

# Knowledge Base Unification via Sense Embeddings and Disambiguation

EMNLP2015

Claudio Delli Bovi @uniroma1 罗马大学

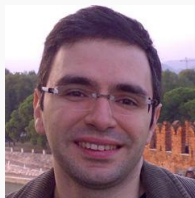
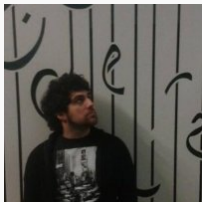
Luis Espinosa-Anke @upf 庞培法布拉大学

Roberto Navigli @uniroma1

hanzhe@icst-wip

2015 年 10 月 30 日

- summary
  - 作者简介
  - 相关工作
  - 论文动机、方法概述
  - 实验使用的工具 (穿插说)
- 归一化 (Unification) 的方法
  - 实体消歧 (entity disambiguation)
  - 关系对应 (relation alignment)
- 实验/总结



### Claudio Delli Bovi

罗马大学博士，计算语言学方向，主要研究语法、句法结构，目前在做 WSD

### Luis Espinosa-Anke

庞培法布拉大学博士，在做一些信息抽取，语义网有关的东西

### Roberto Navigli

Claudio 的导师，07 年博士毕业，BabelNet/Babelfy

## summary/author

- Winner of the prestigious *META prize 2015* for BabelNet, Riga Summit 2015.



### Claudio Delli Bovi

罗马大学博士，计算语言学方向，主要研究语法、句法结构，目前在做 WSD

### Luis Espinosa-Anke

庞培法布拉大学博士，在做一些信息抽取，语义网有关的东西

### Roberto Navigli

Claudio 的导师，07 年博士毕业，BabelNet/Babelfy

# 概述

— 动机/方法简述

## summary/motivation

实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.baidubaik:

李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网  
球)-夫婿-姜山 (湖北人)

### 如何整合多个知识库?

假设

- 全部在 subject-predicate-object 数据库上
- 知识库可能是完全非结构化的 (主语只是普通字符串)
- 给定一个完备的语义集合 (左上的实体集合)
- 其他的数据库可能存在歧义, 结构不标准

# summary/motivation

实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.baidubaike:

李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网  
球)-夫婿-姜山 (湖北人)

## 如何整合多个知识库?

过程

- 怎么把李娜-丈夫-姜山转换成标准的结构?
- 消歧 + 谓词统一
  - 消除其他数据库中主体、客体的歧义
  - 将不同数据库中含义相同的谓词合并

# summary/method

实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.baidubaike:

李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网  
球)-夫婿-姜山 (湖北人)

## 如何整合多个知识库?

方法

- 先对**实体消歧**，再对应不同数据库的**relation 归一**
- **实体消歧【1】**
  - 对一条三元组的主体、客体的语义候选的所有组合，如果存在一组**已经非常好了**，选为**种子**
  - 找出好的 relation（主体的向量集中，比如都表示人；客体的向量集中，比如都表示地点）



# summary/method

## 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

## 【一】

微软-CEO-纳德拉

搜狐-CEO-张朝阳

苹果 (?) -CEO-库克

...

zh.baidubaikē:

李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网  
球)-夫婿-姜山 (湖北人)

## 如何整合多个知识库?

方法

### ● 实体消歧【2】

- 【一】对于好的 relation, 该关系内的所有三元组的主体、客体相互间相似, 可以作为其中一个主体做消歧时的文本

# summary/method

## 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

## 【二】

花果山-所在-连云港

全国政协-所在-北京

大众-所在-黑龙江

(这里指大众乡) ...

zh.baidubaikē:  
李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网  
球)-夫婿-姜山 (湖北人)

## 如何整合多个知识库?

方法

### ● 实体消歧【2】

- 【一】对于好的 relation, 该关系内的所有三元组的主体、客体相互间相似, 可以作为其中一个主体做消歧时的文本
- 【二】对于不好的 relation 对应的一条三元组的主语或客体, 只能通过 relation 名字来消歧

## summary/method

实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.baidubaikē:

李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网

球)-夫婿-姜山 (湖北人)

