

**ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

**NGUYỄN VĂN TÚ**

**HỎI ĐÁP TỰ ĐỘNG SỬ DỤNG NHIỀU NGUỒN  
TRI THỨC**

Chuyên ngành: Hệ thống thông tin

Mã số: 9480104.01

**TÓM TẮT LUẬN ÁN TIẾN SĨ CÔNG NGHỆ  
THÔNG TIN**

**Hà Nội – 2020**

Công trình được hoàn thành tại: Trường Đại học Công nghệ,  
Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học: PGS.TS. Nguyễn Hà Nam

PGS.TS. Lê Anh Cường

Phản biện: .....

.....

Phản biện: .....

.....

Phản biện: .....

.....

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm  
luận án tiến sĩ họp tại

vào hồi            giờ            ngày            tháng            năm

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

# MỞ ĐẦU

## 1. Tính cấp thiết của luận án

Ý tưởng về xây dựng hệ thống hỏi đáp tự động ra đời từ những năm 1960. Điểm chung trong các hệ thống hỏi đáp (Question Answering - QA) giai đoạn này là sử dụng cơ sở dữ liệu được thiết kế bằng tay bởi các chuyên gia trong lĩnh vực được chọn để trích rút câu trả lời. Giai đoạn những năm 1970 - 1980, có nhiều dự án lớn hướng đến việc “hiểu văn bản” và xây dựng hệ thống QA dựa trên các mô hình ngôn ngữ thống kê. Hội nghị TREC (Text REtrieval Conference) diễn ra hàng năm (bắt đầu từ cuối những năm 1990) thu hút sự tham gia của rất nhiều các nhóm nghiên cứu cũng đã góp phần rất lớn trong việc thúc đẩy các nghiên cứu về hệ thống QA. Cuối những năm 1990, World Wide Web (WWW) ra đời và nhanh chóng phát triển bùng nổ trở thành một kho ngữ liệu khổng lồ. Các nhà nghiên cứu về hệ thống QA cũng bắt đầu khai thác web như là một nguồn thông tin hữu ích cho việc tìm kiếm câu trả lời. Các kĩ thuật mới đòi hỏi tốc độ cao, khả năng xử lý lượng dữ liệu web lớn rất được quan tâm. Cùng với thời gian, các câu hỏi của người dùng dành cho các hệ thống QA ngày càng phức tạp, đôi khi người dùng đã không nhận được câu trả lời thích hợp từ các hệ thống QA này. Để giải quyết những khó khăn này, các hệ thống hỏi đáp dựa trên cộng đồng (community Question Answering - cQA) đã được phát triển. Thay vì phải trích rút các câu trả lời từ một kho lưu trữ được xây dựng trước, các cQA sử dụng các câu trả lời từ các chuyên gia và cộng đồng người dùng cQA. Không những vậy, cQA còn cho phép người dùng đánh giá về chất lượng của các câu trả lời cũng như chọn câu trả lời tốt nhất cho mỗi câu hỏi. Hiện nay, một số hệ thống cQA đã được sử dụng rất rộng rãi trên thế giới như StackOverflow<sup>1</sup>, Yahoo!Answers<sup>2</sup>, Quora<sup>3</sup>.

Đối với một hệ thống cQA, chất lượng của hệ thống phụ thuộc vào nhiều yếu tố như: loại câu hỏi mà hệ thống cQA có thể trả lời, thời gian hệ thống trả về các câu trả lời, chất lượng của các câu trả lời, vv. Đối với vấn đề thời gian hệ thống trả về câu trả lời có thể cải thiện thông qua việc tìm kiếm các câu hỏi tương tự trong kho lưu trữ cQA, nếu tìm được các câu hỏi tương tự thì lấy câu

---

<sup>1</sup><https://stackoverflow.com/>

<sup>2</sup><https://answers.yahoo.com/>

<sup>3</sup><https://www.quora.com/>

trả lời của các câu hỏi này làm câu trả lời cho câu hỏi mới. Tuy nhiên vấn đề này mới chỉ được giải quyết dựa trên thông tin cung cấp bởi các cặp câu hỏi. Vì vậy, nghiên cứu để có thể xây dựng các mô hình mới tích hợp được cả các thông tin của các câu hỏi với các nguồn thông tin mới để tìm kiếm các câu hỏi tương tự trong kho lưu trữ cQA là rất cần thiết. Đối với vấn đề đánh giá chất lượng của các câu trả lời, các nghiên cứu trước đây mới chỉ đánh giá dựa trên những thông tin đã có trên câu hỏi và câu trả lời hoặc các thông tin cung cấp bởi người dùng (số lượng votes). Vì vậy, các nghiên cứu nhằm khai thác thêm thông tin từ những nguồn tri thức mới và phát triển các phương pháp hiệu quả hơn dựa trên các nguồn tri thức này để có thể đánh giá tốt hơn chất lượng của các câu trả lời là vấn đề cấp thiết trong xây dựng cQA. Với mong muốn nâng cao hiệu quả của các hệ thống cQA, chúng tôi đã chọn đề tài luận án với tiêu đề: “Hỏi đáp tự động sử dụng nhiều nguồn tri thức”.

## 2. Mục tiêu của luận án

Mục tiêu của luận án là đề xuất các phương pháp dựa trên học máy và tích hợp nhiều nguồn tri thức để nâng cao chất lượng của các hệ thống cQA. Để thực hiện được mục tiêu này, luận án đi vào giải quyết các bài toán sau:

- Thứ nhất: Nghiên cứu, đề xuất một số phương pháp để tìm kiếm và xếp hạng các câu hỏi trong cơ sở dữ liệu liên quan đến câu hỏi mới.
- Thứ hai: Nghiên cứu, đề xuất một số phương pháp để đánh giá độ tin cậy của các câu trả lời trong hệ thống cQA.
- Thứ ba: Nghiên cứu, đề xuất một số phương pháp để tích hợp thêm nguồn tri thức vào mô hình học sâu để đánh giá độ tương tự giữa các câu hỏi.

## 3. Đóng góp của luận án

Từ các mục tiêu và nội dung chính đã đặt ra, luận án đã tiến hành thực hiện và hoàn thành. Những đóng góp chính của luận án bao gồm:

- Đề xuất một số phương pháp để tìm kiếm và xếp hạng các câu hỏi trong cơ sở dữ liệu liên quan đến câu hỏi mới. Đối với việc xác định độ tương tự giữa các câu hỏi, luận án đề xuất phương pháp mới để xem xét nhiều khía cạnh khác nhau để hiểu câu hỏi. Luận án sử dụng nhiều hơn các khía cạnh hiệu quả bao gồm word embedding và các loại câu hỏi (question category). Luận án sử dụng mô hình biểu diễn từ để tạo các word embedding, mô-đun phân loại câu hỏi để xác định loại câu hỏi mới. Sau đó, tất cả các đặc trưng thu được này được kết hợp lại và sử dụng làm đầu vào cho một mô hình học máy để nhận được điểm số tương tự của các cặp câu hỏi và sử dụng điểm số này

để xếp hạng các câu hỏi. Đóng góp này đã được chúng tôi công bố ở *Tạp chí khoa học công nghệ thông tin và truyền thông, học viện Công nghệ Bưu chính Viễn thông* năm 2017 (công trình số 4); kỷ yếu hội thảo quốc tế *Information system Design and Intelligent Applications (INDIA)* năm 2017 (công trình số 5).

