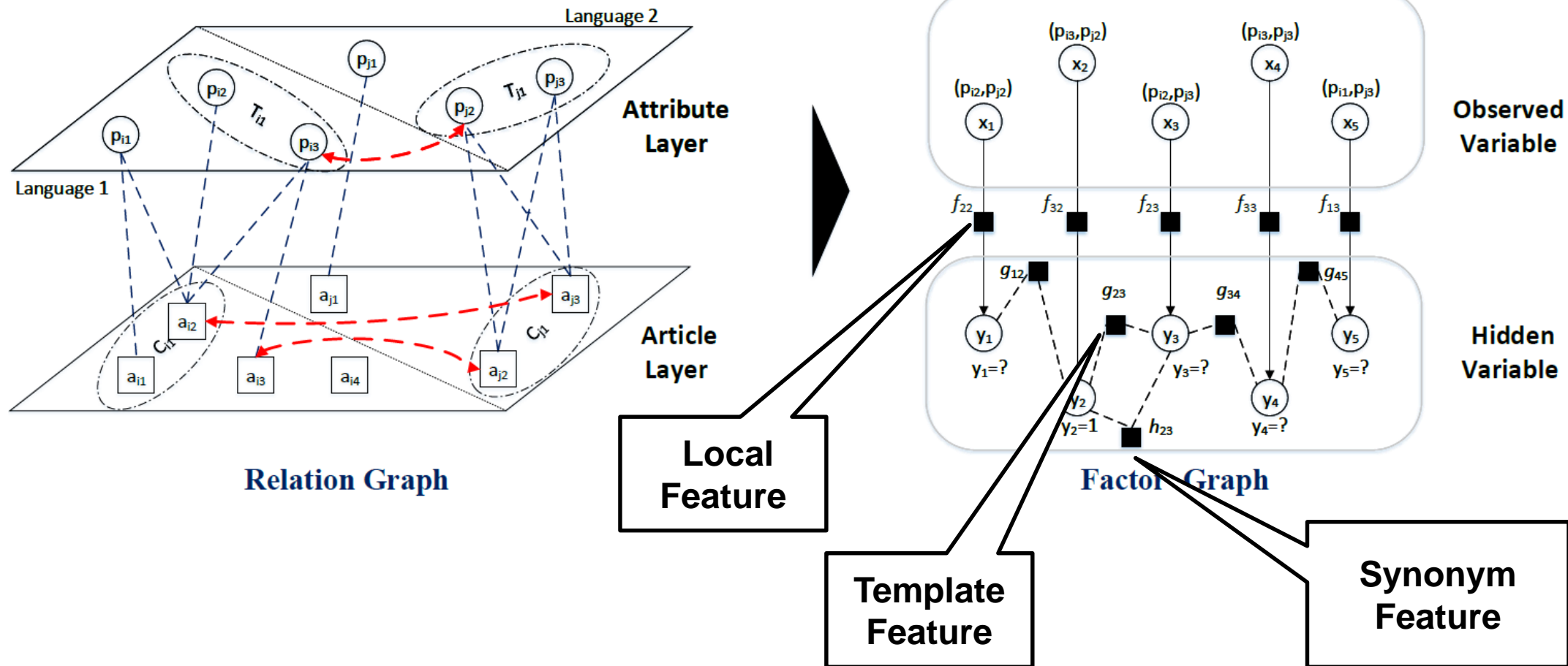


# Outline

- Background: Wikipedia infobox alignment
- **Our Approach: Entity-Attribute Factor Graph**
- Experiments & Conclusion
- Future Work



# Overview



# Model Features

## ■ Local Features

### □ Label similarity

- Levenshtein Distance
- Word embedding distance

### □ Subject-object similarity

### □ Article-usage feature

### □ Category similarity feature

## ■ Template Feature

- use the template attribute co-occurrence relations

## ■ Synonym Feature

- use the semantic relatedness of attributes

# Learning and Inference

## ■ Joint Distribution

$$p(Y) = \frac{1}{Z} \prod_i \exp\{\theta^T(\mathbf{f}(y_i, y_j), \sum_{y_j} \mathbf{g}(y_i, y_j), \sum_{y_j} \mathbf{h}(y_i, y_j))\}$$

## ■ Object Function

$$\log(p(Y^L))$$



# Outline

- Background: Wikipedia infobox alignment
- Our Approach: Entity-Attribute Factor Graph
- Experiments & Conclusion
- Future Work

# Experiments

## ■ Data Set

Data set	#Attribute Pairs	#Related Articles	#Related Categories
EN-ZH	2000	EN:96,331 ZH:54,195	EN:13,763 ZH:9,132
EN-FR	2000	EN:103,915 FR:89,012	EN:15,698 FR:12,371

