



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

BÀI 7: TRÍCH RÚT THÔNG TIN

Nội dung

1. Kiến trúc hệ thống trích rút thông tin
2. Nhận diện thực thể có tên
3. Trích rút quan hệ không giám sát
4. Giám sát từ xa
5. Phân giải đồng tham chiếu

1. Kiến trúc hệ thống trích rút thông tin

- Trích rút thông tin là quá trình tìm kiếm các thực thể và mối quan hệ giữa các thực thể này trong văn bản
- Trích rút thông tin phục vụ khai phá văn bản ở mức chính xác và cô đọng hơn các tác vụ như phân loại văn bản hay gán nhãn văn bản
- Các loại thực thể và quan hệ được định nghĩa trước

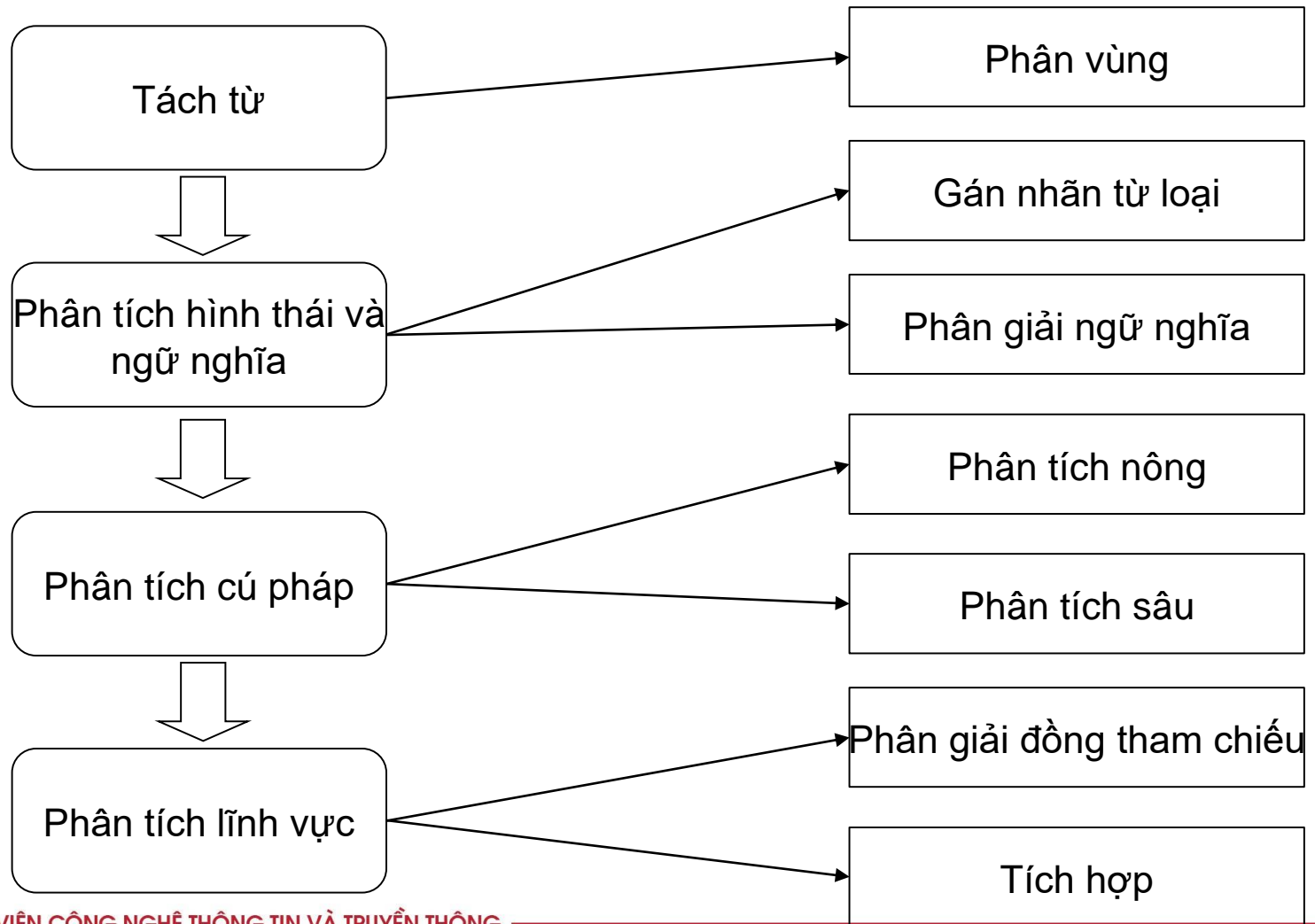
Các giả thiết của trích rút thông tin

- Thông tin được thể hiện một cách tường minh và không yêu cầu suy diễn
- Một số lượng nhỏ khuôn mẫu có thể tóm tắt được nội dung của văn bản
- Thông tin cần thiết xuất hiện cục bộ trong văn bản

Các loại thông tin được trích rút

- Thực thể: Con người, tổ chức, địa điểm,...
- Thuộc tính (của thực thể): Chức danh, tuổi, loại tổ chức...
- Thực tế: quan hệ giữa nhân viên và công ty, quan hệ giữa virus và bệnh,...
- Sự kiện: hai công ty sát nhập, động đất, khủng bố,...

Kiến trúc hệ thống trích rút thông tin



Nhận diện thực thể có tên

- Phát hiện các thực thể có tên trong văn bản và phân loại vào các lớp được định nghĩa trước

[Forbes]_{ORG} : [Việt Nam]_{LOC} có 4 tỷ phú

Phân cụm

- Phát hiện các cụm danh từ và động từ trong câu

Trong đó , Việt Nam có 4 đại diện là Chủ tịch Vingroup Phạm Nhật Vượng , CEO VietJet Air Nguyễn Thị Phương Thảo , Chủ tịch Thaco Trần Bá Dương và Chủ tịch Techcombank Hồ Hùng Anh .

Trích rút quan hệ

- Trích rút các quan hệ giữa các thực thể (thuộc tính, thực tế, sự kiện)

Goldman Sachs Group thì đi vay tiền của Cục Dự trữ Liên bang Mỹ.

Aikido là một môn võ thuật Nhật Bản hiện đại

Phân giải đồng tham chiếu

- Phát hiện sự xuất hiện của cùng một thực thể dưới dạng các tham chiếu khác nhau

Aikido₁ là một môn võ thuật Nhật Bản hiện đại được phát triển bởi Ueshiba Morihei₂ như một sự tổng hợp các nghiên cứu võ học, triết học và tín ngưỡng tôn giáo của ông₂. Aikido₁ thường được dịch là "con đường hợp thông (với) năng lượng cuộc sống" hoặc "con đường của tinh thần hài hòa". Mục tiêu của Ueshiba₂ là tạo ra một nghệ thuật₁ mà các môn sinh₃ có thể sử dụng để tự bảo vệ mình₃ trong khi vẫn bảo vệ người tấn công₄ khỏi bị thương. Các kĩ thuật của Aikido₁ bao gồm: irimi (nhập thân), chuyển động xoay hướng (tenkan - chuyển hướng đà tấn công của đôi phương₄), các loại động tác ném và khóa khớp khác nhau.