- Đề xuất một số phương pháp để đánh giá độ tin cậy của các câu trả lời. Luận án sử dụng thêm nguồn thông tin từ Wikipedia để đánh giá độ tin cậy của câu trả lời. Luận án xây dựng mô hình mới kết hợp cả các thông tin từ câu hỏi và câu trả lời, đồng thời khai thác thêm thông tin từ nguồn tri thức bên ngoài (wikipedia) để đánh giá độ tin cậy của các câu trả lời trong cQA. Đóng góp này đã được công bố ở kỷ yếu hội thảo quốc tế *Integrated Uncertainty in Knowledge Modelling and Decision Making (IUKM)* năm 2016 (công trình số 3).
- Đề xuất áp dụng phương pháp học sâu để nâng cao hiệu quả cho bài toán đánh giá độ tương tự giữa hai đoạn văn bản ngắn (short text) với tập dữ liệu nhỏ. Luận án xây dựng mô hình tích hợp nguồn tri thức bên ngoài vào các mô hình học sâu để đánh giá độ tương tự giữa hai câu hỏi (coi mỗi câu hỏi như là một đoạn văn bản ngắn). Luận án xây dựng các mô hình dựa trên mạng nơ-ron để trích rút các đặc trưng thông qua các tầng (layer) của mạng và sau đó tích hợp các nguồn tri thức bên ngoài vào các mô hình học sâu này để tăng hiệu suất của việc đánh giá độ tương tự. Đóng góp này đã được công bố ở kỷ yếu hội thảo quốc tế *Integrated Uncertainty in Knowledge Modelling and Decision Making (IUKM)* năm 2018 (công trình số 6) và tạp chí quốc tế *International Journal of Machine Learning and Computing* năm 2020 (công trình số 7).

Các nội dung và kết quả nghiên cứu trình bày trong luận án (từ Chương 2 đến Chương 4) đã được công bố trong 07 công trình. Trong đó có 02 bài báo đăng ở tạp chí trong nước có phản biện, 02 bài báo đăng ở tạp chí quốc tế và 03 bài báo đăng trong kỷ yếu của hội nghị quốc tế có phản biện, được xuất bản bởi nhà xuất bản Springer.

# Chương 1

## TỔNG QUAN

Chương này trình bày tổng quan về những vấn đề nghiên cứu của luận án, bao gồm: tổng quan về hệ thống hỏi đáp, phân loại các hệ thống hỏi đáp, một số hệ thống cQA thông dụng cũng như phân tích một số nghiên cứu liên quan về cQA. Phần cuối chương sẽ thảo luận về một số vấn đề còn tồn tại khi xây dựng các hệ thống cQA mà luận án sẽ tập trung giải quyết và xác định nội dung nghiên cứu của luận án.

### 1.1 Tổng quan về hệ thống hỏi đáp

### 1.2 Phân loại các hệ thống hỏi đáp

### 1.3 Một số cQA thông dụng

#### 1.3.1 Yahoo!Answer

#### 1.3.2 StackOverflow

#### 1.3.3 Quora

#### 1.3.4 Một số cQA tiếng Việt

### 1.4 Tình hình nghiên cứu về cQA

#### 1.4.1 Các nghiên cứu liên quan về tìm kiếm và xếp hạng câu hỏi

#### 1.4.2 Các nghiên cứu liên quan về đánh giá độ tin cậy của câu trả lời

### 1.5 Thảo luận

Các hệ thống cQA đang trở thành nguồn thông tin ngày càng quan trọng, nơi người dùng có thể chia sẻ kiến thức về các chủ đề khác nhau thông qua việc hỏi

và trả lời/bình luận các câu hỏi. Mặc dù nền tảng này mang đến cơ hội mới cho người dùng tìm kiếm trợ giúp hoặc cung cấp giải pháp, nhưng chúng cũng đặt ra nhiều thách thức khi quy mô ngày càng tăng của cộng đồng người dùng, các câu hỏi đa dạng và phức tạp. Nói chung, chất lượng của một hệ thống cQA phụ thuộc vào nhiều yếu tố khác nhau như: loại câu hỏi mà hệ thống cQA có thể trả lời, thời gian hệ thống trả về các câu trả lời, chất lượng của các câu trả lời, số lượng của các câu trả lời cho mỗi câu hỏi, vv. Tuy nhiên các nghiên cứu giải quyết các nhiệm vụ này hiện nay vẫn chưa cho hiệu quả cao. Từ những phân tích, đánh giá các khó khăn khi xây dựng hệ thống cQA ở phần mở đầu cũng như phân tích các nghiên cứu liên quan, chúng tôi nhận thấy một số vấn đề còn tồn tại khi xây dựng các hệ thống cQA, cụ thể như sau:

Thứ nhất, đối với bài toán tìm kiếm và xếp hạng các câu hỏi trong cơ sở dữ liệu liên quan đến câu hỏi mới. Các nghiên cứu trước đó đã xây dựng nhiều mô hình khác nhau dựa trên các thông tin cung cấp bởi câu hỏi nhằm đánh giá được độ tương tự giữa câu hỏi mới với các câu hỏi trong kho lưu trữ của hệ thống cQA. Việc đánh giá độ tương tự này có thể sử dụng thêm nhiều phương pháp biểu diễn khác nhau như biểu diễn dựa trên tập nhúng từ cũng như bổ sung thêm thông tin về loại câu hỏi. Đồng thời cũng có thể dựa vào câu trả lời của câu hỏi trong cơ sở dữ liệu để xác định độ tương tự của câu hỏi mới với các câu hỏi trong cơ sở dữ liệu. Ngoài ra, trong vài năm gần đây các mô hình học sâu đã được áp dụng và cho thấy nhiều thành công trong lĩnh vực xử lý ngôn ngữ tự nhiên như: phân tích ngữ nghĩa, tìm kiếm thông tin. Trong nghiên cứu này luận án sẽ xây dựng các mô hình học sâu cho vấn đề xác định độ tương tự giữa hai câu hỏi trong cQA. Nghiên cứu này sẽ xây dựng các mô hình học sâu dựa trên CNN và LSTM để đánh giá độ tương tự giữa hai câu hỏi. Sau đó, luận án nghiên cứu sử dụng thêm một số nguồn thông tin mới và tích hợp thêm các nguồn thông tin mới này vào các mô hình học sâu để đạt được kết quả tốt nhất.

Thứ hai, đối với bài toán đánh giá độ tin cậy của các câu trả lời trong các hệ thống cQA. Để có được một hệ thống cQA tốt thì yêu cầu bắt buộc là mỗi câu hỏi phải nhận được câu trả lời có độ chính xác cao. Để đánh giá chất lượng của các câu trả lời, các nghiên cứu trước đây thường dựa trên việc đánh giá độ tương tự giữa câu hỏi và câu trả lời hoặc dựa trên thông tin cung cấp bởi người dùng. Câu trả lời có độ tương tự với câu hỏi lớn hơn thì được coi là câu trả lời tốt. Tuy nhiên, việc đánh giá như vậy chỉ mới xác định được sự giống nhau giữa câu hỏi và câu trả lời, chưa xác định được đó có phải là câu trả lời đáng tin tưởng hay không. Để đánh giá chất lượng của các câu trả lời luận án không chỉ sử dụng thông tin về độ tương tự giữa câu hỏi và câu trả lời, thông tin cung cấp bởi người dùng mà còn nghiên cứu sử dụng thêm nguồn thông tin mới bên ngoài thông tin hỏi - đáp. Wikipedia là nguồn thông tin hữu ích được nghiên cứu sử dụng như thông tin bổ sung để đánh giá độ tin cậy của các câu trả lời trong các hệ thống cQA.

## Chương 2

# TÌM VÀ XẾP HẠNG CÁC CÂU HỎI LIÊN QUAN

### 2.1 Giới thiệu

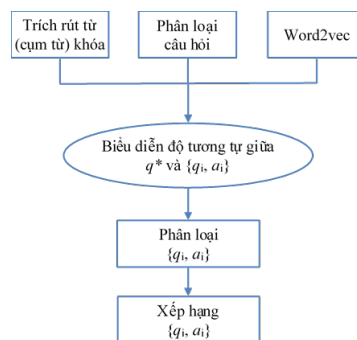
### 2.2 Mô tả bài toán

Cho câu hỏi mới  $q^*$  và một tập  $D$  các cặp câu hỏi-câu trả lời trong cơ sở dữ liệu. Các câu hỏi, câu trả lời được trình bày dưới dạng ngôn ngữ tự nhiên. Tập  $D$  gồm các cặp câu hỏi-câu trả lời được lưu trữ trong các hệ thống cQA, kí hiệu  $D = \{q_i, a_i\}$ . Cần tìm trong tập  $D$  các câu hỏi tương tự (tương tự về mặt ngữ nghĩa) với câu hỏi mới  $q^*$ , sau đó xếp hạng các câu hỏi tương tự này theo độ tương tự của chúng với câu hỏi mới  $q^*$ .

