

Improving Entity Linking by Modeling Latent Relations between Mentions

Презентацию подготовил

Чернявский Антон, 151

Named entity linking (named entity disambiguation)

Paris is the capital of France

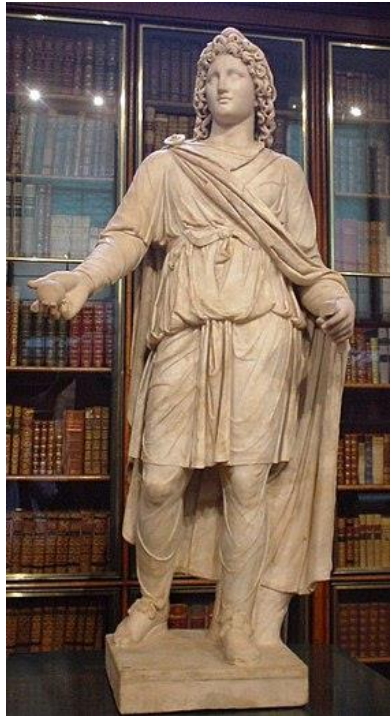
Name entity recognition – пытаемся определить к какому классу (личность, организация, географическая локация и т.д.) относится именованная сущность.

Named entity linking (named entity disambiguation)

Paris is the capital of France



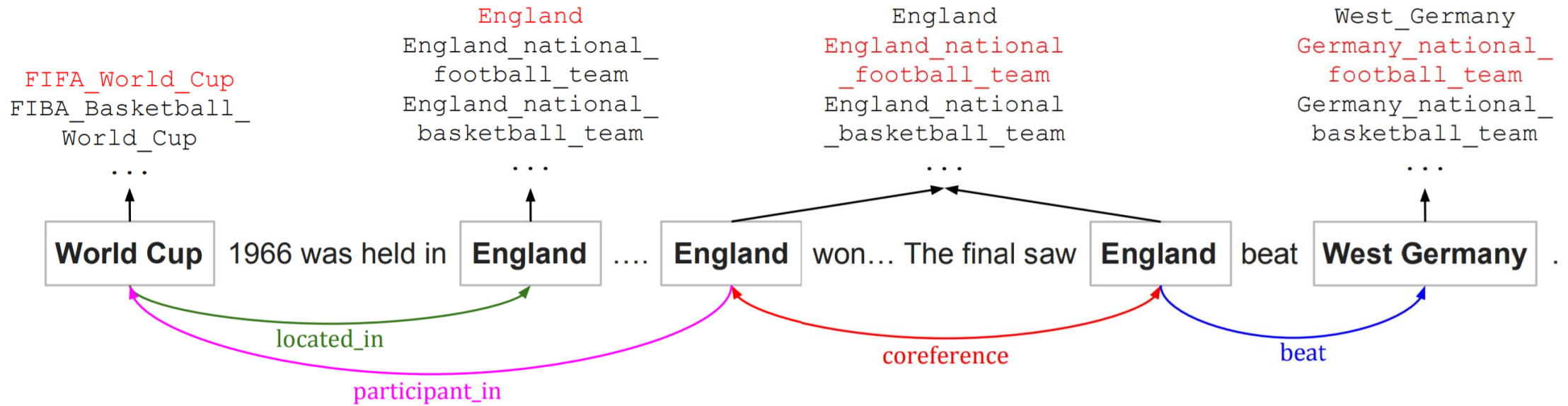
Чему соответствует в базе знаний



Отношения между упоминаниями (mentions)

- Наиболее популярное – coreference. Два упоминания являются кореферентными, если они ссылаются на одну и ту же сущность
- Определение других и их использование – сложная задача
- Для нахождения часто используются созданные вручную правила
- Однако в некоторых языках нет экспертов, способных определить такие правила

Использование отношений между упоминаниями



Применение Named Entity Linking

- Важный аспект понимания естественного языка
- Улучшает качество Information Retrieval систем
- Предварительная предобработка для Question Answering, Information Extraction