2. Nhận diện thực thể có tên

- Dựa trên từ điển:
 - Có thể phát hiện được các thực thể phổ biến
 - Yêu cầu xây dựng từ điển tên riêng
 - Không xử lý được nhập nhằng
- Dựa trên biểu thức chính quy
 - Sử dụng kiến thức chuyên gia
 - Có thể phát hiện được các mẫu phổ biến

Dựa trên học máy

- Yêu cầu dữ liệu huấn luyện
- Độ chính xác không thay đổi nhiều giữa các lĩnh vực
- Quy về bài toán gán nhãn chuỗi BIO
 - Đầu vào là một câu
 - Đầu ra là nhãn của mỗi từ trong câu

Gán nhãn chuỗi BIO

- B: Begin
- I: Inside
- O: Outside

B-ORG I-ORG I-ORG O O O O O B-ORG I-ORG I-ORG I-ORG
↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑
Goldman Sachs Group thì đi vay tiền của Cục Dự_trữ Liên_bang Mỹ

Tập đặc trưng

- Các từ trong cửa sổ $[-k, k]$ ($k = 2, 3$)
- Hình thái từ:
 - Viết hoa, viết thường
 - Chữ số
 - Dấu câu
- Loại từ: Đầu ra của bài toán gán nhãn từ loại
- Phạm vi từ: Đầu ra của bài toán phân cụm

NER dựa trên CRF

- [1]: Sử dụng PoS và phân cụm chuẩn
- [2, 3]: PoS và phân cụm tự động bởi công cụ *NNVLP* và *Underthesea*
- [4]: Không sử dụng PoS và phân cụm

Table 4. Accuracy of our NER system with default and generated PoS, chunking tags; and without PoS and chunking tags

Setting	Precision	Recall	F_1
Default PoS and chunking tags	93.87	93.99	93.93
PoS and chunking tags generated by NNVLP 7	90.21	86.72	88.43
PoS and chunking tags generated by Underthesea	90.28	88.35	89.3
Without PoS, chunking tags	89.91	90.15	90.03

Đánh giá kết quả

- [1]: Sử dụng PoS chuẩn
- [2-6]: PoS tự động từ các công cụ
- [7]: Không sử dụng PoS và phân cụm

Table 5. Proposed NER systems without chunking tag-based features. We compare default PoS with PoS generated by other tools.

Setting	Precision	Recall	F_1
Default PoS tags	90.13	90.55	90.34
PoS by NNVLDP [7]	90.05	85.65	88.31
PoS by Underthesea	90.27	88.58	89.42
PoS by Pyvi	90.16	88.72	89.43
PoS by Vtik	89.62	86.42	87.99
PoS by VnMarMoT [19]	90.51	89.15	89.83
Without PoS, chunking tags	89.91	90.15	90.03

Đánh giá kết quả (tiếp)

- [1]: Sử dụng tách từ chuẩn
- [2,3]: Tách từ tự động sử dụng *UETSegmenter* và *RDRSegmenter*

Table 6. Accuracy of NER system with default and generated word segmentation. We did not use features based on PoS, chunking tags here.

Setting	Precision	Recall	F_1
Default Word segmentation	89.91	90.15	90.03
Word segmentation generated by UETSegmenter	87.67	84.95	86.29
Word segmentation generated by RDRsegmenter	89.05	84.98	86.97

Đánh giá kết quả (tiếp)

- [1]: Mô hình dựa trên tiếng (không tách từ)
- [2]: Sử dụng tách từ chuẩn
- [3]: Tách từ tự động bằng công cụ *RDRSegmenter*

Table 7. Accuracy of NER system with syllable-based and word-based model. We do not use features based on PoS and chunking tags. “ws” stands for word segmentation

Setting	Precision	Recall	F_1
Syllable-based model	88.78	82.94	85.76
Word-based model with gold ws	89.91	90.15	90.03
Word-based model with ws generated by RDRsegmenter	89.05	84.98	86.97

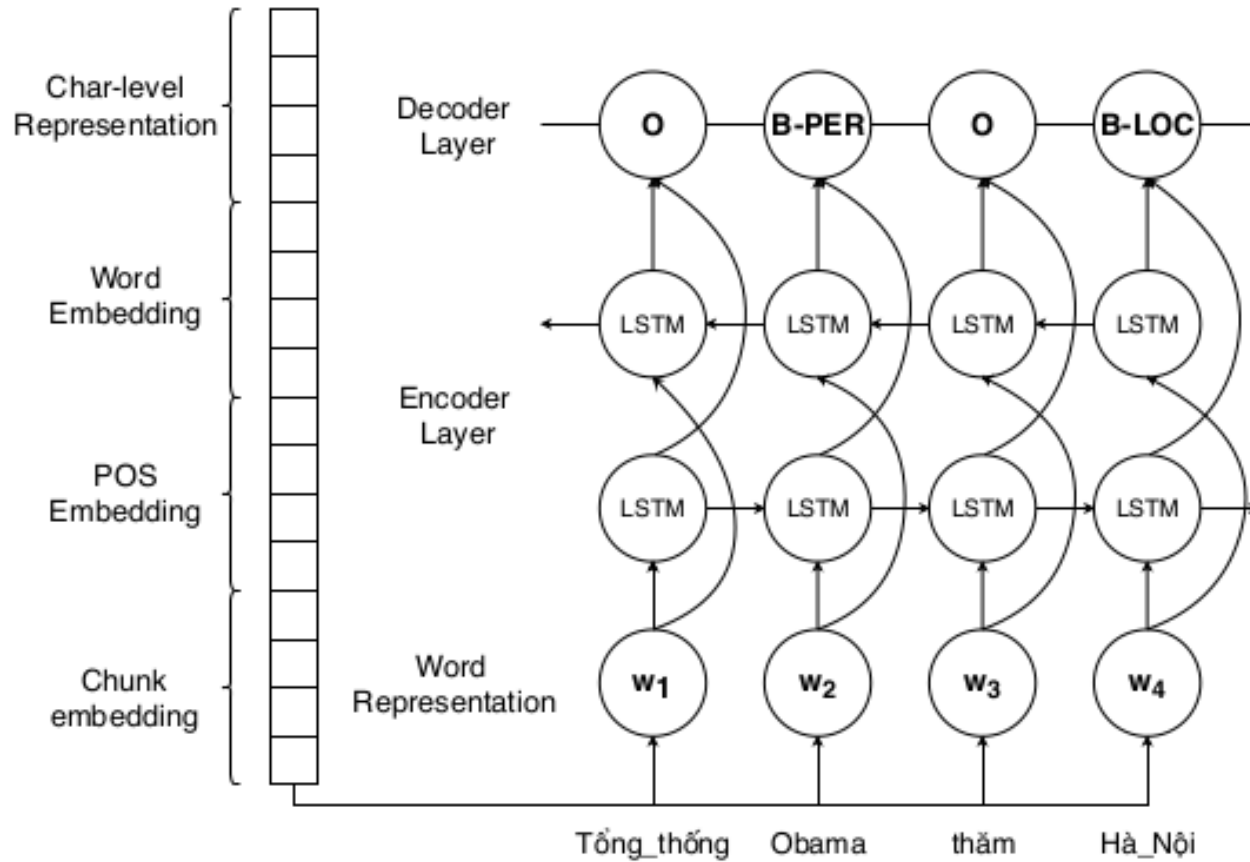
Đánh giá kết quả (tiếp)

- Word: Các từ trong cửa sổ
- Word shapes: Hình thái từ
- w2v: Biểu diễn từ nhúng
- Cluster: Biểu diễn phân cụm Brown

Table 8. Impact of word representation-based features. w2v denotes features based on word embeddings. “cluster” denotes cluster-based features.