## ■ Result of Attribute Matching

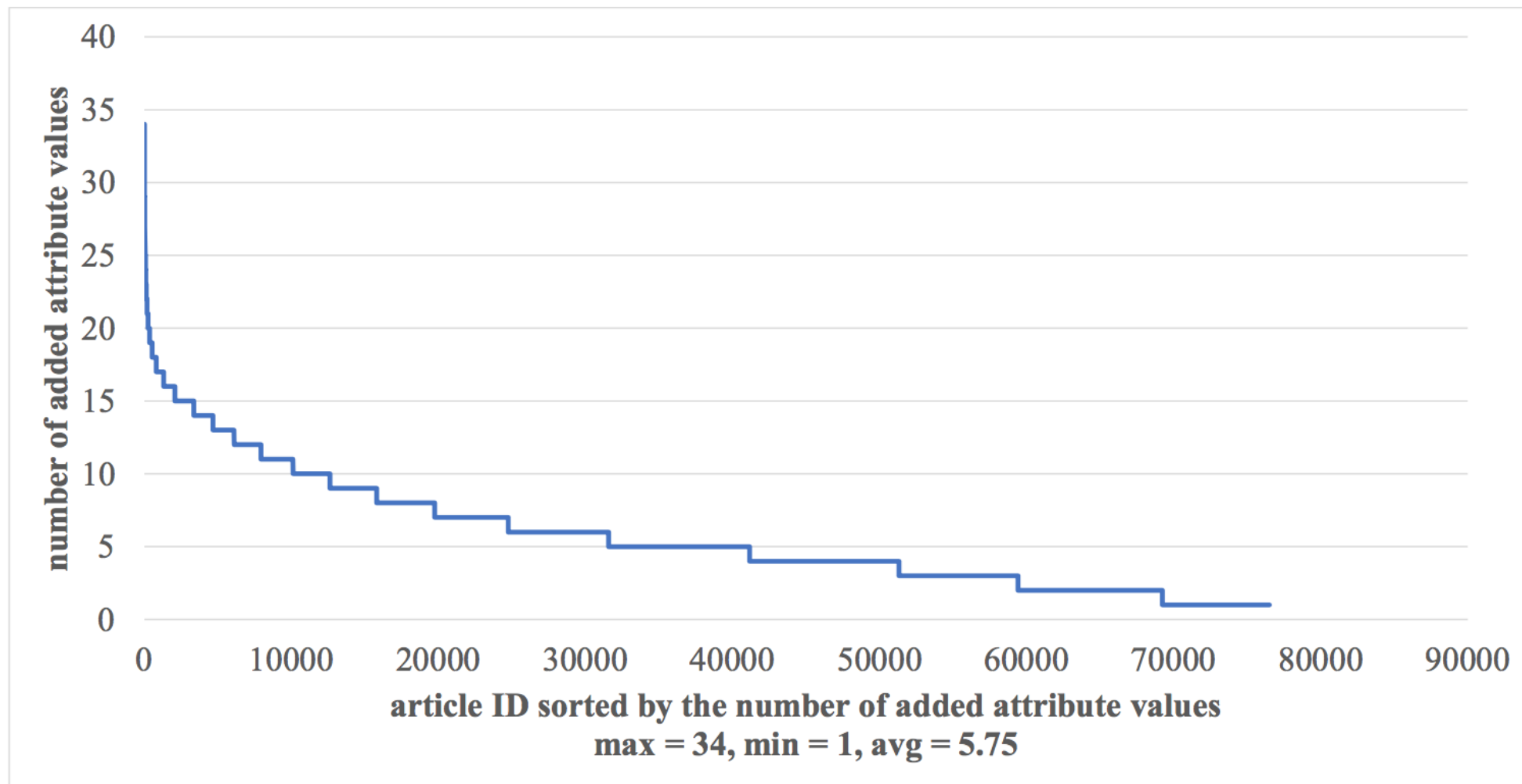
Method	English-Chinese			English-French		
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure
LM	<b>0.973</b>	0.261	0.412	<b>0.982</b>	0.271	0.425
SA	0.749	0.673	0.709	0.764	0.662	0.709
SVM	0.875	0.752	0.809	0.883	0.755	0.814
LR-ADAR	0.907	0.746	0.819	0.917	0.739	0.818
EAFG(NT)	0.863	0.771	0.814	0.877	0.774	0.822
EAFG	0.911	<b>0.805</b>	<b>0.855</b>	0.913	<b>0.802</b>	<b>0.854</b>

# Experiments

## ■ Example

Classes	English	Chinese	French
Person	Alma mater Spouse(s) Title Nationality	母校 配偶 头衔 国籍	Formation Conjoint Activité principale Nationalité
Location	Party Secretary Completed Population Mayor	书记 竣工年份 人口 市长	Secrétaire du PCC Fin des travaux Population Maire
Film	Directed by Screenplay by Running time Country	导演 编剧 片长 产地	Réalisation Scénario Durée Pays d'origine

# Wikipedia Infobox Complementing





# Outline

- Background: Wikipedia infobox alignment
- Our Approach: Entity-Attribute Factor Graph
- Experiments & Conclusion
- **Future Work**



# Future Work

- Attribute Alignment in Xlore (<http://xlore.org>)
- Design a model to match entities & attributes iteratively



**Thank you!**

**For more questions,  
please feel free to contact us.**