Формальная постановка задачи

- Документ D содержит mentions m_1, m_2, \dots, m_n
- Им требуется сопоставить $e_1, e_2, \dots, e_n \in KB$ (базы знаний)
- В случае отсутствия искомой сущности в базе знаний $e_k = NULL$

Предварительная эвристика *candidate selection* оставляет для каждого упоминания набор $C_i = (e_{1i}, \dots, e_{il_i})$

Локальные методы

Игнорируют связи (*coherence*) между решениями, опираясь только на локальный контекст

- c_i – локальный контекст m_i
- $\Psi(e_i, c_i)$ – функция локальной оценки
- $\forall i \in \{1, \dots, n\} \quad e_i^* = \operatorname{argmax}_{e_i \in C_i} \Psi(e_i, c_i)$

Глобальные методы

Помимо локальной информации использует *coherency*, измеряющуюся функцией $\Phi(E, D)$, где $E = (e_1, \dots, e_n)$

$$\begin{aligned} E^* &= \operatorname{argmax}_{E \in C_1 \times \dots \times C_n} \sum_{i=1}^n \Psi(e_i, c_i) + \Phi(E, D) = \\ &= \operatorname{argmax}_{E \in C_1 \times \dots \times C_n} \sum_{i=1}^n \Psi(e_i, c_i) + \sum_{i \neq j} \Phi(e_i, e_j, D) \end{aligned}$$

Недостаток: NP-сложная задача (можно *приблизленно* решать декомпозицией по e_i)

Representation learning (возможные подходы)

1. Косинусное сходство между названиями и контекстами
2. Использование эмбедингов

$$\Psi(e_i, c_i) = \mathbf{e}_i^T \mathbf{B} f(c_i)$$

$$\Phi(e_i, e_j, D) = \frac{1}{n-1} \mathbf{e}_i^T \mathbf{R} \mathbf{e}_j$$

$\mathbf{e}_i \in \mathbb{R}^d$ — эмбединг e_i

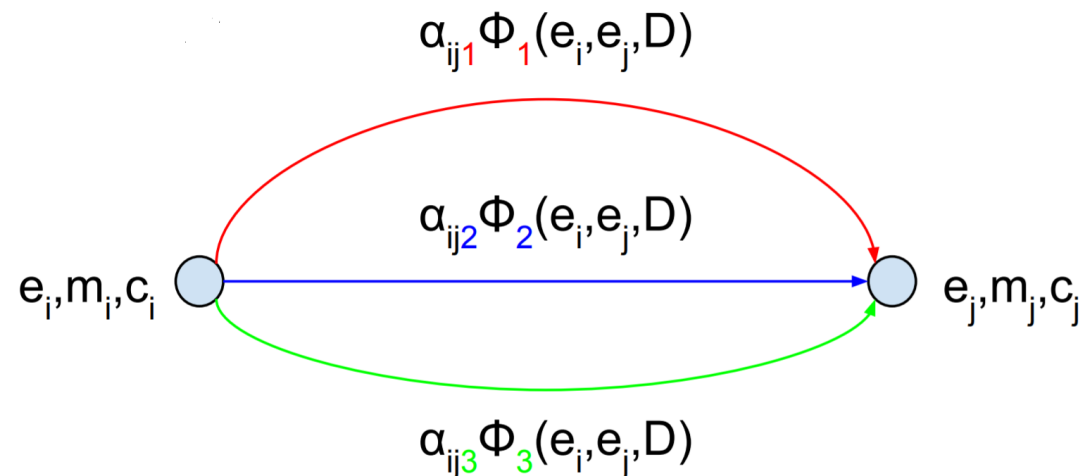
$\mathbf{B}, \mathbf{R} \in \mathbb{R}^{d \times d}$ — диагональные матрицы

$f(c_i)$ — attention к контексту

Multi-relational models (общий вид)