## 如何整合多个知识库?

方法

### ● 实体消歧【2】

- 【一】对于好的 relation，该关系内的所有三元组的主体、客体相互间相似，可以作为其中一个主体做消歧时的文本
- 【二】对于不好的 relation 对应的一条三元组的主语或客体，只能通过 relation 名字来消歧

## summary/method

实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.baidubaik:

李娜-丈夫-姜山



李娜 (网球运动员)-配偶-姜山



zh.hudong: 李娜 (网  
球)-夫婿-姜山 (湖北人)

### 如何整合多个知识库?

方法

- 不同知识库的 **relation 归一**
- 通过每两个知识库的每任两个 relation 的 **主客体特征向量** 计算相似性

## 相关工作

- Open Information Extraction


- 从 Web-scale 级别的自然语言信息中抽取结构化/格式化的数据
  - e.g. DBpedia, Freebase, YAGO,...
- 提升效果/去除噪声数据的方法
  - matrix factorization, distant supervision, multi-instance,...
- 知识库补全 (Knowledge Base completion)
  - 少量结构化数据和大量的半结构化数据相互提升准确率

## ● BabelNet

- 罗马大学开发的跨语言语义网，以“姚明”为例
  - “entity”都是实体，有中文标签；“relation”纯英文
  - “relation”有重复，无意义的

### 姚明

中国篮球运动员

IS A	人 •  basketballer
BIRTH PLACE	中华人民共和国 • 上海市
CAREER POSITION	中锋
COUNTRY OF CITIZENSHIP	中华人民共和国
DRAFT TEAM	休斯敦火箭
HIGHLIGHTS	NBA最佳新秀阵容
MEMBER OF SPORTS TEAM	休斯敦火箭 • 上海东方大鲨鱼篮球俱乐部
OCCUPATION	篮球运动员
PLACE OF BIRTH	上海市
⊖ Less relations	
POSITION PLAYED ON TEAM	中锋
SEX OR GENDER	男性
SPOUSE	叶莉
STAT LABEL	得分 • 盖帽 • 篮板球
SURNAME	姚
TEAM	休斯敦火箭 • 中国男子篮球职业联赛 • 上海东方大鲨鱼篮球俱乐部

EXPLORE NETWORK

### ● Babelfy

- 罗马大学开发的多语言的消歧和实体链接集成的工具
  - 链接到 BabelNet
  - 分词效果不好;
  - 有一定的消歧能力：能找出网球运动员李娜，找不到歌手李娜

The screenshot shows the Babelfy web interface. At the top left is the Babelfy logo. To its right is a search bar containing the text '李娜 (1982年2月26日 - )，出生于中华人民共和国湖北省武汉市，两届网球大满贯单打得主，前职业女子网球选手，1999年转为职业球员'.

Below the search bar, there are two buttons: 'Enable partial matches: ☐' and 'CHINESE'. To the right of these is a green button labeled 'BABELFY!'.

Below the buttons, there are two links: 'expanded view' and 'compact view'.

The main part of the interface displays a search result for '李娜' (Li Na). The result is shown as a sequence of entities and relations: '李娜' (Li Na) (entity), '（' (left parenthesis), '1982年' (1982) (year), '2月26日' (February 26) (date), '）' (right parenthesis), '，' (comma), '出生' (birth) (relation), '于' (at) (preposition), '中华' (China) (entity). Below this sequence are four circular images: a tennis player (Li Na), a person in a colorful patterned dress, a man in a suit (likely a historical figure), and a person lying down.



