

Coreference Resolution

Референциальные цепочки

Четверо студентов из Индии утонули в Волге. Они не умели плавать
время публикации: 09 августа 2020 г., 15:12 | последнее обновление: 09 августа 2020 г., 21:24

Четверо граждан Индии, учащихся волгоградского вуза, утонули в реке Волга. Они отдыхали с сокурсниками на берегу Волги на пляже хутора Бобры Среднеахтубинского района, передает "Интерфакс" со ссылкой на пресс-службу следственного управления СК по региону.

По предварительной версии следствия, около 2 часов утра 8 августа один из них решил искупаться. Когда он отплыл достаточно далеко от берега, течение стало его уносить, и он начал звать на помощь.

Сначала к тонущему подплыл другой индийский студент и начал тащить его, но у него не получилось, тогда к ним добрались еще двое. В итоге по цепочке терпящий бедствие потащил за собой всех товарищей на дно. Как выяснилось, все четверо не умели плавать, выбраться из воды им не удалось.

Тела погибших обнаружили спустя сутки сотрудниками спасательной службы.

Водолазы смогли к 7 утра по местному времени найти и поднять тела всех погибших. Они находились на дне примерно в одном месте.

По факту произошедшего Среднеахтубинским межрайонным следственным отделом Следственного управления СКР по области организована доследственная проверка.

Подробнее: <https://www.newsru.com/russia/09Aug2020/volgariver.html>

Entities

Волге
следственного управления СК по р
региону
один из них
другой индийский студент
двое
Тела погибших
сотрудниками спасательной службы

Parent Entities

Четверо студентов из Индии
ним
всех товарищей

Метрики

- MUC
- B^3
- CEAFE
- CONLL-2012
- LEA

ML-подходы

- (Mention detection)

- Mention - pair

(классификация пар упоминаний)

- Entity - mention

(классификация пара – еще одно упоминание)

- Mention ranking

(для каждого упоминания ищется наиболее вероятный antecedent)

- Cluster ranking

(ранжируем упоминания и решаем, включать в кластер для сущности или нет)

[John]₁¹ is a musician. [He]₃¹ played [a new song]₄⁴.
[A girl]₅⁵ was listening to [the song]₆⁴.
“[It]₇⁴ is [my]₉¹ favourite,” [John]₁₀¹ said to [her]₁₁⁵.



End-to-end coreference resolution

- End-to-end Neural Coreference Resolution ([Lee et al. 2017](#))
- Mention ranking model
- Подход используется до сих пор (67.2 -> 82.2 F1)