- Пусть существует K различных отношений
- Каждое присваиваем паре (m_i, m_j) с коэффициентом $\alpha_{ijk} \geq 0$

$$\Phi(e_i, e_j, D) = \sum_{k=1}^K \alpha_{ijk} \Phi_k(e_i, e_j, D)$$



Multi-relational models (общий вид)

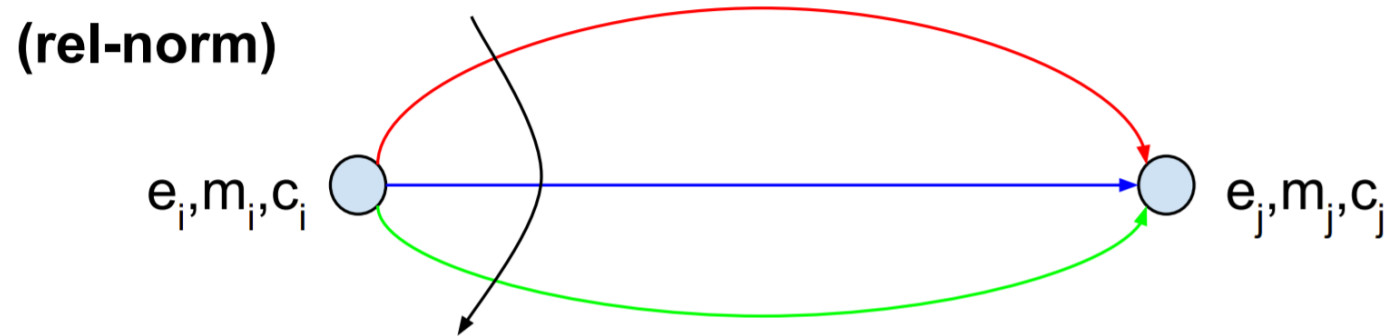
- Пусть существует K различных отношений
- Каждое присваиваем паре (m_i, m_j) с коэффициентом $\alpha_{ijk} \geq 0$

$$\Phi(e_i, e_j, D) = \sum_{k=1}^K \alpha_{ijk} \Phi_k(e_i, e_j, D)$$

$$\Phi_k(e_i, e_j, D) = \mathbf{e}_i^T \mathbf{R}_k \mathbf{e}_j$$

$$\alpha_{ijk} = \frac{1}{Z_{ijk}} \exp \left\{ \frac{f^T(m_i, c_i) D_k f(m_j, c_j)}{\sqrt{d}} \right\}$$

Коэффициент нормализации



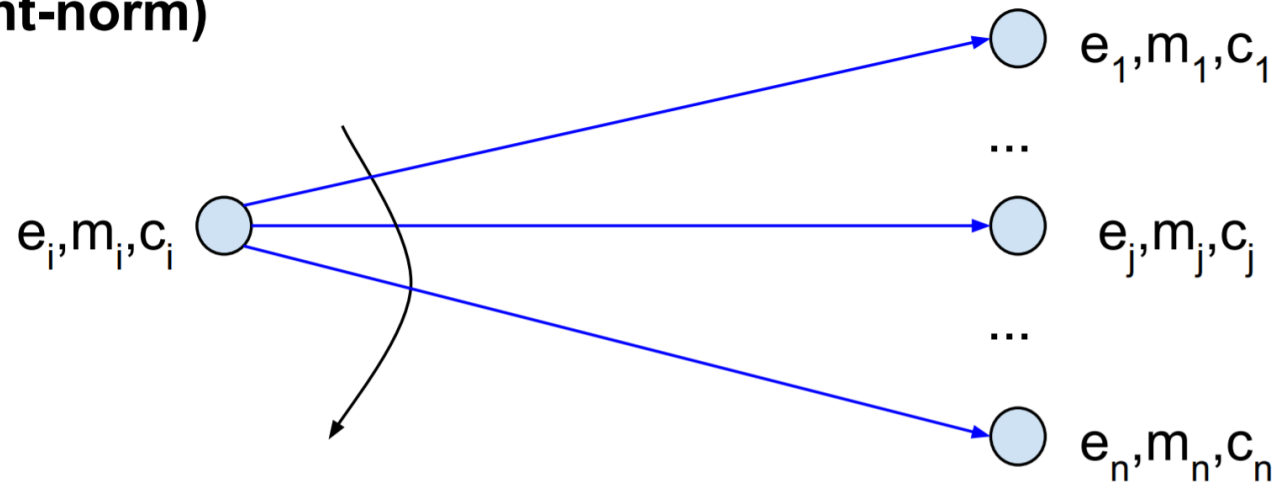
normalize over relations: $\alpha_{ij1} + \alpha_{ij2} + \alpha_{ij3} = 1$