### 2.3 Các phương pháp tiếp cận và đề xuất

#### 2.3.1 Kiến trúc mô hình đề xuất

Mô hình tìm kiếm và xếp hạng các câu hỏi trong cQA được minh họa như Hình 2.1.



**Hình 2.1:** Kiến trúc của mô hình tìm kiếm và xếp hạng các câu hỏi



## 2.3.2 Trích rút từ (cụm từ) khóa

### 2.3.3 Biểu diễn độ tương tự giữa các câu hỏi, giữa câu hỏi và câu trả lời

Biểu diễn dựa trên mô hình n-gram

Biểu diễn dựa trên tính chất của câu hỏi và câu trả lời

Biểu diễn dựa trên mô hình word2vec

Các mô hình word2vec và các ứng dụng của nó gần đây đã thu hút rất nhiều sự chú ý từ cộng đồng học máy. Sự biểu diễn véc-tơ dày đặc của các từ đã học được từ word2vec đã được chứng minh là có ý nghĩa ngữ nghĩa và rất hữu ích trong nhiều trường hợp sử dụng, từ xử lý ngôn ngữ tự nhiên đến phân tích luồng dữ liệu mạng. Cách biểu diễn véc-tơ của các từ được học bởi word2vec đặc biệt hữu ích trong trường hợp dữ liệu thưa như dữ liệu trong cQA.

Trong nghiên cứu này, mô hình continuous Skip-gram của bộ công cụ word2vec<sup>1</sup> được sử dụng để sinh ra sự biểu diễn véc-tơ của các từ trong cơ sở dữ liệu. Đầu tiên, tất cả các câu trong câu hỏi mới và trong các câu hỏi, câu trả lời trong cơ sở dữ liệu được phân tích thành các từ tố và các từ được chuyển thành các véc-tơ từ sử dụng mô hình huấn luyện lại word2vec. Để xây dựng các véc-tơ câu hỏi, véc-tơ câu trả lời từ các véc-tơ từ này nghiên cứu sử dụng hai phương pháp khác nhau.

Biểu diễn dựa trên loại câu hỏi

Nghiên cứu này cũng sử dụng mô hình biểu diễn véc-tơ từ để mô hình hóa mối quan hệ ngữ nghĩa giữa loại của câu hỏi mới với loại của câu hỏi trong cơ sở dữ liệu. Loại của câu hỏi mới đạt được bằng cách sử dụng mô-đun phân loại câu hỏi. Tập dữ liệu  $D$  bao gồm các câu hỏi được trích xuất từ các hệ thống cQA, trong đó mỗi câu hỏi trong  $D$  được gán một nhãn loại. Mô-đun phân loại câu hỏi nhằm phân loại câu hỏi mới  $q^*$  vào một trong các loại của các câu hỏi trong tập dữ liệu  $D$ .

- Bước 1: chuẩn bị tập dữ liệu huấn luyện bao gồm các câu hỏi trong tập dữ liệu  $D$ , chúng đã được gán nhãn loại (nhãn ở đây là loại câu hỏi).
- Bước 2: các câu hỏi trong tập huấn luyện được biểu diễn dưới dạng véc-tơ của các đặc trưng.
- Bước 3: một phương pháp học máy được sử dụng (ở đây sử dụng thuật toán phân loại SVM) để xây dựng mô hình phân loại.

<sup>1</sup><https://code.google.com/p/word2vec>

### 2.3.4 Các bộ phân loại

### 2.3.5 Xếp hạng câu hỏi

## 2.4 Thực nghiệm

### 2.4.1 Cài đặt thực nghiệm

Các thực nghiệm sử dụng tập dữ liệu cQA cung cấp bởi SemEval 2016 task 3 subtask B<sup>2</sup>. Tập dữ liệu bao gồm 337 câu hỏi mới, 3369 câu hỏi liên quan cùng với 33690 câu trả lời của nó. Tập dữ liệu được phân chia thành 267 câu hỏi mới và 2669 câu hỏi liên quan được sử dụng như là tập dữ liệu huấn luyện, cũng như 70 câu hỏi mới và 700 câu hỏi liên quan được sử dụng như là tập dữ liệu kiểm tra. Mỗi điểm dữ liệu là một cặp câu hỏi (câu hỏi mới và câu hỏi liên quan) và một nhãn tương tự, đó là “relevant” hoặc “irrelevant”. Chúng tôi cần dự đoán nhãn nhị phân trong đó 1 thuộc về lớp “relevant”, 0 thuộc về lớp “irrelevant” và xếp hạng một tập hợp các câu hỏi có liên quan theo độ đo sự giống nhau của chúng đối với câu hỏi mới.

### 2.4.2 Độ đo đánh giá

### 2.4.3 Kết quả thực nghiệm

**Thực nghiệm 1:** Thực nghiệm này kiểm tra hiệu suất của mô hình khi biểu diễn độ tương tự giữa các câu hỏi sử dụng mô hình n-gram.

**Bảng 2.1:** Các kết quả phân loại và xếp hạng của thực nghiệm 1

N-gram từ	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F1	MAP	AveRec	MRR
1-gram (Unigram)	62,43	42,02	33,91	37,53	55,98	76,22	61,85
2-gram (Bigram)	61,43	39,66	30,47	34,47	53,33	72,72	57,07
3-gram (Trigram)	64,29	44,97	32,62	37,81	53,69	73,31	56,49

**Thực nghiệm 2:** Thực nghiệm này nhằm kiểm tra hiệu suất của mô hình khi biểu diễn độ tương tự giữa câu hỏi mới với các câu hỏi trong cơ sở dữ liệu dựa trên tính chất của câu hỏi và câu trả lời. Thực nghiệm này sử dụng cả hai bộ phân loại SVM và MLP.

**Thực nghiệm 3:** Thực nghiệm này sử dụng những đặc trưng như trong thực nghiệm 2 nhưng bổ sung thêm các đặc trưng mới dựa trên mô hình biểu diễn véc-tơ từ (sử dụng công cụ word2vec) để biểu diễn câu hỏi và câu trả lời.

**Thực nghiệm 4:** Thực nghiệm này kiểm tra tính hiệu quả của mô hình khi bổ sung thêm các đặc trưng trích rút từ loại câu hỏi.

<sup>2</sup><http://alt.qcri.org/semeval2016/task3/index.php?id=data-and-tools>

**Bảng 2.2:** Các kết quả phân loại và xếp hạng của thực nghiệm 2

Biểu diễn độ tương tự dựa trên	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F1	MAP	AveRec	MRR
Tính chất của câu hỏi (QP), Tính chất của câu trả lời (AP)	78,86 (79,14)	67,63 (65,67)	69,96 (69,86)	68,78 (67,70)	77,53 (75,53)	90,86 (90,12)	83,98 (81,57)

**Bảng 2.3:** Các kết quả phân loại và xếp hạng của thực nghiệm 3

Biểu diễn độ tương tự dựa trên	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F1	MAP	AveRec	MRR
QP, AP, Biểu diễn véc-tơ từ (WVR) (Phương pháp 1)	80,71 (80,86)	70,59 (67,38)	72,10 (73,02)	71,34 (70,09)	78,21 (75,91)	92,12 (90,73)	85,64 (81,69)
QP, AP, Biểu diễn véc-tơ từ (WVR) (Phương pháp 2)	81,57 (81,86)	71,49 (69,53)	74,25 (74,31)	72,84 (71,84)	77,64 (77,99)	91,72 (91,68)	84,93 (86,23)