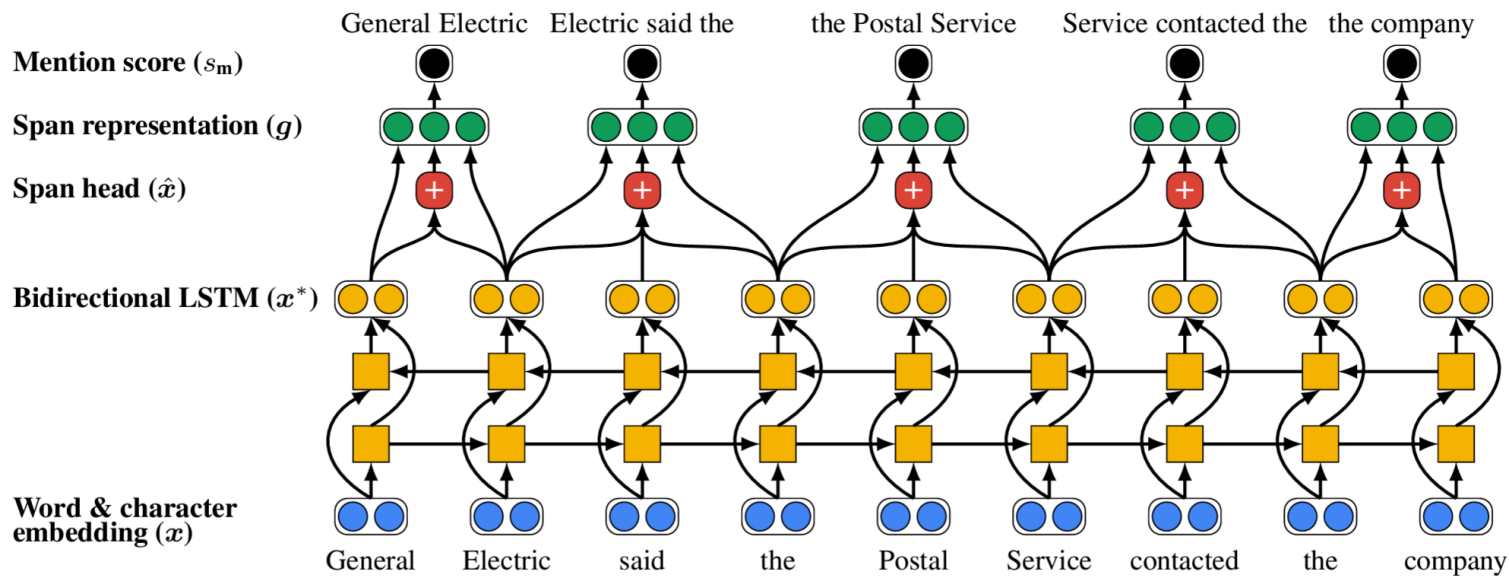
End-to-end coreference resolution

- **Mention extraction**
 - **Mention scoring**
 - **Mention pruning**
 - **Mention pair scoring**
 - **Clustering**
- *It is a period of civil war. Rebel spaceships, striking from a hidden base, have won their first victory against the evil Galactic Empire.*
 - Кандидаты в упоминания: It, It is, It is a, It is a period, It is a period of, It is a period of civil, It is a period of civil war. ...
 - [start, end, weighted sum of all tokens, features]
 - 0.2, -0.6, -5.2, -0.5, -7.1 ...
 - берем λT упоминаний с лучшими оценками (T – длина документа)

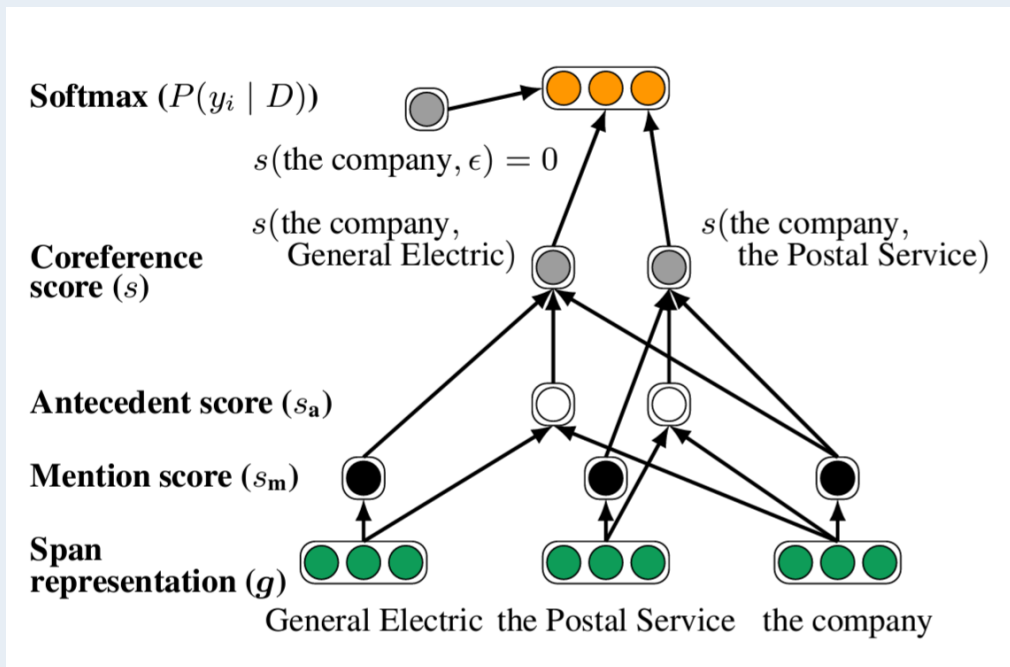
End-to-end coreference resolution

- Берем K ближайших antecedентов: $[\lambda T, K, \text{pair embedding size}]$
- Оцениваем матрицу, используя FFNN: $[\lambda T, K]$
- Итоговый coreference score $S(a, b) = S_m(a) + S_m(b) + S_a(a, b)$
- Добавим пустой antecedент со скором 0: $[\lambda T, K + 1]$
- Для каждого спана выберем antecedент с самым высоким скором (если это окажется наша пустышка, значит, у спана нет antecedента)

End-to-end coreference resolution



End-to-end coreference resolution





+ efficiency

- Lee et al., 2018
- Добавим coarse scoring function:

$$S(a, b) = S_m(a) + S_m(b) + S_c(a, b) + S_a(a, b)$$

$$S_c(a, b) = aW_cb^T$$

- Позволит нам учесть не ближайшие K antecedентов, а наилучшие варианты
- 
- 
- 