Интуиция: α_{ijk} – вероятность k -го отношения для пары (m_i, m_j)

$$\Phi(e_i, e_j, D) = \sum_{k=1}^K \alpha_{ijk} \mathbf{e}_i^T \mathbf{R}_k \mathbf{e}_j = \mathbf{e}_i^T \mathbf{R}_{ij} \mathbf{e}_j, \text{ где } \mathbf{R}_{ij} = \sum_{k=1}^K \alpha_{ijk} \mathbf{R}_k$$

Коэффициент нормализации

(ment-norm)



normalize over mentions:

$$\alpha_{i12} + \alpha_{i22} + \dots + \alpha_{ij2} + \dots + \alpha_{in2} = 1$$

Интуиция: для каждого m_i ищем m_j связанное отношением k (схоже с multi-head attention)

Ment-norm vs. rel-norm

Ment-norm:

- 1) Если α_{ijk} маленький для всех k , значит упоминания не связаны
- 2) Если α_{ijk} велик для нескольких k , значит упоминания связаны несколькими отношениями

Rel-norm:

- 1) none-relation – отсутствие связи
- 2) Регулируем относительные вероятности за счет none-relation

Ment-norm vs. rel-norm

Главный недостаток ment-norm:

Модель использует все К отношений даже в тех случаях, когда некоторые отношения неприменимы.