#### 2.4.4 Đánh giá các kết quả thực nghiệm

Kết quả của thực nghiệm 1 cho thấy khi chỉ sử dụng mô hình n-gram để biểu diễn mối quan hệ giữa câu hỏi mới với các câu hỏi  $q_i$  cho kết quả phân loại và xếp hạng không cao. Thực nghiệm này chỉ đạt kết quả cao nhất với các độ đo *Accuracy* và *MAP* lần lượt là 64,29% và 55,98%. Trong thực nghiệm 2, chúng tôi sử dụng sự biểu diễn dựa trên tính chất của các câu hỏi và câu trả lời, sử dụng cả hai bộ phân loại SVM và MLP. Chúng tôi đã đạt được kết quả phân loại và xếp hạng cao với các độ đo *Accuracy* tăng 14,85% và *MAP* tăng 21,55% so với khi chỉ sử dụng mô hình biểu diễn n-gram. Thực nghiệm 3 sử dụng thêm sự biểu diễn véc-tơ từ của cả câu hỏi và câu trả lời. Trong cách biểu diễn này, chúng tôi sử dụng hai phương pháp khác nhau để xây dựng véc-tơ câu hỏi, véc-tơ câu trả lời từ sự biểu diễn véc-tơ từ. Trong thực nghiệm này chúng tôi thấy rằng cách xây dựng các véc-tơ tổng hợp cho các câu hỏi, câu trả lời sử dụng phương pháp 2 cho các kết quả phân loại và xếp hạng cao hơn phương pháp 1. Đồng thời các kết quả của thực nghiệm này cũng tăng lên đáng kể so với khi chỉ sử dụng sự biểu diễn dựa trên tính chất của câu hỏi và câu trả lời như trong thực nghiệm 2. Các độ đo *Accuracy* và *MAP* tăng lần lượt là 2,71% và 0,46%. Thực nghiệm 4 thực hiện như trong thực nghiệm 3 nhưng có bổ sung thêm sự biểu diễn độ tương tự giữa câu hỏi mới với các câu hỏi  $q_i$  dựa trên loại câu hỏi. Các kết quả của thực nghiệm này cho thấy mô hình đề xuất của chúng tôi (sử dụng mô đun phân loại câu hỏi) cho kết quả cao hơn, cả ở các độ đo phân

**Bảng 2.4:** Các kết quả phân loại và xếp hạng của thực nghiệm 4

Biểu diễn độ tương tự dựa trên	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F1	MAP	AveRec	MRR
QP, AP, WVR (Phương pháp 1), loại câu hỏi(QC)	80,86 (81,57)	72,20 (69,10)	69,10 (73,85)	70,61 (71,40)	78,27 (76,45)	92,14 (91,08)	85,64 (84,37)
QP, AP, WVR (Phương pháp 2), loại câu hỏi(QC)	81,86 (82,29)	73,25 (70,39)	71,67 (74,89)	72,45 (72,57)	77,75 (78,35)	91,81 (91,93)	85,64 (86,23)

loại và xếp hạng. Kết quả của thực nghiệm 4 cho thấy các độ đo *Accuracy* và *MAP* tăng tương ứng là 0,43% và 0,14% so với các kết quả của thực nghiệm 3.

## 2.5 Kết luận chương

Chương này đã trình bày những nội dung, kết quả nghiên cứu về bài toán tìm kiếm và xếp hạng các câu hỏi trong cơ sở dữ liệu liên quan đến câu hỏi mới trong các cQA. Các phương pháp đề xuất trong chương này có những ưu điểm sau: (1) Nghiên cứu đã sử dụng một cách hiệu quả hơn mô hình biểu diễn véc-tơ từ (ở đây là word2vec) để biểu diễn véc-tơ tổng hợp cho câu hỏi và câu trả lời, từ đó tính toán được chính xác hơn độ tương tự giữa câu hỏi mới với các câu hỏi trong cơ sở dữ liệu. (2) Việc sử dụng mô-đun phân loại câu hỏi đã bổ sung thêm thông tin để có thể biểu diễn tốt hơn độ tương tự giữa câu hỏi mới với các câu hỏi trong cơ sở dữ liệu.

## Chương 3

# ĐÁNH GIÁ ĐỘ TIN CẬY CỦA CÂU TRẢ LỜI

### 3.1 Giới thiệu

### 3.2 Mô tả bài toán

Đánh giá độ tin cậy của các câu trả lời trong các hệ thống cQA có thể được xem xét như là một vấn đề phân loại. Đưa ra một tập  $Q$  các câu hỏi, ở đó mỗi câu hỏi  $q_i \in Q$  có một tập các câu trả lời ứng viên  $\{a_{i1}, a_{i2}, \dots, a_{in}\}$  ( $n = 1, 2, \dots$ ). Việc đánh giá độ tin cậy của các câu trả lời cho câu hỏi  $q_i$  chính là gán nhãn cho các câu trả lời  $\{a_{i1}, a_{i2}, \dots, a_{in}\}$  các nhãn tương ứng là  $\{l_{i1}, l_{i2}, \dots, l_{in}\}$  trong đó  $l_{ij} = 1$  nếu  $a_{ij}$  là câu trả lời đúng cho câu hỏi  $q_i$ , ngược lại  $l_{ij} = 0$ .

### 3.3 Các đặc trưng

Có nhiều loại đặc trưng khác nhau được trích rút để phục vụ cho việc đánh giá độ tin cậy của các câu trả lời trong cQA. Phần này trình bày việc trích rút các đặc trưng này.

#### 3.3.1 Các đặc trưng n-gram

#### 3.3.2 Các đặc trưng dựa trên thuộc tính của câu hỏi

Số từ trong câu hỏi và câu trả lời, số lượng các sentence trong câu trả lời, chồng chéo từ giữa câu hỏi và câu trả lời, độ tương tự cosin giữa câu hỏi và câu trả lời.

### 3.3.3 Đặc trưng dựa trên thông tin người dùng

### 3.3.4 Đặc trưng dựa trên tập nhúng từ (Word Embedding)

Sử dụng sự biểu diễn véc-tơ từ để mô hình hóa mối quan hệ giữa câu hỏi và các câu trả lời của nó.

## 3.4 Sử dụng wikipedia như nguồn tài nguyên bên ngoài

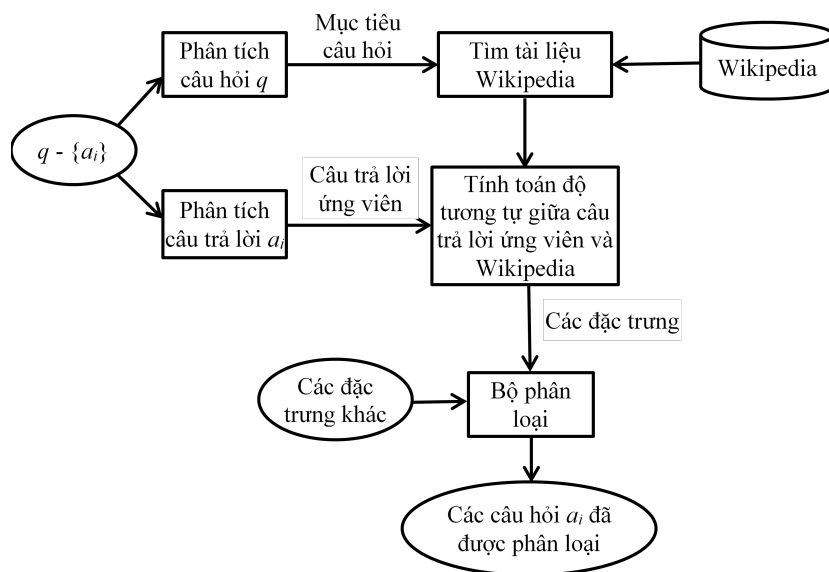
### 3.4.1 Wikipedia

Wikipedia - bách khoa toàn thư mở được chọn như là nguồn tài nguyên bên ngoài uy tín nhất dùng để đánh giá độ tin cậy của các câu trả lời. Các nội dung trên các trang Wikipedia rất phong phú, đa dạng và thường xuyên được cập nhật sẽ là nguồn tài nguyên có độ chính xác cao, chúng có thể bổ sung thông tin trong việc đánh giá chất lượng của các câu trả lời trong các cQA.