Setting	Precision	Recall	F_1
(1) = all features with default PoS, Chunk	93.87	93.99	93.93
(2) = (1) - cluster - w2v	91.66	92.02	91.84
(4) = word + word shapes + default PoS	88.01	87.95	87.98
(5) = word + word shapes + cluster + w2v	89.91	90.15	90.03
(6) = word + word-shapes	88.17	88.08	88.13
(7) = word + word-shapes + w2v	88.69	88.72	88.70
(8) = word + word-shapes + cluster	88.96	89.99	89.97

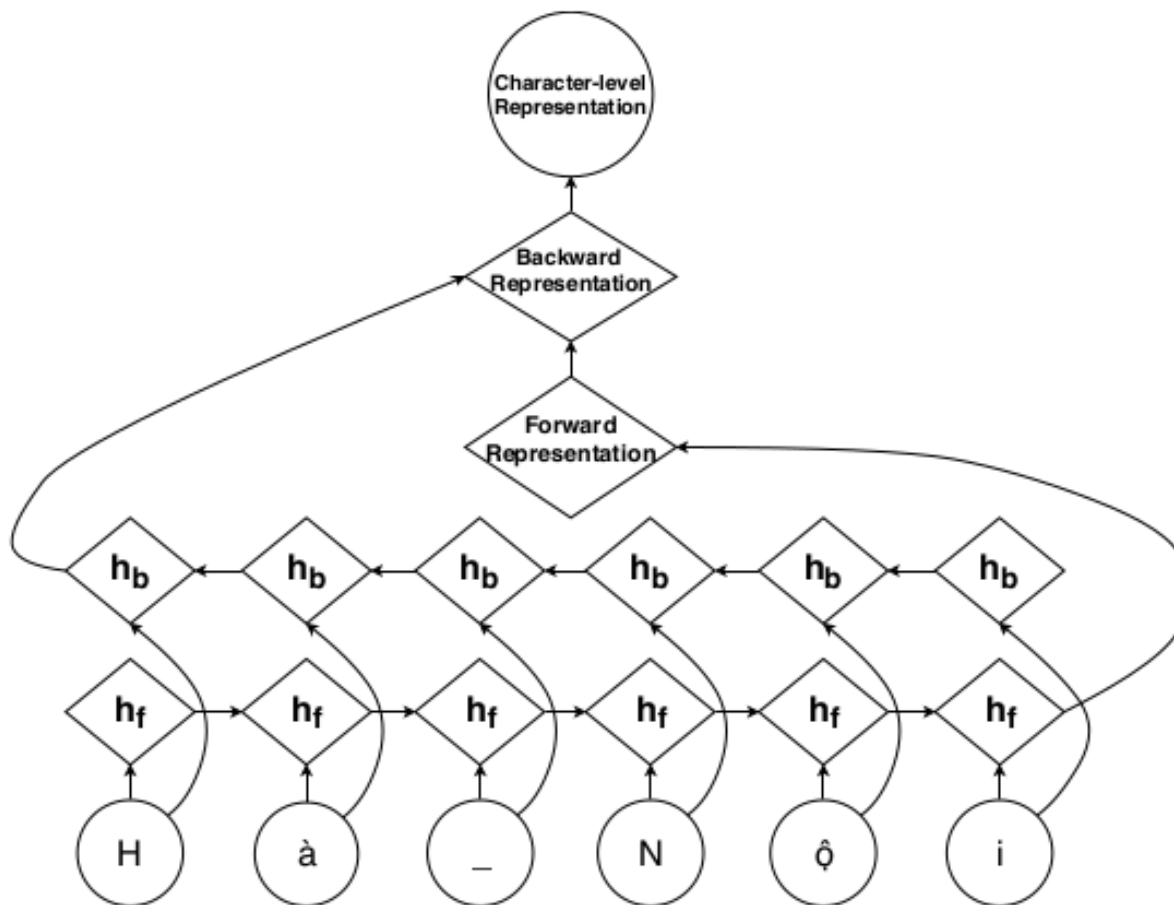
NER dựa trên RNN



Tầng đầu vào

- Biểu diễn nhúng kết hợp:
 - Biểu diễn từ: Sử dụng từ nhúng huấn luyện trước bởi word2vec trên 2 triệu văn bản
 - Biểu diễn ký tự: Sử dụng mạng LSTM hai chiều để học biểu diễn ký tự với khởi tạo ngẫu nhiên
 - Biểu diễn từ loại: Biểu diễn one-hot
 - Biểu diễn cụm: Biểu diễn one-hot

Học biểu diễn ký tự



LSTM hai chiều

- Sử dụng hai mạng LSTM theo chiều tiến và chiều lùi
 - Mục đích: Các từ ở đầu câu có thể sử dụng cả thông tin ở cuối câu để dự đoán và ngược lại
- Đầu ra được ghép nối để đưa vào tầng đầu ra

Tầng đầu ra

- Dự đoán các nhãn BIO ứng với các loại thực thể
 - VD: Với 3 loại thực thể ORG, PER, LOC, tập nhãn có 7 nhãn (B-ORG, I-ORG, B-PER, I-PER, B-LOC, I-LOC, O)
- Tầng đầu ra có thể được đưa vào một mô hình CRFs để thể hiện quan hệ với nhãn ở thời điểm trước thông qua xác suất chuyển đổi

Đánh giá kết quả

Method	P	R	F1	F1 (w.o char)
Feature-rich CRFs [25]	93.87	93.99	93.93	-
NNVLP [7]	92.76	93.07	92.91	-
BiLSTM-CRFs	90.97	87.52	89.21	76.43
BiLSTM-CRFs + POS	90.90	90.39	90.64	86.06
BiLSTM-CRFs + Chunk	95.24	92.16	93.67	87.13
BiLSTM-CRFs + POS + Chunk	95.44	94.33	94.88	91.36

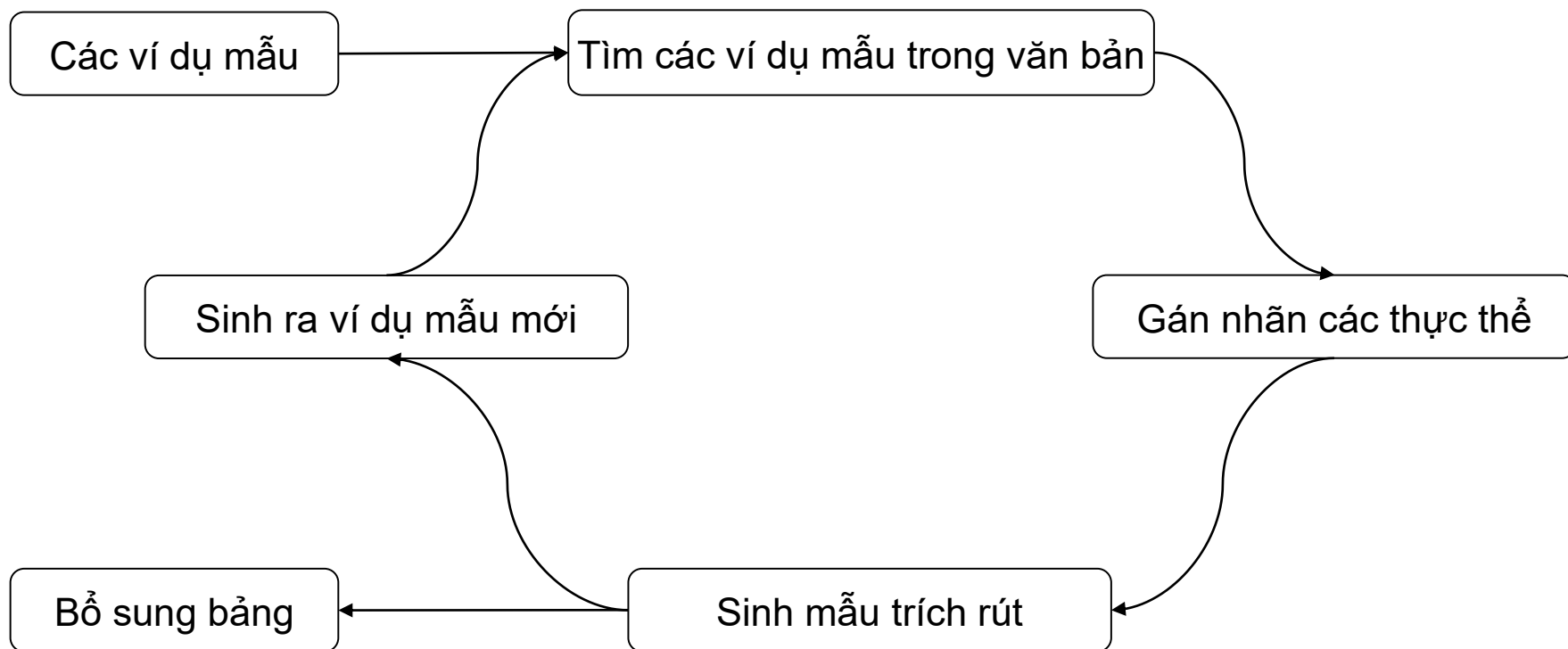
BiLSTM-CRFs sử dụng thêm các thông tin PoS và phân cụm