“West Germany” ни с чем не связана кореференцией

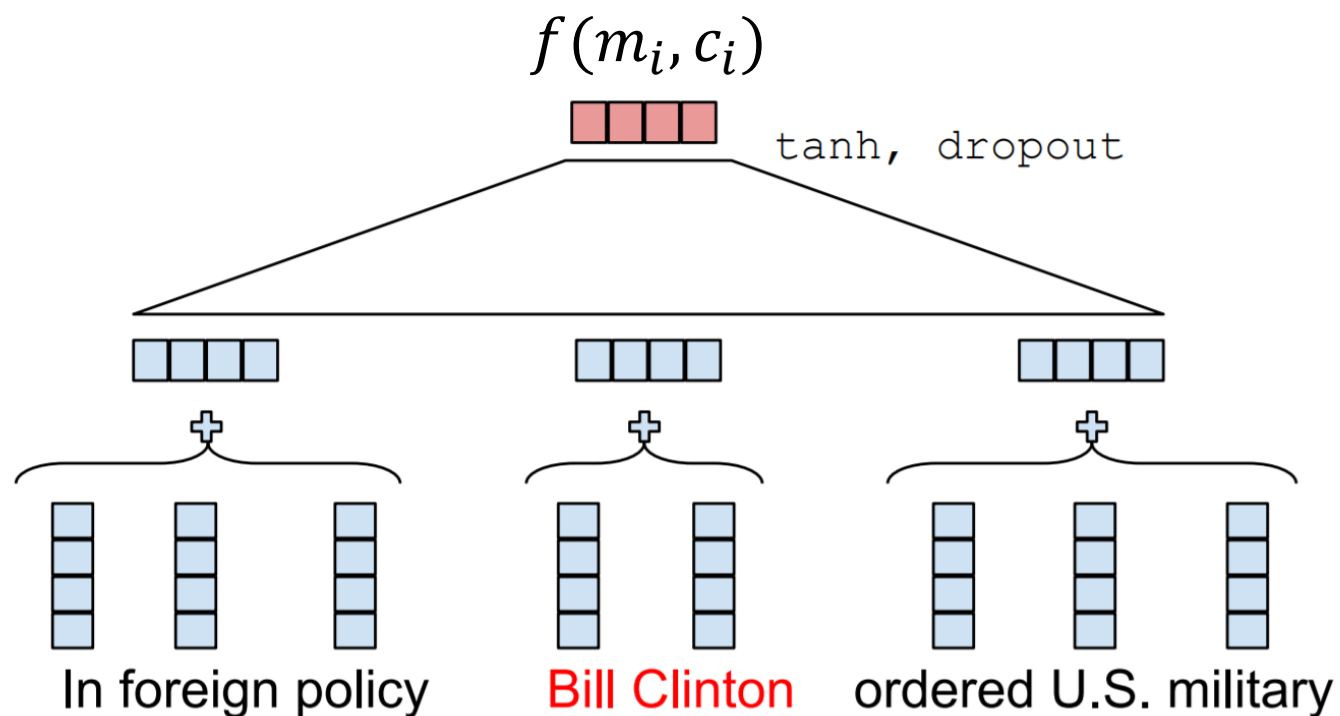
Ment-norm vs. rel-norm

Решение проблемы: *mention padding*

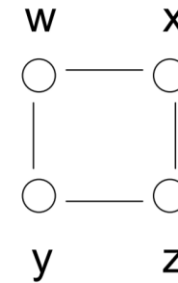
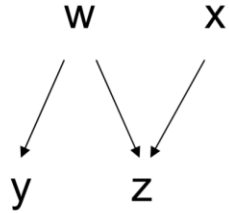
Можно добавить mention m_{pad} связанное с сущностью e_{pad} . За счет него модель может смещать свою вероятностную массу.

Реализация

$$\alpha_{ijk} = \frac{1}{Z_{ijk}} \exp \left\{ \frac{f^T(m_i, c_i) D_k f(m_j, c_j)}{\sqrt{d}} \right\}$$

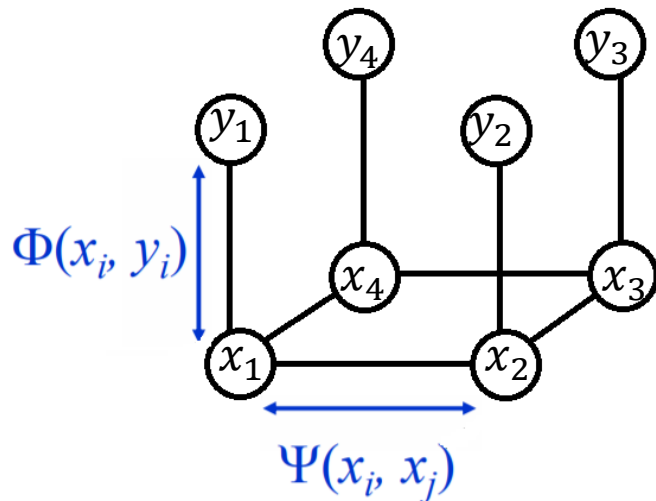


Графовые модели



$$P(w, x, y, z) = P(w)P(x)P(y|w)P(z|w, x)$$

$$P(w, x, y, z) = \frac{1}{Z} f_{wx}(w, x) f_{xz}(x, z) f_{yz}(y, z) f_{wy}(w, y)$$



$$P(x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_N) = \prod_{(i,j)} \Psi(x_i, x_j) \prod_p \Phi(x_p, y_p)$$