### 两个数据库间的归一/对应

- Dutta et al. (2014): NELL 数据集的 augment 链接到 DBpedia 的实体
  - 使用一阶逻辑和马尔可夫网络
- Grycner and Weikum (2014) :PATTY 的 pattern 和 wordNet 的谓词对应
- Lin et al. (2012) : 使用 freebase 的类别来链接 reverb 中的潜在实体
  - reverb record:[2047542 **Bilberry** also\_contains **vitamin\_C** bilberry also\_contain vitamin\_c 1 0.94124 [http:...](#)]

能适应多个数据库归一的方法

- Riedel et al. (2013): 训练含有隐式特征 (latent feature) 的实体向量、关系向量
- Dong et al. (2014): 使用 freebase 数据训练一个概率模型来判断抽取结果的准确度

在关系抽取和知识补全上使用 Embedding models

- Socher et al., 2013; Weston et al., 2013; Bordes et al., 2013
- 作者认为这些模型停留在 (surface level), 不能表达普世的语义 (common semantic)

能适应多个数据库归一的方法

- Riedel et al. (2013): 训练含有隐式特征 (latent feature) 的实体向量、关系向量
- Dong et al. (2014): 使用 freebase 数据训练一个概率模型来判断抽取结果的准确度

在关系抽取和知识补全上使用 Embedding models

- Socher et al., 2013; Weston et al., 2013; Bordes et al., 2013
- 作者认为这些模型停留在 (surface level), 不能表达普世的语义 (common semantic) 【不能区分多义词?】

# 归一化方法

## — 实体消歧

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

## 【实体消歧】准备

- BabelNet 是一个语义集合 (sense inventory), 与常见知识库的实体对应
- 消歧结束
  - 知识库的主语/客体对应到 BabelNet 上
- 实体/语义向量
  - 利用一个庞大的标注语料训练词向量
  - 采用 SENSEMBED (罗马大学 2015ACL) 训练而非 word2vec, 因为后者不能区分多义词

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

## 【实体消歧】准备

- BabelNet 是一个语义集合 (sense inventory), 与常见知识库的实体对应
- 消歧结束
  - 知识库的主语/客体对应到 BabelNet 上
- 实体/语义向量
  - 利用一个庞大的标注语料训练词向量
  - 采用 SENSEMBED (罗马大学 2015ACL) 训练而非 word2vec, 因为后者不能区分多义词

按照论文的图, 作者是拿维基百科训练的。

而维基百科内部链接做的很好, 很多地方被人工消歧了。

可以训练出多个词向量 {苹果、苹果公司}?

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

## 【实体消歧】准备

- BabelNet 是一个语义集合 (sense inventory), 与常见知识库的实体对应
- 消歧结束
  - 知识库的主语/客体对应到 BabelNet 上
- 实体/语义向量
  - 利用一个庞大的标注语料训练词向量
  - 采用 SENSEMBED (罗马大学 2015ACL) 训练而非 word2vec, 因为后者不能区分多义词

## iPhone [编辑]

(重定向自 iPhone)



本文介绍的是苹果公司推出的iPhone系列。关于iPhone的其他意思，详见“[iPhone \(消歧义\)](#)”

**iPhone**，是苹果公司旗下研发的**移动电话**系列，它搭载苹果公司研发的**iOS**(以前是**OS**)手机操作系统。当时苹果公司CEO的**史蒂夫·乔布斯**发布<sup>[14]</sup>，并在同年6月29日正式发售<sup>[15][16]</sup>。2015年9月9日，推出第九代产品**iPhone 6S**，于同年9月25日正式发售。苹果公司旗下智能手机iPhone，

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

## 【实体消歧】准备

- BabelNet 是一个语义集合 (sense inventory), 与常见知识库的实体对应
- 消歧结束
  - 知识库的主语/客体对应到 BabelNet 上
- 实体/语义向量
  - 利用一个庞大的标注语料训练词向量
  - 采用 SENSEMBED (罗马大学 2015ACL) 训练而非 word2vec, 因为后者不能区分多义词



## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山

李娜 (网球运动员)-丈夫-姜山

## 【实体消歧一】找出种子三元组

- 三元组  $\langle e_d, r, e_g \rangle$ ,  $e_d, e_g$  的所有语义  $\mathbf{s}_d = \{s_d^1, \dots, s_d^m\}$  and  $\mathbf{s}_g = \{s_g^1, \dots, s_g^{m'}\}$ , 对应的词向量  $\mathbf{v}_d = \{v_d^1, \dots, v_d^m\}$  and  $\mathbf{v}_g = \{v_g^1, \dots, v_g^{m'}\}$
- 任意两个组合中词向量余弦相似度最大的一对作为该三元组消歧的默认最优解
$$\langle v_d^*, v_g^* \rangle = \operatorname{argmax}_{v_d \in \mathbf{v}_d, v_g \in \mathbf{v}_g} \frac{v_d \cdot v_g}{\|v_d\| \|v_g\|}$$
- 如果最优解的相似度大于一个阈值  $\zeta$ , 那么这个最有解是“完美”的, 该三元组即为种子三元组