BiLSTM-CRFs không kết hợp biểu diễn mức kí tự

3. Trích rút quan hệ không giám sát

- Học có giám sát có độ chính xác cao nhưng đòi hỏi DL huấn luyện
- Học không giám sát tận dụng được lượng DL lớn nhưng có độ chính xác thấp hơn
- Giám sát từ xa tận dụng được cơ sở tri thức và cải thiện độ chính xác so với học không giám sát

Snowball



Các ví dụ mẫu

- Do người dùng cung cấp
- Sau đó hệ thống tự động trích rút ra từ văn bản
- VD: Quan hệ <tập đoàn, trụ sở>
 - <Microsoft, Redmond>
 - <Exxon, Irving>
 - <IBM, Armonk>

Tìm các ví dụ mẫu trong văn bản

- “Hệ thống máy chủ của **Microsoft** nằm ở trụ sở chính **Redmon**”
- “**Exxon, Irving** đang dần trở thành tập đoàn dầu khí...”
- “Tin đồn rút nhân viên khỏi Iraq đến từ trụ sở chính của **Exxon, Irving...**”
- “... vừa nhận được email từ trụ sở chính của **Boeing** ở **Seattle.**”

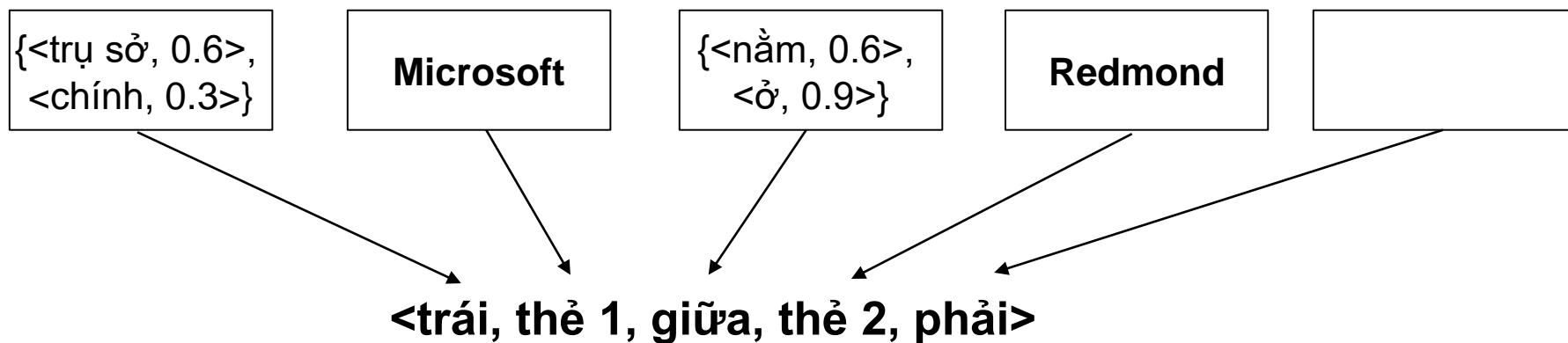
Gán nhãn thực thể

- “Hệ thống máy chủ của <ORG> nằm ở trụ sở chính <LOC>”
- “<ORG>, <LOC> đang dần trở thành tập đoàn dầu khí...”
- “Tin đồn rút nhân viên khỏi Iraq đến từ trụ sở chính của <ORG>, <LOC>...”
- “... vừa nhận được email từ trụ sở chính của <ORG> ở <LOC>.”

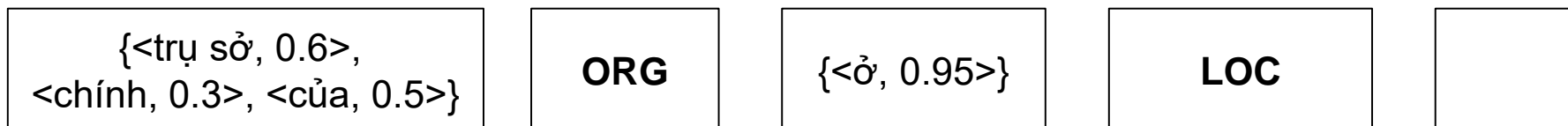
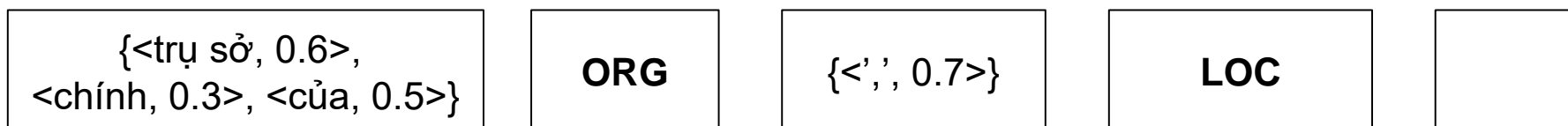
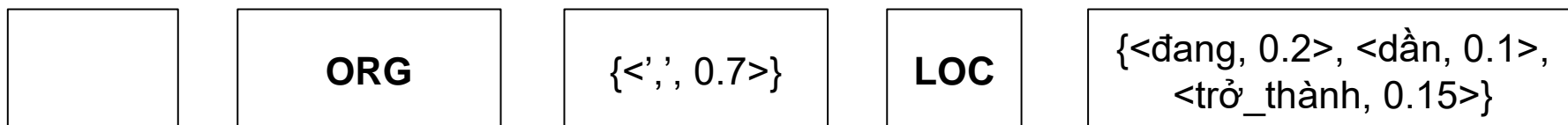
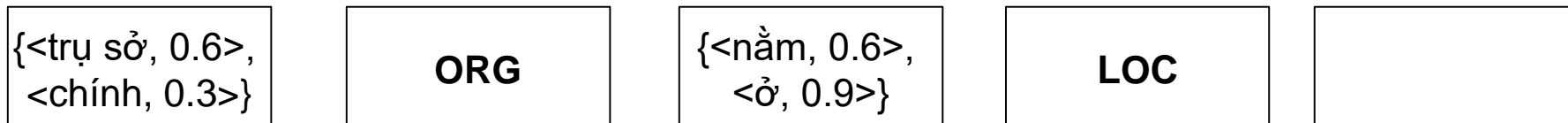
Sinh 5-tuple

- 5-tuple: <trái, thẻ 1, giữa, thẻ 2, phải>
- Trái: k từ ở bên trái cùng với véc-tơ trọng số
- Thẻ 1: thực thể thứ nhất
- Giữa: các từ ở giữa cùng với véc-tơ trọng số
- Thẻ 2: thực thể thứ hai
- Phải: k từ ở bên phải cùng với véc-tơ trọng số

Sinh 5-tuple (tiếp)



Sinh 5-tuple (tiếp)



Sinh mẫu trích rút

- Cho 2 5-tuple có cùng thẻ tag_1 và tag_2 :
 - $t = \{l, tag_1, m, tag_2, r\}$
 - $t' = \{l', tag_1, m', tag_2, r'\}$
- Độ tương đồng: $match(t, t') = l \cdot l' + m \cdot m' + r \cdot r'$
- Phân cụm các 5-tuple dựa trên độ tương đồng
- Với mỗi cụm, lấy centroid của c làm mẫu trích rút
 $p = \{l_c, tag_1, m_c, tag_2, r_c\}$