Belief propagation

- В случае графов без циклов дает точное решение
- Иначе “loopy” BP находит приблизительное решение
- Итеративный процесс передачи сообщений между вершинами
- В sum-product оцениваются маргинальные вероятности
- В max-product оцениваются MAP для всего MRF

Реализация

$$E^* = \operatorname{argmax}_{E \in C_1 \times \dots \times C_n} \sum_{i=1}^n \Psi(e_i, c_i) + \sum_{i \neq j} \Phi(e_i, e_j, D)$$

Реализуется, как CRF:

$$q(E|D) \propto \exp \left\{ \sum_{i=1}^n \Psi(e_i, c_i) + \sum_{i \neq j} \Phi(e_i, e_j, D) \right\}$$

Для оценки max-marginal вероятностей max-product loopy Belief Propagation

$$\hat{q}_i(e_i|D) \approx \max_{\substack{e_1, \dots, e_{i-1}, \\ e_{i+1}, \dots, e_n}} q(E|D)$$

Реализация

Для каждого m_i : $\hat{p}(e|m_i)$ – вероятность выбора e , обуславливаясь только на m_i

Финальная оценка: $\rho_i(e) = g(\hat{q}_i(e|D), \hat{p}(e|m_i))$

Пусть e^* – ground truth. Функция потерь по параметрам θ :

$$h(m_i, e) = \max(0, \gamma - \rho_i(e^*) + \rho_i(e))$$

$$L(\theta) = \sum_{D \in \mathcal{D}} \sum_{m_i \in D} \sum_{e \in C_i} h(m_i, e) + \lambda_1 \sum_{i,j} \text{dist}(R_i, R_j) + \lambda_2 \sum_{i,j} \text{dist}(D_i, D_j)$$