### 3.4.2 Sử dụng Wikipedia cho đánh giá độ tin cậy của câu trả lời

#### Kiến trúc hệ thống

Kiến trúc mô hình của hệ thống đánh giá độ tin cậy của câu trả lời được mô tả như trong Hình 3.1. Mô hình thực hiện phân tích câu hỏi để tìm ra mục tiêu (tên thực thể) mà câu hỏi muốn hỏi, phân tích câu trả lời ứng viên và sử dụng nguồn tài nguyên từ Wikipedia để đánh giá độ tin cậy của các câu trả lời ứng viên này.



Hình 3.1: Kiến trúc của hệ thống đánh giá câu trả lời

### Phân tích câu hỏi

Mô-đun này là một thành phần quan trọng trong hệ thống đánh giá độ tin cậy của các câu trả lời. Kết quả của mô-đun này là việc xác định thông tin lòng cốt (mục tiêu câu hỏi) mà người hỏi muốn hỏi hay chính là chủ đề của câu hỏi (tên thực thể). Đồng thời từ việc phân tích câu hỏi này cũng sẽ xác định được kiểu của câu trả lời cho câu hỏi.

### Tìm kiếm các tài liệu Wikipedia liên quan

Mô-đun này thực hiện tìm kiếm các tài liệu Wikipedia mà có liên quan đến câu hỏi từ một tập lớn các tài liệu Wikipedia cho trước. Dựa trên mục tiêu của câu hỏi để tìm các tài liệu Wikipedia mà tiêu đề của chúng so khớp được với mục tiêu này. Khi một tài liệu Wikipedia được tìm thấy, nó sẽ được sử dụng trong các mô-đun tiếp theo để đánh giá sự tin cậy của các câu trả lời

### Phân tích các câu trả lời

Mỗi câu trả lời trong cQA có thể gồm nhiều sentence. Mục đích của mô-đun này là phân tích các câu trả lời ứng viên để trích rút ra những sentence tốt nhất tương ứng cho câu hỏi từ một tập của các câu trả lời ứng viên

### Trích rút các đặc trưng từ Wikipedia

Mô-đun này thực hiện trích rút các đặc trưng dựa trên nguồn thông tin mở rộng wikipedia bằng cách tính toán độ tương tự giữa các tài liệu Wikipedia tìm được với các sentence trả lời cho câu hỏi

## 3.5 Thực nghiệm

### 3.5.1 Tập dữ liệu

Dữ liệu thực nghiệm được thu thập được từ các trang cQA Yahoo!Answer và StackOverflow. Bảng 3.1 là một số thống kê trên tập dữ liệu được sử dụng.

**Bảng 3.1:** Một số thống kê về tập dữ liệu

	Câu hỏi - câu trả lời	Tỉ lệ câu trả lời / câu hỏi	TB số từ	TB số kí tự
Dữ liệu huấn luyện	138 - 523	3,79	8,56	43,41
Dữ liệu kiểm tra	34 - 130	3,82	8,22	42,88

### 3.5.2 Kết quả thực nghiệm

**Thực nghiệm 1:** thực nghiệm này sử dụng các đặc trưng: unigram, bigram, số từ trong câu hỏi, số từ trong câu trả lời, số sentence trong câu trả lời, số lượng bình chọn cho câu trả lời, đặc trưng so khớp từ, đặc trưng dựa trên sự biểu diễn véc-tơ từ. Bảng 3.2 chỉ ra các kết quả của thực nghiệm này.

**Bảng 3.2:** Độ chính xác của bộ phân loại SVM với sự kết hợp của nhiều loại đặc trưng

Đặc trưng sử dụng	Acc	P	R	F1
Unigram, Bigram, số từ trong câu hỏi, số từ trong câu trả lời, số sentence trong câu trả lời, số lượng bình chọn của câu trả lời	76,15	53,06	76,47	62,65
Unigram, Bigram, số từ trong câu hỏi, số từ trong câu trả lời, số sentence trong câu trả lời, số lượng bình chọn của câu trả lời, đặc trưng so khớp từ, cosin giữa $q$ và $a$	78,46	55,10	81,82	65,85
Unigram, Bigram, số từ trong câu hỏi, số từ trong câu trả lời, số sentence trong câu trả lời, số lượng bình chọn của câu trả lời, đặc trưng dựa trên sự biểu diễn véc-tơ từ	76,92	55,10	77,14	64,29
Sử dụng tất cả các đặc trưng trên	79,23	55,10	84,38	66,67

**Thực nghiệm 2:** thực nghiệm này kiểm tra đề xuất sử dụng thêm nguồn tài nguyên bên ngoài để đánh giá sự tin cậy của các câu trả lời. Các kết quả đạt được được trình bày trong Bảng 3.3.

**Bảng 3.3:** Độ chính xác của bộ phân loại SVM khi thêm các đặc trưng từ Wikipedia

Đặc trưng sử dụng	Acc	P	R	F1
Tất cả các đặc trưng trong thực nghiệm 1, các đặc trưng trích rút từ tài liệu Wikipedia (Phương pháp 1: dựa trên so khớp từ)	83,08	57,14	96,55	71,79
Tất cả các đặc trưng trong thực nghiệm 1, các đặc trưng trích rút từ tài liệu Wikipedia (Phương pháp 2: dựa trên sự biểu diễn véc-tơ từ)	84,62	61,22	96,77	75,00

### 3.5.3 Đánh giá các kết quả thực nghiệm

Các kết quả thực nghiệm đã được trình bày trong các Bảng 3.2 và Bảng 3.3. Trong Bảng 3.2, khi sử dụng tất cả các đặc trưng trích rút từ câu hỏi và câu trả lời (Unigram, Bigram, số từ trong câu hỏi, số từ trong câu trả lời, số sentence trong câu trả lời, số lượng bình chọn của câu trả lời, đặc trưng so khớp từ đặc trưng dựa trên sự biểu diễn véc-tơ từ) kết quả đạt được độ chính xác phân loại (*accuracy*) là 79,23%. Trong thực nghiệm thứ 2, khi bổ sung thêm các đặc



trung trích rút từ các tài liệu Wikipedia kết quả đã đạt được các kết quả phân loại tốt hơn như trong Bảng 3.3. Khi đánh giá độ tương tự giữa các câu trả lời và tài liệu Wikipedia, sử dụng hai phương pháp là so khớp từ và dựa trên biểu diễn véc-tơ từ kết quả cho thấy phương pháp dựa trên sự biểu diễn véc-tơ từ cho kết quả phân loại cao hơn với độ chính xác (*accuracy*) đạt được cao nhất là 84,62%. Từ các kết quả đạt được của thực nghiệm 1 và thực nghiệm 2 chỉ ra rằng bằng cách sử dụng thêm nguồn tài nguyên bên ngoài (trong nghiên cứu này là Wikipedia) đạt được độ chính xác cao hơn.

### 3.6 Kết luận chương

Chương này đã trình bày những nội dung, kết quả nghiên cứu về bài toán đánh giá độ tin cậy của các câu trả lời trong các cQA. Nghiên cứu đã đề xuất sử dụng thêm nguồn tài nguyên bên ngoài trong việc đánh giá độ tin cậy của các câu trả lời. Các phương pháp đề xuất trong chương này có những ưu điểm sau: (1) Chúng tôi đã trích rút nhiều loại đặc trưng khác nhau bao gồm các đặc trưng dựa trên tính chất của câu hỏi và câu trả lời, các đặc trưng dựa trên thông tin của người sử dụng. Chúng tôi cũng đã sử dụng mô hình biểu diễn véc-tơ từ (word2vec) để biểu diễn tốt hơn mối quan hệ giữa câu hỏi và câu trả lời. (2) Với việc sử dụng thêm nguồn tài nguyên bên ngoài, chúng tôi đã có thêm nguồn thông tin tin cậy để giúp xác định được tốt hơn chất lượng của các câu trả lời. Chúng tôi đã trích rút các đặc trưng dựa trên nguồn dữ liệu mở rộng là các tài liệu Wikipedia, các đặc trưng này sau đó được kết hợp với các đặc trưng khác và sử dụng làm véc-tơ đầu vào cho bộ phân loại SVM.

