

CTT405
Báo cáo cuối kỳ

Combining Natural Logic and Shallow Reasoning
for Question Answering

Vũ Ngọc Minh Hoàng - 1412186 Bùi Quốc Bảo -

January 6, 2017

Mục lục

1	Tổng quan	2
2	Các công trình liên quan	2
3	Phương pháp	2
4	Thực nghiệm và kết quả	2

1 Tổng quan

Bài toán của chúng ta phát biểu như sau:

1. Cho một bộ dữ liệu lớn để làm knowledge base
2. Dữ liệu dùng để đánh giá là các câu hỏi có nhiều lựa chọn
3. Nhiệm vụ của máy là tìm dẫn chứng từ knowledge base để tìm xem lựa chọn nào phù hợp

Có khá nhiều cách để máy xác định lựa chọn, trong đó có phương pháp Combining Natural Logic and Shallow Reasoning sẽ được giới thiệu kĩ ở các phần sau.

Phương pháp được giới thiệu sau đây đóng góp rất lớn cho việc tìm kiếm thông tin hoặc trích xuất thông tin trong một cơ sở tri thức.

Bài toán của chúng ta khá giống với bài thi Reading trong IELTS. Chúng ta sẽ được đưa cho nhiều mệnh đề và phải dựa vào bài đọc để xác định xem mệnh đề đó thuộc nhãn TRUE, FALSE hay NOT GIVEN.

2 Các công trình liên quan

3 Phương pháp

4 Thực nghiệm và kết quả

Đánh giá dựa trên bộ dữ liệu SCITEXT. Gồm các câu hỏi khoa học dành cho học sinh lớp 4. Quá trình train sẽ tạo ra cây quan hệ như ở phần NaturalLI đã giới thiệu bằng các câu trả lời đã được gán nhãn, ở phần test sẽ chỉ đưa các câu hỏi vào cây để tìm ra câu support cho nó.

Đặt vấn đề

Lấy ví dụ về trường hợp sau:

P: Food serves mainly for growth, energy and body repair, maintenance and protection.

H: Animals get energy for growth and repair from food.

Từ 2 câu trên ta thấy được form của chúng như sau:

Food serves mainly for x -> Animals get x from food

Mặc dù có đủ thông tin nhưng NaturalLI sẽ không tìm thấy câu support. Vì lý do đó, chúng ta cần một đánh giá để xác định độ tốt của thuật toán.

Đánh giá độc lập các keyphrase

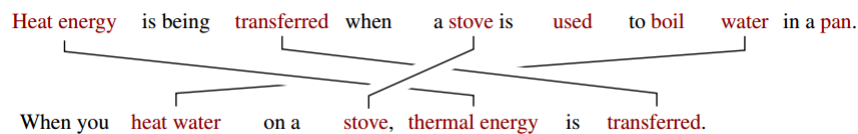
Chúng ta sẽ đánh giá độc lập các keyphrase, phương pháp này gồm 2 bước là xác định keyphrase và match giữa 2 câu. Keyphrase là một cụm từ trong câu, có thể xác định thông qua các cách như sau:

- một danh từ, có thể có tính từ phía trước hoặc được nối giữa of và sở hữu 's .sneaky kitten or pail of water
- một động từ, có thể có trạng từ phía trước .e.g., quietly pounce
- một danh từ có danh động từ phía trước .flowing water
- Động từ to be sẽ không làm một keyphrase

Sau khi đã xác định các keyphrase, ta sẽ match các keyphrase ở 2 câu như sau:

1. Đầu tiên, tất cả các keyphrase giống nhau sẽ được nối với nhau. (stove, transferred)
2. Sau đó, các keyphrase có hậu tố và tiền tố giống nhau sẽ được nối với nhau. (suffix: heat energy – thermal energy)
3. Nếu keyphrase này cũng chứa keyphrase kia thì cũng sẽ được nối (heat water – water)
4. Cuối cùng, các keyphrase bị “kẹp” giữa 2 keyphrase được nối khác thì cũng sẽ được nối.

Ví dụ như trường hợp sau:



Sự tương đồng giữa 2 câu sẽ được tính thông qua một số điểm được tính bằng công cụ Solr. Số điểm này sẽ dựa trên số cặp keyphrase được nối, số cặp keyphrase được match chính xác với nhau và số keyphrase không được nối.

Một đánh giá khác

P: some cats have tails

H: no cats have tails

Ta có thể thấy 2 câu trên có ý đối lập hoàn toàn với nhau, ta cũng sẽ xét sự đối lập của nó. 2 câu được xem là phù hợp với nhau nếu:

- Chúng có cùng “lemmatized” (có cùng gốc từ)
- Chúng giống nhau về “polarity”

System	Train	Test
NaturalLI	65%	61%
NaturalLI cải tiến	74%	67%

Table 1: Kết quả thực nghiệm

Phương pháp đã cải thiện lên hơn 5%. Thông qua phương pháp này, chúng ta nhận ra rằng:

- Áp dụng suy luận vào dependency tree,
- Sử dụng soft evaluation function để tìm ra các lựa chọn gần đúng
- Cuối cùng, chứng tỏ rằng có thể sử dụng relational entailment và meronymy để hỗ trợ cho natural logic.

This document is an example of BibTeX using in bibliography management. Three items are cited: *The L^AT_EX Companion* book [**latexcompanion**], the Einstein journal paper [**einstein**], and the Donald Knuth's website [**knuthwebsite**]. The L^AT_EX related items are [**latexcompanion**, **knuthwebsite**].