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

## 【实体消歧一】找出种子三元组

- 三元组  $\langle e_d, r, e_g \rangle, e_d, e_g$  的所有语义  $s_d = \{s_d^1, \dots, s_d^m\}$  and  $s_g = \{s_g^1, \dots, s_g^{m'}\}$ , 对应的词向量  $\mathbf{v}_d = \{v_d^1, \dots, v_d^m\}$  and  $\mathbf{v}_g = \{v_g^1, \dots, v_g^{m'}\}$
- 任意两个组合中词向量余弦相似度最大的一对作为该三元组消歧的默认最优解
$$\langle v_d^*, v_g^* \rangle = \operatorname{argmax}_{v_d \in \mathbf{v}_d, v_g \in \mathbf{v}_g} \frac{v_d \cdot v_g}{\|v_d\| \|v_g\|}$$
- 如果最优解的相似度大于一个阈值  $\zeta$ , 那么这个最有解是“完美”的, 该三元组即为种子三元组
- 关系  $r$  的三元组的主语、客体的词向量可能是满足一个特定的“角度”才是标注的?

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山

李娜 (网球运动员)-丈夫-姜山

## 【实体消歧一】找出种子三元组

- 三元组  $\langle e_d, r, e_g \rangle$ ,  $e_d, e_g$  的所有语义  $\mathbf{s}_d = \{s_d^1, \dots, s_d^m\}$  and  $\mathbf{s}_g = \{s_g^1, \dots, s_g^{m'}\}$ , 对应的词向量  $\mathbf{v}_d = \{v_d^1, \dots, v_d^m\}$  and  $\mathbf{v}_g = \{v_g^1, \dots, v_g^{m'}\}$
- 任意两个组合中词向量余弦相似度最大的一对作为该三元组消歧的默认最优解
$$\langle v_d^*, v_g^* \rangle = \operatorname{argmax}_{v_d \in \mathbf{v}_d, v_g \in \mathbf{v}_g} \frac{v_d \cdot v_g}{\|v_d\| \|v_g\|}$$
- 如果最优解的相似度大于一个阈值  $\zeta$ , 那么这个最有解是“完美”的, 该三元组即为种子三元组

## 【实体消歧二】找出完美关系

BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

### def. 完美关系

该 relation 对应的主语 (客体) 的词向量很集中, 方差小

- $v_D, v_G$  表示关系  $r$  对应的所有种子三元组的主语、客体的词向量集合

- 该关系的主语、客体中心词向量为

$$\mu_k = \frac{1}{|\mathbf{v}_k|} \sum_{v \in \mathbf{v}_k} \frac{v}{\|v\|}, \quad k \in \{D, G\}$$

- 该关系的主语、客体方差为

$$\sigma_k^2 = \frac{1}{|\mathbf{v}_k|} \sum_{v \in \mathbf{v}_k} (1 - \cos(v, \mu_k))^2$$

- 该关系的总体方差为主客体方差的平均数
- 方差越小, 说明该关系对应的种子三元组越好
- 如果总体方差小于某个阈值  $\delta$ , 则认为其是完美关系

## BabelNet 实体

李娜 (网球运动员)

李娜 (歌手)

李娜 (游泳运动员)

姜山 (网球运动员)

张家辉

...