## Chương 4

# TÍCH HỢP NHIỀU NGUỒN TRI THỨC TRONG MÔ HÌNH HỌC SÂU ĐỂ ĐÁNH GIÁ ĐỘ TƯƠNG TỰ

### 4.1 Giới thiệu

### 4.2 Mô tả bài toán

### 4.3 Các phương pháp tiếp cận và đề xuất

Phần này của luận án sẽ trình bày các phương pháp tiếp cận trong việc giải quyết vấn đề xác định điểm số tương tự giữa các câu hỏi trong cQA sau đó dựa vào điểm số tương tự này để xếp hạng các câu hỏi. Luận án xây dựng các mô hình khác nhau bao gồm: mô hình dựa trên mạng nơ-ron sâu sử dụng CNN, mô hình dựa trên mạng nơ-ron sâu sử dụng BLSTM, mô hình dựa trên mạng nơ-ron sâu tích hợp thêm nguồn tri thức mở rộng khác.

#### 4.3.1 Nguồn tri thức mở rộng (External Knowledge - EK)

Các đặc trưng thông dụng

Trích rút các đặc từ thông tin câu hỏi và câu trả lời.

Từ để hỏi

Sử dụng từ để hỏi ("*what*", "*who*", "*when*", "*how*", "*why*", "*which*", "*where*") trong mỗi câu hỏi như là một đặt trưng.

### Loại câu hỏi (Question Category)

Sử dụng mô hình biểu diễn véc-tơ từ (word2vec) để mô hình hóa mối quan hệ ngữ nghĩa giữa loại của câu hỏi mới với loại của câu hỏi trong cơ sở dữ liệu. Tập dữ liệu  $D$  bao gồm các cặp câu hỏi - câu trả lời được trích xuất từ các hệ thống cQA, trong đó mỗi câu hỏi trong  $D$  được gán một nhãn loại. Để đạt được đặc trưng loại câu hỏi, nghiên cứu thực hiện 2 bước sau: (1) Xác định loại câu hỏi cho mỗi câu hỏi mới (sử dụng mô-đun phân loại câu hỏi). (2) Tính toán độ tương tự giữa loại câu hỏi mới với loại câu hỏi trong cơ sở dữ liệu (sử dụng mô-đun đo độ tương tự).

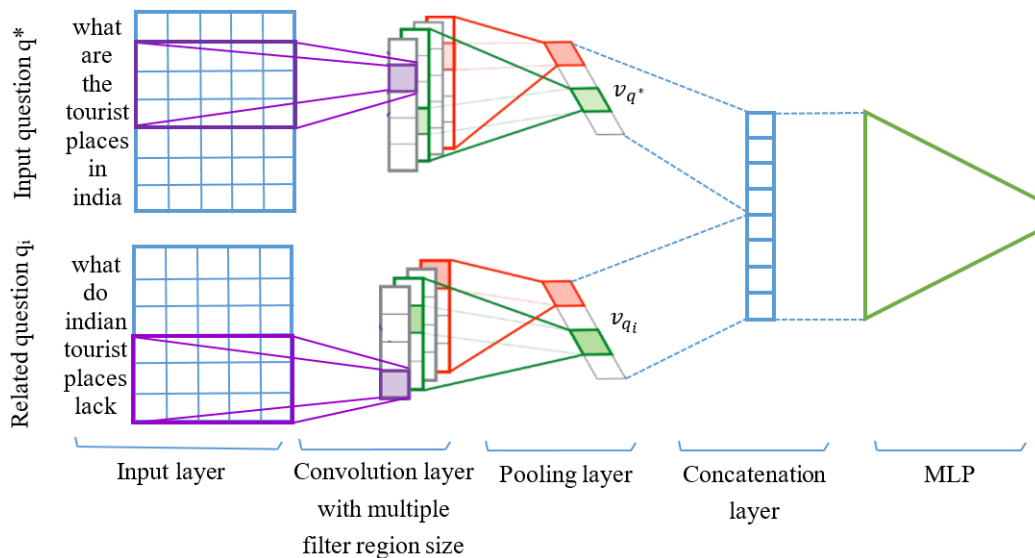
### Đặc trưng dựa trên tập nhúng từ (Word Embedding)

Sử dụng mô hình biểu diễn véc-tơ từ để mô hình hóa mối quan hệ giữa các câu hỏi.

### 4.3.2 Mô hình dựa trên mạng nơ-ron tích chập CNN

#### Kiến trúc chung của mô hình dựa trên mạng nơ-ron tích chập CNN

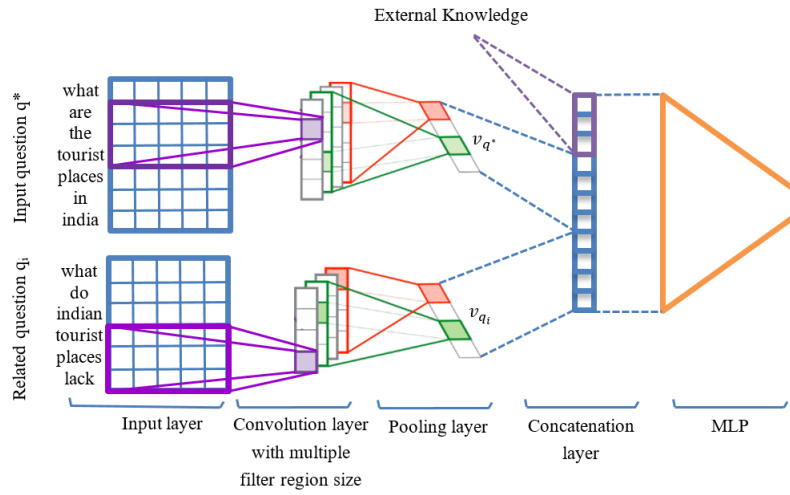
Phần này xây dựng mô hình dựa trên CNN để tính toán điểm tương đồng giữa câu hỏi mới  $q^*$  và câu hỏi  $q_i$  trong cơ sở dữ liệu, sau đó chúng tôi xếp hạng các câu hỏi trong cơ sở dữ liệu dựa trên kết quả điểm tương đồng giữa  $q^*$  và  $q_i$ . Kiến trúc chung của mô hình của chúng tôi được minh họa trong Hình 4.1.



**Hình 4.1:** Minh họa mô hình dựa trên CNN tính toán điểm tương đồng giữa  $q^*$  và  $q_i$

### 4.3.3 Mô hình dựa trên mạng nơ-ron tích chập CNN tích hợp EK

Trong phần này, luận án đề xuất xây dựng một mô hình mới, tích hợp thêm nguồn tri thức mới vào mô hình dựa trên CNN. Hình 4.2 minh họa kiến trúc chung của mô hình tích hợp này.