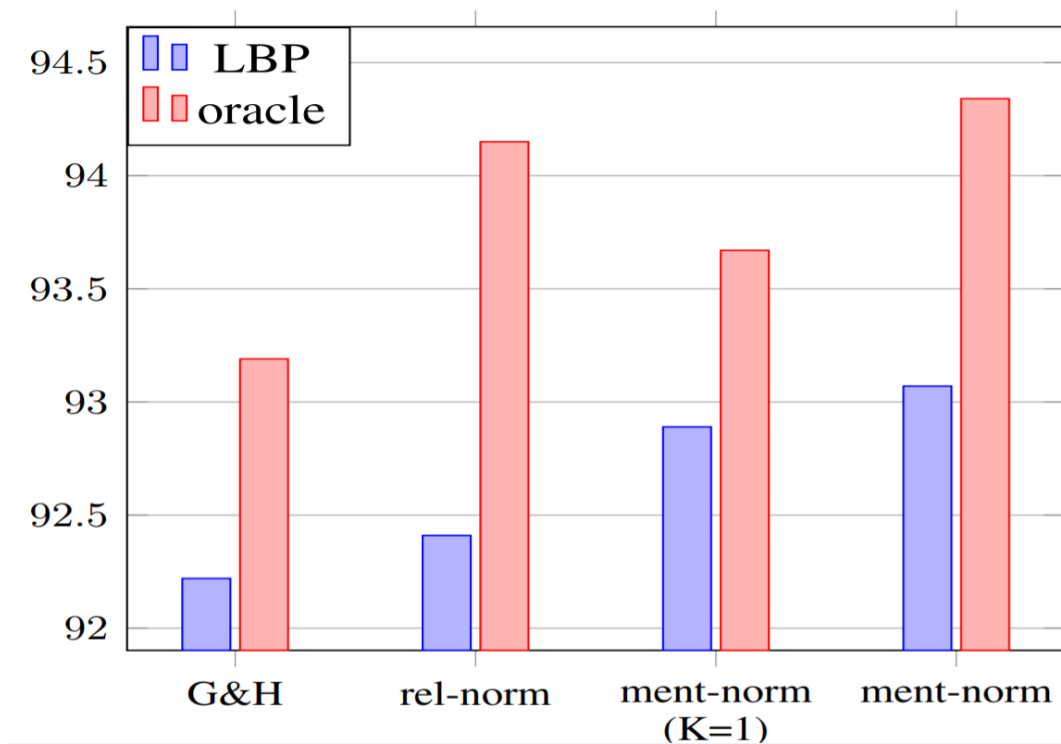
Candidate selection

- 1) Выбираются top-30 по $\hat{p}(e|m_i)$
- 2) По эмбедингам \mathbf{w} слов w контекста d_i упоминания m_i (окно из 50 слов) из них выбирается top-4 по $e^T \sum_{w \in d_i} \mathbf{w}$
- 3) Дополнительно выбираем top-3 по $\hat{p}(e|m_i)$

Результаты

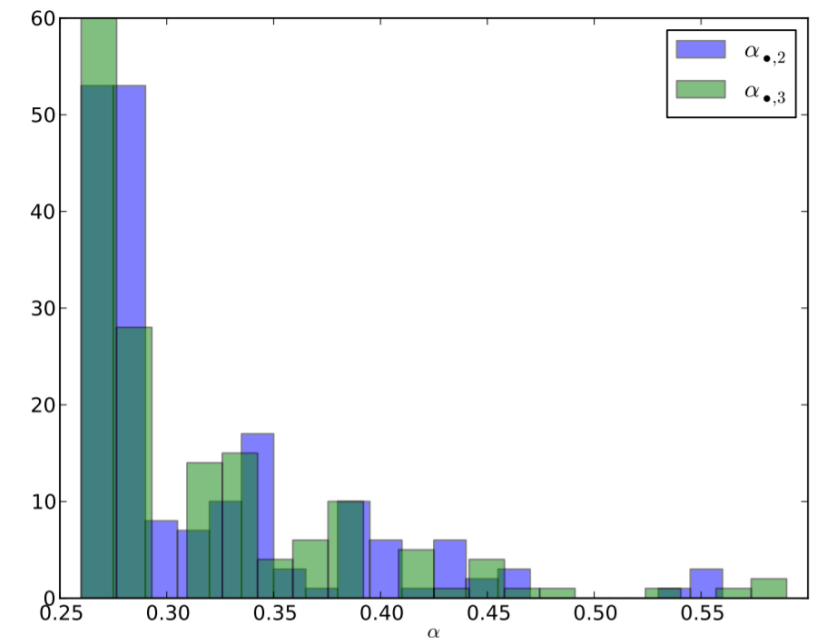
Methods	Aida-B
Chisholm and Hachey (2015)	88.7
Guo and Barbosa (2016)	89.0
Globerson et al. (2016)	91.0
Yamada et al. (2016)	91.5
Ganea and Hofmann (2017)	92.22 ± 0.14
rel-norm	92.41 ± 0.19
ment-norm	93.07 ± 0.27
ment-norm ($K = 1$)	92.89 ± 0.21
ment-norm (no pad)	92.37 ± 0.26

micro-F1 scores



Найденные отношения

rel-norm	on Friday , Liege police said in	ment-norm
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(1) missing teenagers in Belgium .	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(2) UNK BRUSSELS UNK	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(3) UNK Belgian police said on	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(4) , ” a Liege police official told	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(5) police official told Reuters .	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(6) eastern town of Liege on Thursday ,	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(7) home village of UNK .	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(8) link with the Marc Dutroux case , the	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	(9) which has rocked Belgium in the past	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>



ИСТОЧНИКИ

- [1] Phong Le and Ivan Titov, “Improving Entity Linking by Modeling Latent Relations between Mentions”, ACL 2018, <https://arxiv.org/pdf/1804.10637.pdf>
- [2] James Coughlan, “A Tutorial Introduction to Belief Propagation”, http://www.computerrobotvision.org/2009/tutorial_day/crv09_belief_propagation_v2.pdf