Sinh ra ví dụ mẫu mới

Algorithm GenerateTuples(tập mẫu)

1. **foreach** đoạn \in tập văn bản **do**
2. $\{<o, l, <l_s, t_1, m_s, t_2, r_s>\} = \text{CreateOccurrence}(\text{đoạn});$
3. $T_C = <o, l>;$
4. $\text{Sim}_{\text{Best}} = 0;$
5. **foreach** $p \in$ tập mẫu
6. $\text{sim} = \text{Match}(<l_s, t_1, m_s, t_2, r_s>, p);$
7. **if** $(\text{sim} \geq \tau_{\text{sim}})$ **then**
8. $\text{UpdatePatternSelectivity}(p, T_C);$
9. **if** $(\text{sim} \geq \text{Sim}_{\text{Best}})$ **then**
10. $\text{Sim}_{\text{Best}} = \text{sim};$
11. $P_{\text{Best}} = p;$
12. **endif**
13. **endif**
14. **endfor**
15. **if** $(\text{Sim}_{\text{Best}} \geq \tau_{\text{sim}})$ **then**
16. $\text{CandidateTuples}[T_C].\text{Patterns}[P_{\text{Best}}] = \text{Sim}_{\text{Best}};$
17. **endif**
18. **endfor**
19. **return** CandidateTuples;

Đánh giá mẫu

- Với mỗi ví dụ $\langle \text{org}, \text{loc} \rangle$, phân loại
 - Positive nếu đã tồn tại ví dụ mẫu
 - Negative nếu tồn tại ví dụ mẫu $\langle \text{org}, \text{loc}' \rangle$
 - Unknown nếu $\langle \text{org}, * \rangle$ chưa tồn tại
- Độ tin tưởng của mẫu P:

$$\text{conf}(P) = \frac{\text{P.positive}}{\text{P.positive} + \text{P.negative}}$$

- P.positive: số ví dụ positive khớp với P
- P.negative: số ví dụ negative khớp với P

Đánh giá ví dụ

- Độ tin tưởng của ví dụ $T = \{org, loc\}$

$$Conf(T) = 1 - \prod_{i=0}^{|P|} (1 - (Conf(P_i) \cdot Match(C_i, P_i)))$$

- $P = \{P_i\}$ là tập các mẫu sinh ra ví dụ T
- C_i là 5-tuple ứng với đoạn văn bản khớp với P_i với độ tương tự $Match(C_i, P_i)$
- Tập ví dụ mẫu = $\{T \mid Conf(T) > \tau_t\}$

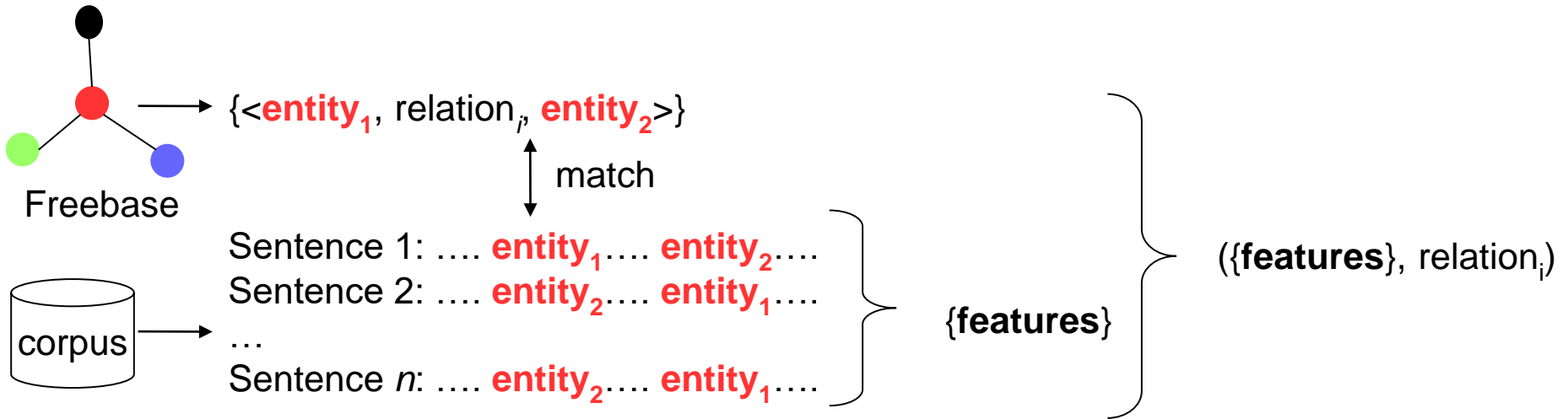
Ưu, nhược điểm

- Ưu điểm:
 - Tận dụng được dữ liệu không có nhãn
 - Chỉ cần một số ít ví dụ mẫu gốc
- Nhược điểm:
 - Vẫn yêu cầu gán nhãn thủ công từ người dùng
 - Quá trình lặp dẫn đến suy giảm chất lượng

4. Giám sát từ xa

- Freebase là cơ sở tri thức lớn và có chất lượng về quan hệ giữa các thực thể
- Freebase được xây dựng từ Wikipedia
- Giám sát từ xa:
 - Freebase giám sát quá trình trích rút quan hệ từ văn bản
 - Freebase + corpus = dữ liệu có nhãn

Mô hình giám sát từ xa



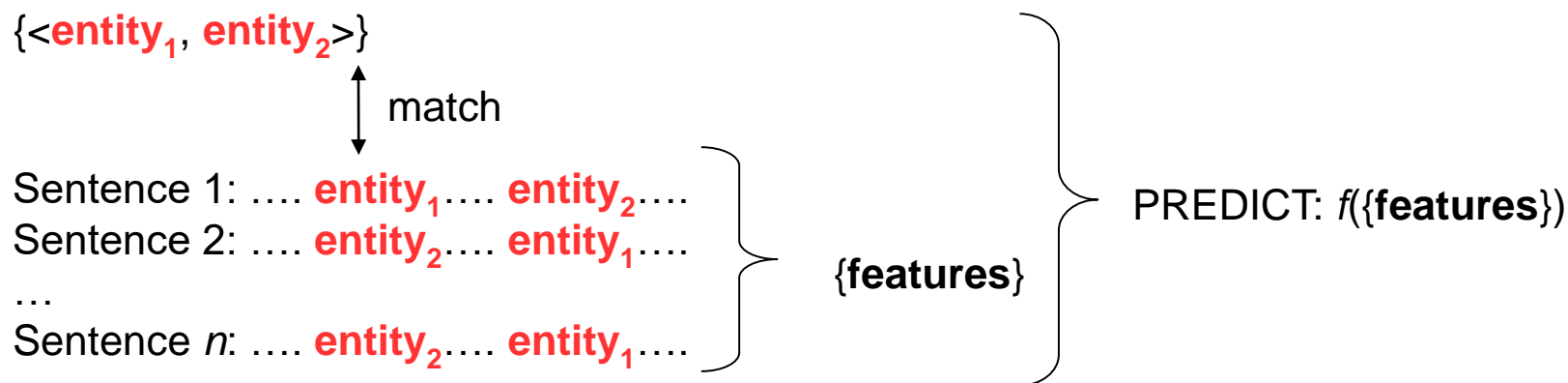
$\{ \langle \text{entity}_1', \text{relation}_j, \text{entity}_2' \rangle \} \rightarrow (\{ \text{features}' \}, \text{relation}_j)$