zh.wikipedia: 李娜-丈夫-姜山



李娜 (网球运动员)-丈夫-姜山

【实体消歧三】实体消歧（分类）对于任意一个三元组  $\langle e_d, r, e_g \rangle$ ，比如想要判断  $e_d$  的含义

- 如果  $r$  是好的关系

那么  $r$  内的种子三元组能正确的反应  $r$  的含义，使用种子三元组的主语可以作为  $e_d$  消歧的依据

- 如果  $r$  不是好的关系

只用当前这条三元组的  $r, e_g$  来作为  $e_d$  分类/消歧的依据

# 归一化方法

— 关系 (relation) 归一

和 disambiguation 过程类似

- 1 对于每个知识库的每种关系，计算其主体、客体的平均向量
- 2 对于任意两个知识库的任意关系对  $r_i, r_j$ ，其相似性为  $(s_D + s_G)/2$ 
  - $s_k$  定义如下， $k \in \{D, G\}$ ，其中  $\mu_D^{r_j}$  表示关系  $r_j$  的主语向量的平均值

$$s_k = \frac{\mu_k^{r_i} \cdot \mu_k^{r_j}}{\|\mu_k^{r_i}\| \|\mu_k^{r_j}\|}$$

# 实验



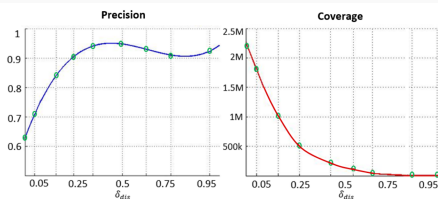
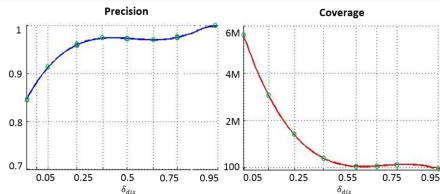
# experiment

【数据集】4 个，后两个内部有链接（linked resources），前两个没有

	$K_U$		$K_D$	
	NELL	REVERB	PATTY	WiSeNet
# relations	298	1 299 844	1 631 531	245 935
# triples	2 245 050	14 728 268	15 802 946	2 271 807
# entities	1 996 021	3 327 425	1 087 907	1 636 307

【额外实验】测试种子三元组的正确率

- 实验的时候认为种子三元组是完全正确的，实际有一点错误
- 测试在 patty 上：左边是在全部种子三元组的情况，右边是去掉本来就无歧义的种子三元组



## 【主要实验一】测试消歧准确率

- 上图的 Baseline 是指不选 seed
  - 任意一个词消歧时只把他所在的那条三元组的主语、客体、谓词作为上下文消歧，和该关系的其他三元组无关
  - $\zeta$  越大表示种子三元组越准确，准确率会更高

$\zeta_{dis}$	SENSEMBED			Baseline		
	0.5-0.7	0.7-0.9	0.9-1.0	0.5-0.7	0.7-0.9	0.9-1.0
<b>PATTY</b>	.980	.980	1.000	.793	.780	1.000
<b>WiSENET</b>	.958	.960	.973	.726	.786	.791
<b>NELL</b>	.955	.995	1.000	.800	.770	.885
<b>REVERB</b>	.930	.940	.950	.775	.725	.920

Table 2: Disambiguation precision for all KBs

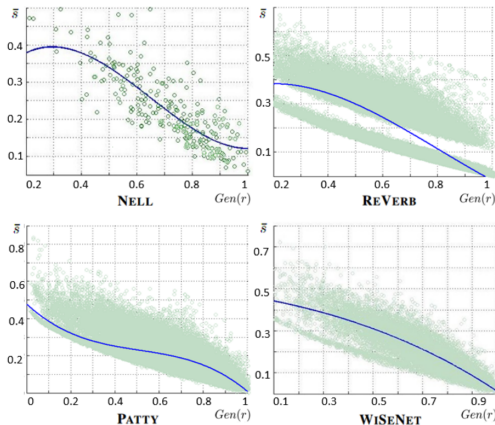
## 【主要实验一】测试消歧准确率

- 下图的“only seed”是指只采用每个类别的种子三元组来表示给类别的主语、客体，进而预测三元组的主语、客体的类别
  - 不论是否“完美”，都按“完美”考虑
- (?)PATTY 和 WISENET 本身对应到维基百科实体，理论不用消歧，准确率是 100%：可能是因为：
  - 一些维基百科本身对应错误
  - 作者的方法其实也可以使用在已链接的主客体上