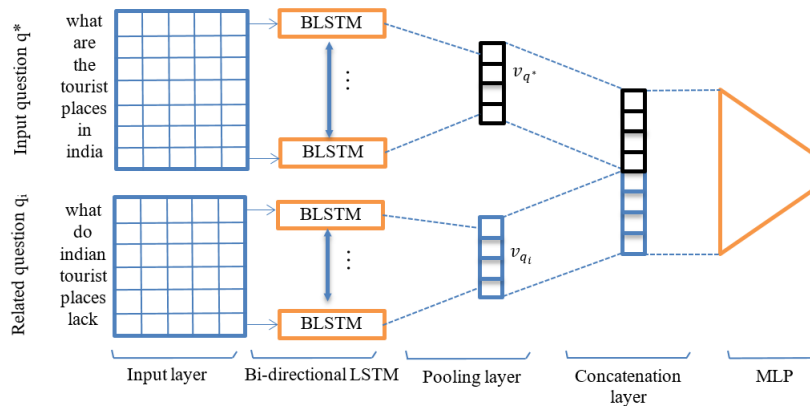


**Hình 4.2:** Minh họa mô hình dựa trên CNN tích hợp EK tính toán điểm tương đồng giữa  $q^*$  và  $q_i$

#### 4.3.4 Mô hình dựa trên mạng BLSTM

##### Kiến trúc chung của mô hình dựa trên mạng BLSTM

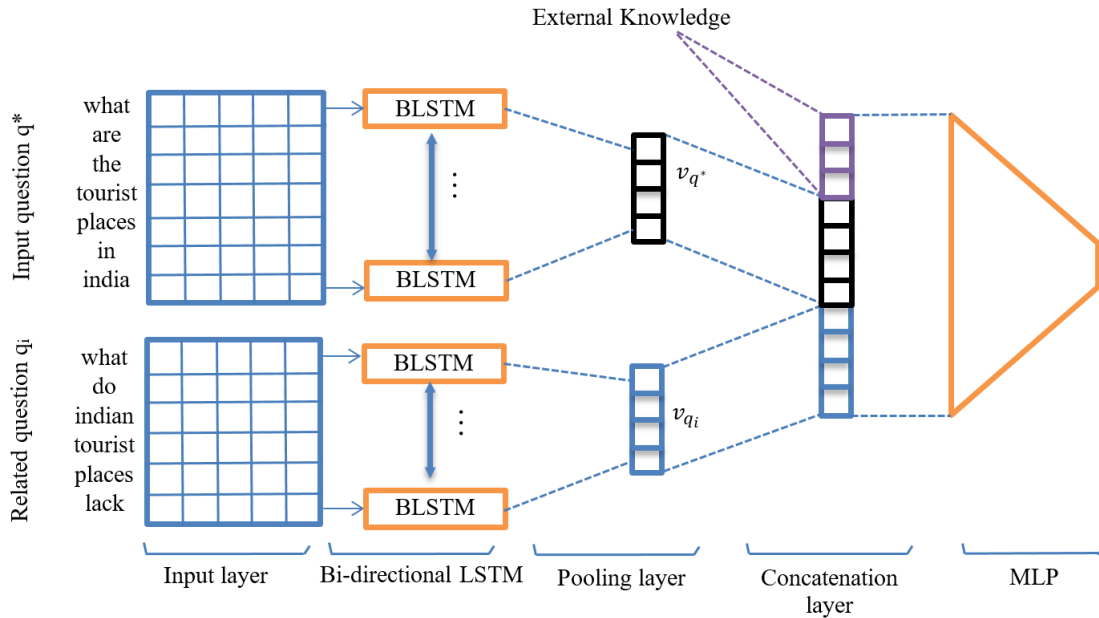
Trong phần này, luận án đề xuất xây dựng các mô hình học sâu dựa trên BLSTM để tính toán điểm tương đồng giữa câu hỏi mới  $q^*$  và câu hỏi  $q_i$  trong cơ sở dữ liệu, sau đó điểm tương đồng này sẽ được sử dụng để xếp hạng các câu hỏi  $q_i$ . Kiến trúc chung của mô hình dựa trên BLSTM được minh họa trong Hình 4.3.



**Hình 4.3:** Minh họa mô hình dựa trên BLSTM tính toán điểm tương đồng giữa  $q^*$  và  $q_i$

#### 4.3.5 Mô hình dựa trên mạng BLSTM tích hợp EK

Nghiên cứu đề xuất xây dựng mô hình mới, tích hợp thêm cơ sở tri thức mở rộng vào mô hình dựa trên BLSTM. Hình 4.4 minh họa kiến trúc mô hình.



**Hình 4.4:** Minh họa mô hình dựa trên BLSTM tích hợp EK tính toán điểm tương đồng giữa  $q^*$  và  $q_i$

## 4.4 Thực nghiệm

### 4.4.1 Tập dữ liệu và các độ đo đánh giá

Để cài đặt các thực nghiệm trong chương này, luận án sử dụng 2 tập dữ liệu cQA. Tập dữ liệu thứ nhất là SemEval 2016 task 3 subtask B như đã được mô tả trong Chương 2. Tập dữ liệu thứ hai được sử dụng là Quora. Tập dữ liệu này được trích rút từ <https://www.quora.com/>, một diễn đàn cQA mà ở đó mọi người có thể gửi những câu hỏi, câu trả lời về nhiều khía cạnh khác nhau. Tập dữ liệu Quora bao gồm 404082 cặp câu hỏi (câu hỏi mới và câu hỏi liên quan). Tập dữ liệu này được phân chia thành 363665 cặp câu hỏi được sử dụng như là tập dữ liệu huấn luyện và 40417 cặp câu hỏi được sử dụng như là tập dữ liệu kiểm tra. Mỗi điểm dữ liệu là một cặp câu hỏi (câu hỏi mới và câu hỏi liên quan) và một nhãn mang giá trị 1 hoặc 0. Giá trị nhãn là 1 tức là hai câu hỏi có tương tự nhau, ngược lại mang giá trị 0.

### 4.4.2 Các thực nghiệm trên tập dữ liệu SemEval 2016

#### Cài đặt thực nghiệm

Tập nhúng từ (Word embedding) được huấn luyện lại bằng cách sử dụng công cụ Gensim word2vec. Dữ liệu huấn luyện cho word2vec là tập dữ liệu được cung cấp bởi SemEval 2016. Các thông số huấn luyện được thiết lập như sau: (1) kích thước véc-tơ từ là 200; (2) khoảng cách tối đa giữa từ hiện tại và từ được dự đoán trong một câu được đặt thành 5; (3) mô hình huấn luyện véc-tơ bỏ qua

tất cả các từ có tần số xuất hiện nhỏ hơn 5. Ngoài các véc-tơ được xây dựng mới, các thực nghiệm cũng sử dụng các véc-tơ 300 chiều đã được huấn luyện và cung cấp bởi word2vec. Các mô hình được huấn luyện theo từng lô nhỏ (kích thước batch size là 64) và độ dài tối đa của các câu hỏi đầu vào và các câu hỏi liên quan là 40. Bất kì kí tự nào vượt ra khỏi phạm vi này sẽ bị hủy.

Trong mô hình dựa trên mạng BLSTM, các thực nghiệm sử dụng các LSTM với số đơn vị ẩn bằng 100, theo sau các LSTM là các tầng tổng hợp (pooling layer) và tầng tỉ lệ bỏ học (dropout layer).

#### Kết quả và thảo luận

Bảng 4.1 tóm tắt kết quả thực nghiệm trên tập dữ liệu SemEval 2016 task 3 subtask B sử dụng mô hình dựa trên CNN và bảng 4.2 tóm tắt kết quả sử dụng mô hình dựa trên BLSTM.

**Bảng 4.1:** Các kết quả thực nghiệm trên tập dữ liệu SemEval 2016 sử dụng mô hình dựa trên CNN

Các mô hình	Kích thước véc-tơ từ	Các độ đo phân loại				Các độ đo xếp hạng		
		Acc	P	R	F1	MAP	AvgRec	MRR
CNN-based	d = 300	73,71	53,65	62,19	57,60	72,95	87,87	78,29
CNN-based	d = 200	74,00	53,65	62,81	57,87	73,21	88,35	79,24
CNN-based + EK	d = 300	82,57	71,24	75,11	73,13	78,37	91,97	86,23
CNN-based + EK	d = 200	82,86	72,10	75,34	73,68	78,38	92,01	86,23

**Bảng 4.2:** Các kết quả thực nghiệm trên tập dữ liệu SemEval 2016 sử dụng mô hình dựa trên BLSTM

Các mô hình	Kích thước véc-tơ từ	Các độ đo phân loại				Các độ đo xếp hạng		
		Acc	P	R	F1	MAP	AvgRec	MRR
BLSTM-based	d = 300	74,14	53,22	63,27	57,81	73,86	87,08	80,51
BLSTM-based	d = 200	74,29	53,65	63,45	58,14	73,88	87,13	80,51
BLSTM-based + EK	d = 300	83,43	71,24	77,22	74,11	78,44	92,01	86,35
BLSTM-based + EK	d = 200	83,57	71,67	77,31	74,39	78,48	92,10	86,35

#### 4.4.3 Các thực nghiệm trên tập dữ liệu Quora

##### Cài đặt thực nghiệm

Để thiết lập các cài đặt thực nghiệm sử dụng tập dữ liệu Quora, mô hình giữ nguyên các cấu hình giống như trong phần 4.4.2, ngoại trừ những khác biệt sau: Thứ nhất, mô hình huấn luyện theo từ lô nhỏ với kích thước là 128; Thứ hai, đặt độ dài tối đa của câu hỏi mới và câu hỏi liên quan là 20 thay vì 40 (do dữ

liệu Quora bao gồm các câu hỏi ngắn hơn); Thứ ba, dữ liệu huấn luyện cho xây dựng các véc-tơ từ là kho dữ liệu Quora. Kho dữ liệu này chứa khoảng 35,65 triệu từ và 77845 từ duy nhất. Mô hình cũng sử dụng các véc-tơ 300 chiều đã được huấn luyện và cung cấp bởi word2vec bằng cách sử dụng một phần của tập dữ liệu Google News.

