



NGỮ NGHĨA HỌC TÍNH TOÁN

ĐỒ ÁN 2 – GÁN NHÃN NGHĨA CỦA TỪ

NGUYỄN TRỌNG CHỈNH



TRÌNH BÀY

1. GIỚI THIỆU ĐỒ ÁN
2. HUẤN LUYỆN MÔ HÌNH
3. BIỂU DIỄN TRI THỨC
4. TRUY VẤN TRI THỨC.



1. GIỚI THIỆU ĐỒ ÁN



1. GIỚI THIỆU ĐỒ ÁN

Đồ án gán nhãn nghĩa của từ:

- Biểu diễn tri thức của một đoạn văn.
- Tự động bổ sung tri thức từ WordNet.
- Truy vấn tri thức theo ngôn ngữ bậc một.



1. GIỚI THIỆU ĐỒ ÁN

Ví dụ: Hãy biểu diễn tri thức của đoạn văn bên dưới và hãy truy vấn tri thức theo nội dung các câu:

- John has a mammal.
- Who loves John?
- A dog is intelligent.

Đoạn văn:

John has a puppy. His name is Ace. Ace is white. He has a small bed. He is smart and playful. John loves him every much. And he loves John, too.



1. GIỚI THIỆU ĐỒ ÁN

Kết quả cần có:

- Kết quả xác định nhãn nghĩa tự động cho các danh từ, động từ, tính từ trong đoạn văn.
- Kết quả biểu diễn tri thức của đoạn văn theo logic vị từ bậc một và bổ sung tri thức của đoạn văn từ nghĩa xác định được với WordNet.



1. GIỚI THIỆU ĐỒ ÁN

Phạm vi đồ án:

- Tri thức được mô tả bằng một đoạn văn miêu tả không có yếu tố thời gian.
- Tri thức trong đoạn văn và câu truy vấn được xác định thủ công.
- Tri thức được bổ sung tự động với WordNet.



1. GIỚI THIỆU ĐỒ ÁN

Yêu cầu đồ án:

1. Tự chọn một đoạn văn miêu tả tiếng Anh từ 7 đến 10 câu.
2. Thu thập một số câu truy vấn có nội dung xoay quanh đoạn văn được chọn.
3. Xây dựng tri thức theo đoạn văn theo phương pháp thủ công.
4. Huấn luyện mô hình gán nhãn nghĩa tự động.
5. Sử dụng WordNet để bổ sung tri thức của đoạn văn.



1. GIỚI THIỆU ĐỒ ÁN

Các nội dung cần trình bày:

1. Ngôn ngữ bậc một.
2. Dịch thủ công đoạn văn sang ngôn ngữ bậc một.
3. Phương pháp khử nhập nhằng nghĩa của từ.
4. Phương pháp bổ sung tri thức tự động.
5. Thủ nghiệm và nhận xét.



2. HUẤN LUYỆN MÔ HÌNH



2. HUẤN LUYỆN MÔ HÌNH

Ngữ liệu huấn luyện: có thể chọn kết hợp các ngữ liệu sau

1. Semcor (Miller và cộng sự, 1993)
2. OMSTI (Taghipour và Ng, 2015)

Ngữ liệu đánh giá:

1. Senseval-2 (Edmons và Cotton, 2001)
2. Đoạn văn bản đã thu thập.



2. HUẤN LUYỆN MÔ HÌNH

Mô hình: chọn một trong các mô hình học máy cho bài toán gán nhãn chuỗi (sequential labeling):

Trích xuất đặc trưng:

- Đặc trưng thủ công.
- Đặc trưng từ neural network: LSTM, Bi-LSTM, BERT, ...

Phân lớp:

- SVM
- Neural network: Linear, Feed-forward, ...



2. HUẤN LUYỆN MÔ HÌNH

Cài đặt thuật toán MFS, dùng nghĩa thường gấp để gán nhãn.

Kết quả này được dùng làm baseline.

Đánh giá:

- Accuracy
- Precision
- Recall
- F_1 .

Nhận xét và giải thích kết quả.



2. HUẤN LUYỆN MÔ HÌNH

Các nội dung cần trình bày:

1. Trình bày phương pháp phân lớp được chọn.
2. Tiền xử lý ngũ liệu.
3. Phương pháp rút trích đặc trưng.
4. Thủ nghiệm và đánh giá.
5. Nhận xét.



3. BIỂU DIỄN TRI THỨC



3. BIỂU DIỄN TRI THỨC

Yêu cầu

1. Chọn một đoạn văn miêu tả tiếng Anh từ 7 đến 10 câu.
2. Xác định từ vựng bậc một từ đoạn văn và các câu hỏi.
3. Định nghĩa một vị từ để kiểm tra tính hợp lệ của một biểu thức logic vị từ bậc một theo từ vựng đã xác định ở bước 2.
4. Xác định các vị từ thể hiện nội dung văn bản.
5. Tính chất và quan hệ giữa các vị từ.



3. BIỂU DIỄN TRI THỨC

Các nội dung cần trình bày:

1. Ngôn ngữ bậc một.
2. Dịch thủ công đoạn văn sang ngôn ngữ bậc một.
3. Phương pháp bổ sung tri thức tự động.



4. TRUY VÂN TRI THỨC



4. TRUY VẤN TRI THỨC

Truy vấn tri thức trên Prolog:

1. Dịch truy vấn từ ngôn ngữ tự nhiên sang FOL:

- Dịch thủ công
- Sử dụng hệ luật tính toán ngữ nghĩa theo DRS và biểu thức lambda.

2. Thực thi biểu thức FOL với tri thức đã có.

- Sử dụng cơ chế chứng minh của Prolog.
- Cài đặt cơ chế chứng minh riêng.



4. TRUY VÂN TRI THỨC

Các nội dung cần trình bày:

1. Tập câu hỏi được thử nghiệm kèm câu trả lời.
2. Kết quả truy vấn tri thức.
3. Nhận xét.