$\{ \langle \text{entity}_1'', \text{relation}_j, \text{entity}_2'' \rangle \} \rightarrow (\{ \text{features}'' \}, \text{relation}_j)$



BỘ PHÂN LOẠI ĐA LỚP f : $\{ \text{relation}_1, \text{relation}_2, \dots, \text{relation}_m \}$

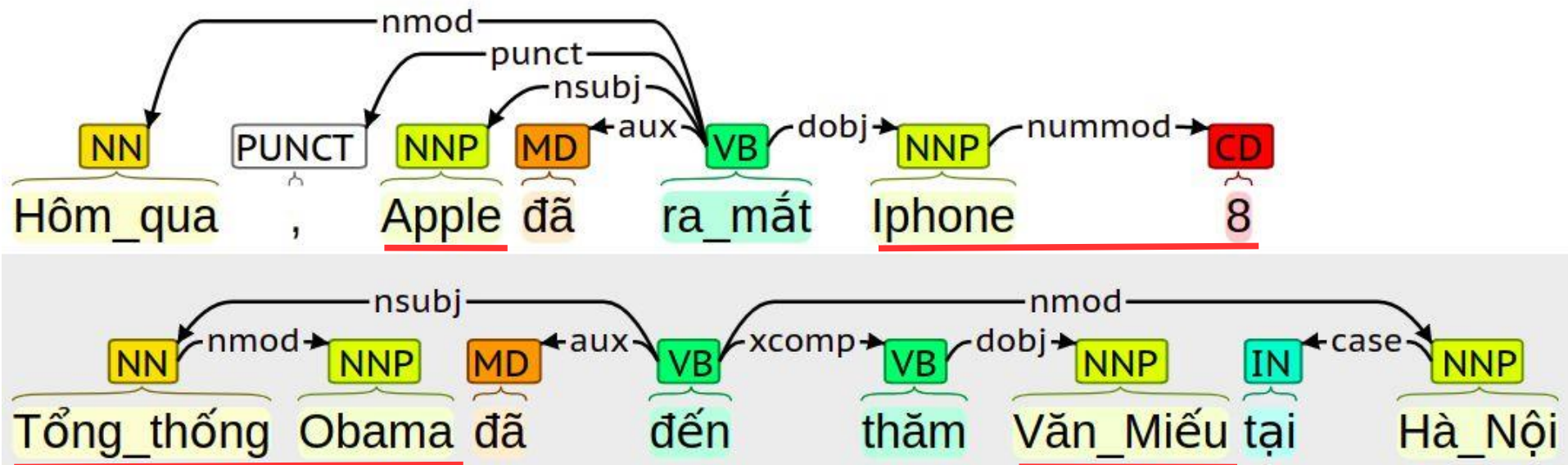
Mô hình giám sát từ xa (tiếp)



Tập đặc trưng

- Những từ và POS ở giữa hai thực thể và PoS
- Thứ tự của hai thực thể
- Từ và POS của k từ ở bên trái
- Từ và POS của k từ ở bên phải
- Loại thực thể
- Đường đi giữa hai thực thể trên cây phụ thuộc

Quan hệ phụ thuộc



5. Phân giải đồng tham chiếu

- Phân giải đồng tham chiếu là quá trình phát hiện một cặp từ hay cụm từ trong văn bản chỉ đến cùng một thực thể
- Đồng tham chiếu là hiện tượng phổ biến trong ngôn ngữ
- Phân giải đồng tham chiếu có vai trò quan trọng với các ứng dụng khai phá văn bản

Các loại đồng tham chiếu

- Đại từ làm chủ ngữ: “**Cô ta** đang học trực tuyến”
- Đại từ làm tân ngữ: “Hãy liên lạc với **anh ấy** ngay”
- Đại từ sở hữu: “Lịch trình của **chúng ta** đã được thống nhất”
- “Anh ta tự làm khó **mình**”

Các loại đồng tham chiếu (tiếp)

- Tên riêng: “**Thủ tướng Nguyễn Xuân Phúc** tuyên bố giãn cách xã hội. **Thủ tướng Phúc** cũng yêu cầu người dân tự giác thực hiện các quy định.”
- Apposition: “**Phạm Nhật Vượng, Chủ tịch Vingroup** là một trong số các tỉ phú được Forbes nêu tên.”
- Động từ ‘là’ : “**Park Hang Seo là HLV trưởng đội tuyển bóng đá nam Việt Nam.**”

Các loại đồng tham chiếu (tiếp)

- Nhóm người: “**Mây Trắng** tuyên bố tái hợp. **Nhóm** dự định ra mắt album mới đầu năm sau.”
- Thuộc tính - giá trị: “**Giá cổ phiếu VIC là 94.800 VND**”
- Thứ tự: “IBM và **Microsoft** là những ứng cử viên cuối cùng, nhưng đại diện nhà đầu tư ưu tiên **ứng cử viên thứ hai.**”
- Bộ phận - toàn thể: “Vinfast mới ra mắt **dòng xe mới. Bộ truyền động** sử dụng công nghệ CVT vô cấp tiên tiến.”

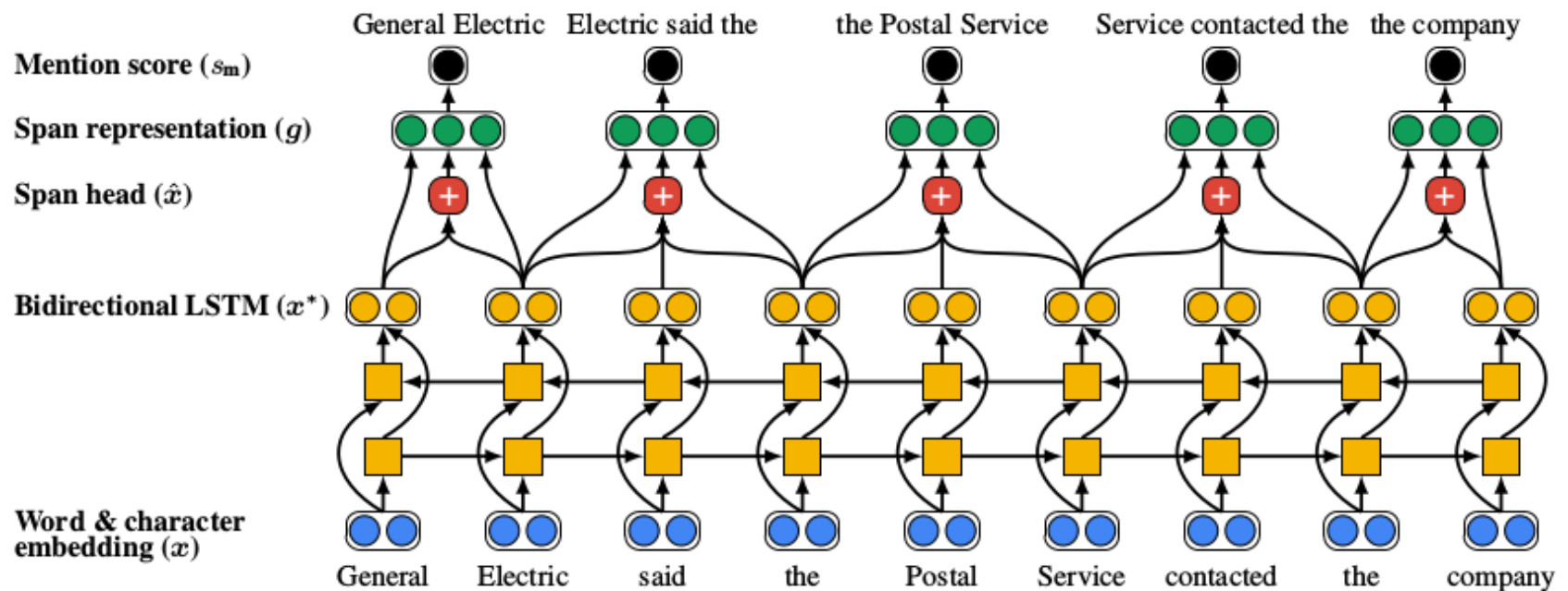
Các phương pháp truyền thống

- Tập trung vào đại từ là loại xuất hiện phổ biến nhất
- Dùng thông tin ngôn ngữ để phát hiện các ứng cử viên đằng trước
- Loại bỏ các ứng cử viên dựa trên tính chất như giới tính, số ít số nhiều, ...
- Tính điểm các ứng cử viên
 - So khớp
 - Luật
 - Học máy