	$\delta_{spec} = 0.8$		$\delta_{spec} = 0.5$		$\delta_{spec} = 0.3$	
	all	only seeds	all	only seeds	all	only seeds
<b>PATTY</b>	62.15	26.60	52.49	24.06	40.75	21.41
<b>WISENET</b>	60.00	37.46	54.44	22.26	53.58	16.62
<b>NELL</b>	76.97	62.98	50.95	20.71	44.70	4.36
<b>REVERB</b>	41.20	38.57	25.14	23.70	13.37	12.75

Table 3: Coverage results (%) for all KBs

## 【额外实验二】不同 relation 对主客体的限制



- $Gen(r) = (\sigma_D + \sigma_G)/2$   
(主客体的方差) 的均值
- $\bar{s} = avg_{v_r=r}(v_d * v_g)$   
(三元组的主客体相似度) 的均值
- $Gen(r)$  越大说明 relation 要求越严格
- $\bar{s}$  越大说明主客体的类别要求越低  
(都是随机出现, 则  $\bar{s} \approx \sqrt{2}/2 \approx 0.7?$ )
- 越普适的 relation, 对主客体的要求越低

## 【额外实验二】不同 relation 对主客体的限制

NELL	
High $Gen(r)$	agent created at location
Low $Gen(r)$	person in economic sector restaurant in city
REVERB	
High $Gen(r)$	is for is in
Low $Gen(r)$	enter Taurus in carry oxygen to
PATTY	
High $Gen(r)$	located in later served to
Low $Gen(r)$	starting pitcher who played league coach for
WiSENET	
High $Gen(r)$	include is a type of
Low $Gen(r)$	lobe-finned fish lived during took part in the Eurovision contest

● 例子

## 【主要实验二】测试 relation 对应的准确率

- 为了方便测试，每个数据集只取前 1w 个关系，相互对应
  - 随机选取 150 对，人工判定

	PATTY-WiSENET		PATTY-REVERB		NELL-REVERB	
$\delta_{align}$	0.7	0.9	0.7	0.9	0.7	0.9
<b>Prec.</b>	.68	<b>.80</b>	.58	<b>.74</b>	.61	<b>.75</b>
<b># Align.</b>	128k	1.2k	47k	643	2.6k	88

	PATTY-NELL		WiSENET-NELL		WiSENET-REVERB	
$\delta_{align}$	0.7	0.9	0.7	0.9	0.7	0.9
<b>Prec.</b>	.66	<b>1.00</b>	.70	<b>.84</b>	.59	<b>.87</b>
<b># Align.</b>	2.6k	57	381	34	9.9k	169

## 【主要实验二】测试 relation 对应的准确率

### ● 对齐的样例

PATTY-WISENET		$\zeta_{align}$
portrayed	's character	0.84
debuted in	first appeared in	0.86
PATTY-REVERB		$\zeta_{align}$
language in	is spoken in	0.81
mostly known for	plays the role of	0.70
NELL-REVERB		$\zeta_{align}$
bookwriter	is a novel by	0.88
personleadscity	is the mayor of	0.60
NELL-PATTY		$\zeta_{align}$
worksfor	was hired by	0.72
riveremptiesintoriver	tributary of	0.89
NELL-WISENET		$\zeta_{align}$
animaleatfood	feeds on	0.72
teamhomestadium	play their home games at	0.88
REVERB-WISENET		$\zeta_{align}$
has a selection of	offers	0.82
had grown up in	was born and raised in	0.85

## 总结



## ● 出发点

- 同一个关系的不同实例应该有相似的词向量，主语互相比拟相像，客体也是
- 不同知识库如果使用同一套主语、客体 entity，那么不同知识库之间的相同关系应该有相似的主语、客体形式

## ● 实验效果

- 对种子三元组和完美关系这两处筛选/排序使得好的数据起到更大的作用
- (缺点) 没有统一的评价体系，相互间的优劣不能对比：“我们和别的系统不同；不能照搬；无法比较效果”，很多手工筛选检查

问题?