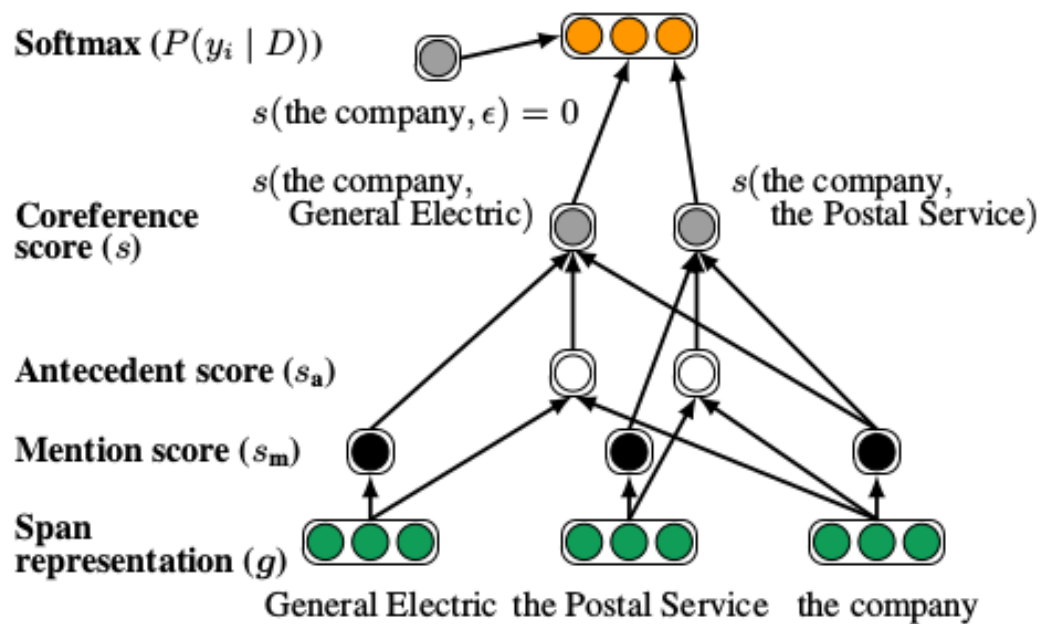
P² dựa trên mạng nơ-ron

- Hạn chế sử dụng các đặc trưng phức tạp
- Hạn chế sử dụng phân tích cú pháp
- Tận dụng biểu diễn học trước
- Thách thức:
 - Sử dụng thông tin thay thế cho thông tin cú pháp
 - Biểu diễn cụm từ, ngữ cảnh
 - Phân giải đồng tham chiếu bản chất là một bài toán phân cụm cứng trong phạm vi văn bản

Kiến trúc mô hình



Kiến trúc mô hình (tiếp)



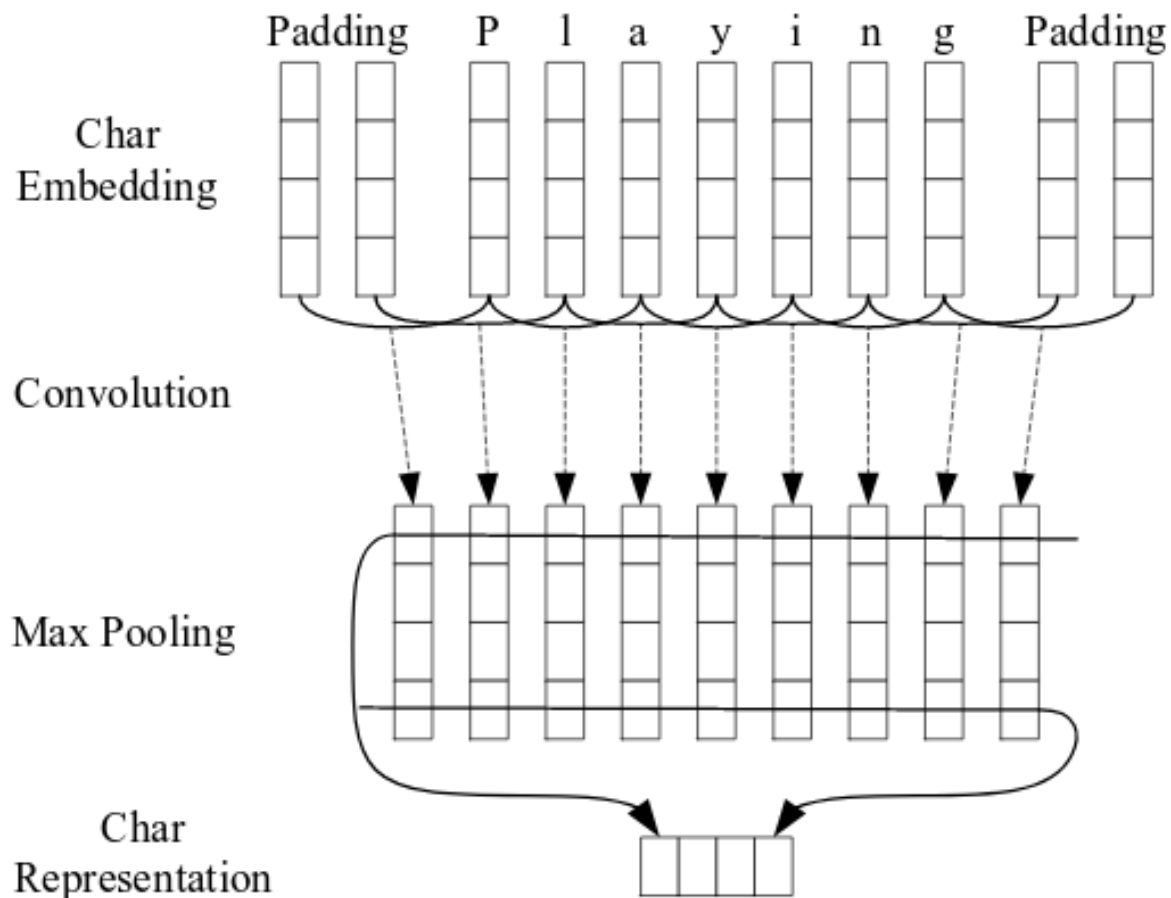
Phát biểu bài toán

- Một văn bản D gồm một chuỗi các từ w_1, w_2, \dots, w_T
- D chứa $N = T(T+1)/2$ đoạn có độ dài từ 1 tới T
- Các đoạn được sắp xếp theo vị trí của từ bắt đầu $START(i)$; các đoạn có cùng từ bắt đầu được sắp xếp theo vị trí của từ kết thúc $END(i)$
- Với mỗi đoạn i , tìm đoạn j đứng trước nó thể hiện một thực thể mà i tham chiếu tới: $j = y_i$
 - Nếu i không tham chiếu tới đoạn nào, $y_i = \varepsilon$

Tầng đầu vào

- Biểu diễn từ:
 - Kết hợp Glove 300 chiều và Turian et al. (2010)
 - OOV: Véc-tơ 0
- Biểu diễn kí tự dựa trên CNN:
 - Kí tự đầu vào có 8 chiều
 - Các cửa sổ $\{3, 4, 5\}$ kí tự, mỗi cửa sổ có 50 bộ lọc

Biểu diễn kí tự dựa trên CNN



Biểu diễn ngữ cảnh

- Biểu diễn từ được đưa vào hai mạng LSTM
 - LSTM tiến: Thể hiện phụ thuộc của từ hiện tại vào các từ phía trước trong câu
 - LSTM lùi: Thể hiện phụ thuộc của từ hiện tại vào các từ phía sau trong câu
 - Biểu diễn cuối cùng là ghép nối của hai biểu diễn này

Biểu diễn đoạn

- $\mathbf{g}_i = [\mathbf{x}_{\text{START}(i)}^*, \mathbf{x}_{\text{END}(i)}^*, \hat{\mathbf{x}}_i, \Phi(i)]$
- $\mathbf{x}_{\text{START}(i)}^*$: Biểu diễn của từ đầu tiên
- $\mathbf{x}_{\text{END}(i)}^*$: Biểu diễn của từ cuối cùng
- $\hat{\mathbf{x}}_i$: Biểu diễn “*mềm*” của từ chính trong đoạn dựa trên cơ chế chú ý
- $\Phi(i)$: Biểu diễn độ dài của i (số từ trong i)

Biểu diễn “mềm” của từ chính

$$\alpha_t = \mathbf{w}_\alpha \cdot \text{FFNN}_\alpha(\mathbf{x}_t^*)$$
$$a_{i,t} = \frac{\exp(\alpha_t)}{\sum_{k=\text{START}(i)}^{\text{END}(i)} \exp(\alpha_k)}$$
$$\hat{\mathbf{x}}_i = \sum_{t=\text{START}(i)}^{\text{END}(i)} a_{i,t} \cdot \mathbf{x}_t$$

- FFNN_α : Mạng nơ-ron tiến học trọng số chú ý
- \mathbf{w}_α : Trọng số liên kết của FFNN_α
- α_t : Đầu ra của mạng FFNN_α tại thời điểm t

Tính điểm mention

- $s_m(i) = \mathbf{w}_m \cdot \text{FFNN}_m(\mathbf{g}_i)$
- \mathbf{g}_i : Biểu diễn của đoạn i
- FNNN_m : Mạng nơ-ron tiến tính điểm mention
- \mathbf{w}_m : Trọng số liên kết của mạng FNNN_m

Tính điểm tương đồng

- $s_a(i, j) = \mathbf{w}_a \cdot \text{FFNN}_a([\mathbf{g}_i, \mathbf{g}_j, \mathbf{g}_i \circ \mathbf{g}_j, \Phi(i, j)])$
- FNNN_a : Mạng nơ-ron tiến tính điểm tương đồng giữa hai đoạn i và j
- \mathbf{w}_a : Trọng số liên kết của mạng FNNN_a
- $\mathbf{g}_i \circ \mathbf{g}_j$: Phép nhân véc-tơ từng thành phần
- $\Phi(i, j)$: Biểu diễn thông tin người nói và giới tính, và khoảng cách giữa hai đoạn i và j

Hàm lỗi

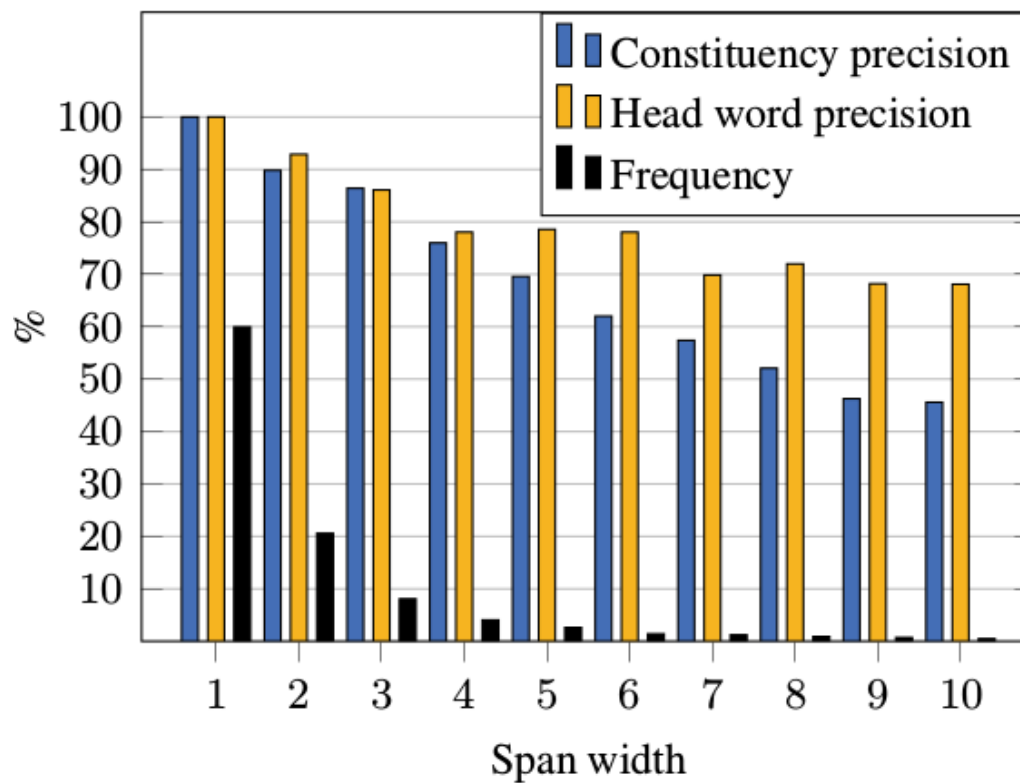
$$\begin{aligned} P(y_1, \dots, y_N | D) &= \prod_{i=1}^N P(y_i | D) \\ &= \prod_{i=1}^N \frac{\exp(s(i, y_i))}{\sum_{y' \in \mathcal{Y}(i)} \exp(s(i, y'))} \end{aligned}$$

- Xác suất biên của các đoạn thể hiện các thực thể
- $s(i, y_i)$: Khả năng i tham chiếu tới y_i

Đánh giá

	Avg. F1	Δ
Our model (ensemble)	69.0	+1.3
Our model (single)	67.7	
– distance and width features	63.9	-3.8
– GloVe embeddings	65.3	-2.4
– speaker and genre metadata	66.3	-1.4
– head-finding attention	66.4	-1.3
– character CNN	66.8	-0.9
– Turian embeddings	66.9	-0.8

Đánh giá (tiếp)



Đánh giá (tiếp)

1 (A **fire in a Bangladeshi garment factory**) has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee (**the blaze**) in the four-story building.

A fire in (**a Bangladeshi garment factory**) has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in (**the four-story building**).

2 We are looking for (**a region of central Italy bordering the Adriatic Sea**). (**The area**) is mostly mountainous and includes Mt. Corno, the highest peak of the Apennines. (**It**) also includes a lot of sheep, good clean-living, healthy sheep, and an Italian entrepreneur has an idea about how to make a little money of them.

3 (**The flight attendants**) have until 6:00 today to ratify labor concessions. (**The pilots'**) union and ground crew did so yesterday.

4 (**Prince Charles and his new wife Camilla**) have jumped across the pond and are touring the United States making (**their**) first stop today in New York. It's Charles' first opportunity to showcase his new wife, but few Americans seem to care. Here's Jeanie Mowth. What a difference two decades make. (**Charles and Diana**) visited a JC Penney's on the prince's last official US tour. Twenty years later here's the prince with his new wife.

5 Also such location devices, (**some ships**) have smoke floats (**they**) can toss out so the man overboard will be able to use smoke signals as a way of trying to, let the rescuer locate (**them**).



25 YEARS ANNIVERSARY
SOICT

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

**Thank you for
your attentions!**



soict.hust.edu.vn/



fb.com/groups/soict