#### Kết quả và thảo luận

Bảng 4.3 trình bày các kết quả của các thực nghiệm trên tập dữ liệu Quora sử dụng mô hình dựa trên CNN. Bảng 4.4 trình bày các kết quả của các thực nghiệm trên tập dữ liệu Quora sử dụng mô hình dựa trên BLSTM.

**Bảng 4.3:** Các kết quả thực nghiệm trên tập dữ liệu Quora sử dụng mô hình dựa trên mạng nơ-ron tích chập CNN

Các mô hình	Kích thước véc-tơ từ	Acc	P	R	F1
CNN-based	d = 300	77,85	60,54	71,54	65,58
CNN-based	d = 200	79,40	70,31	70,51	70,41
CNN-based + EK	d = 300	86,81	70,89	89,05	78,91
CNN-based + EK	d = 200	87,54	74,78	87,65	80,71

**Bảng 4.4:** Các kết quả thực nghiệm trên tập dữ liệu Quora sử dụng mô hình dựa trên mạng BLSTM

Các mô hình	Kích thước véc-tơ từ	Acc	P	R	F1
BLSTM-based	d = 300	78,92	65,01	71,84	68,25
BLSTM-based	d = 200	79,40	76,72	68,17	72,19
BLSTM-based + EK	d = 300	87,55	73,68	88,69	80,49
BLSTM-based + EK	d = 200	87,79	71,87	91,24	80,41

## 4.5 Kết luận chương

Chương này đã trình bày những nội dung, kết quả nghiên cứu về vấn đề xây dựng các mô hình học sâu cho bài toán có tập dữ liệu nhỏ, tích hợp thêm các nguồn tri thức mới vào mô hình học sâu để giúp xác định tốt hơn độ tương tự giữa các câu hỏi. Các mô hình đề xuất này đã nâng cao hiệu quả của việc giải quyết bài toán tìm kiếm và xếp hạng các câu hỏi liên quan đến câu hỏi mới trong các cQA.

# KẾT LUẬN

Trong phần này, chúng tôi tóm lược lại các kết quả chính và những đóng góp của luận án. Ngoài ra, chúng tôi còn trình bày một số hạn chế của luận án và thảo luận về hướng phát triển cho các nghiên cứu tiếp theo trong tương lai.

## 1. Tóm lược các kết quả và đóng góp của luận án

Trong 4 chương của luận án, ngoài Chương 1 trình bày tổng quan; nội dung và các kết quả nghiên cứu được trình bày ở các chương chính là Chương 2, Chương 3 và Chương 4. Các đóng góp chính và kết quả của luận án được tóm tắt như sau:

Thứ nhất, luận án đã xây dựng mô hình để tìm kiếm và xếp hạng các câu hỏi trong cơ sở dữ liệu liên quan đến câu hỏi mới. Mô hình xây dựng đã xem xét các khía cạnh khác nhau để hiểu câu hỏi. Bên cạnh các đặc trưng truyền thống, luận án đã sử dụng các khía cạnh hiệu quả hơn bao gồm sự biểu diễn véc-tơ từ (word2vec) và các loại câu hỏi. Luận án sử dụng mô hình word2vec để biểu diễn mỗi từ dưới dạng một véc-tơ, mô-đun phân loại câu hỏi để xác định loại (category) cho câu hỏi mới. Sau đó, tất cả các loại đặc trưng này được kết hợp và làm đầu vào cho một mô hình học máy để có được điểm số tương tự giữa câu hỏi mới với các cặp câu hỏi-câu trả lời trong cơ sở dữ liệu và sử dụng điểm số này để xếp hạng các câu hỏi. Các kết quả thực nghiệm trên bộ dữ liệu SemEval 2016 task 3 subtask B cho thấy kết quả thu được đạt độ chính xác phân loại (*accuracy*) là 81,86% và độ đo xếp hạng (*MAP*) là 78,27% khi sử dụng bộ phân loại SVM; 82,29% và 78,35% khi sử dụng bộ phân loại MLP.

Thứ hai, luận án đã đề xuất sử dụng thêm nguồn tài nguyên bên ngoài cho vấn đề đánh giá độ tin cậy của các câu trả lời trong cQA. Wikipedia - bách khoa toàn thư mở được lựa chọn như là nguồn tài nguyên bên ngoài đáng tin cậy nhất dùng để bổ sung thông tin cho việc xác minh độ tin cậy của các câu trả lời. Luận án xem xét nhiệm vụ đánh giá độ tin cậy của các câu trả lời trong cQA như là một vấn đề phân loại và sử dụng bộ phân loại SVM để xác định mỗi ứng viên trả lời là “good” hoặc “bad”. Luận án đã trích rút các loại đặc trưng truyền thống và sau đó bổ sung thêm các loại đặc trưng mới dựa trên việc đo lường sự giống nhau giữa các câu trả lời ứng viên và các tài liệu Wikipedia. Luận án cũng đã sử dụng mô hình biểu diễn véc-tơ từ (word2vec) để biểu diễn các từ



trong câu trả lời, các tài liệu Wikipedia nhằm phục vụ tốt hơn cho việc tính toán sự giống nhau này. Kết quả thực nghiệm cho thấy đề xuất sử dụng thêm nguồn tài nguyên bên ngoài mang lại kết quả tốt hơn (*accuracy* là 84,62% và *F1 – measure* là 75%) so với việc chỉ sử dụng các đặc trưng truyền thống (*accuracy* là 79,23% và *F1 – measure* là 66,67%).

Thứ ba, luận án đã nghiên cứu, xây dựng mô hình học sâu để nâng cao hiệu suất cho bài toán có tập dữ liệu nhỏ, dữ liệu thưa. Luận án đã xây dựng các mô hình dựa trên mạng CNN, BLSTM để đánh giá độ tương tự giữa các câu hỏi trong các cQA. Để nâng cao hiệu quả của việc áp dụng các mô hình học sâu cho bài toán có tập dữ liệu nhỏ, nghiên cứu đã sử dụng thêm nguồn tri thức mới để tích hợp với các đặc trưng được sinh ra từ các tầng của mạng CNN, BLSTM và được sử dụng làm đầu vào cho MLP. Vì các câu hỏi trong cQA thường thưa thớt, nhiễu và mơ hồ. Do đó, kết hợp cơ sở tri thức với biểu diễn véc-tơ của câu hỏi sẽ giúp xác định tốt hơn sự giống nhau giữa các câu hỏi. Kết quả thực nghiệm trên bộ dữ liệu SemEval 2016 task 3 subtask B cho kết quả *accuracy* và *MAP* cao nhất là 82,86% và 78,38% khi sử dụng mô hình tích hợp dựa trên mạng CNN. Các kết quả này sử dụng mô hình tích hợp dựa trên mạng BLSTM là 83,57% và 78,48%. Các kết quả thực nghiệm sử dụng tập dữ liệu Quora đạt được độ chính xác *accuracy* và *F1 – measure* cao nhất là 87,54% và 80,71% trên mô hình học sâu dựa trên mạng CNN. Đối với mô hình tích hợp nguồn tri thức mới sử dụng mạng BLSTM, các độ đo *accuracy* và *F1 – measure* cao nhất đạt được là 87,79% và 80,49%.

## 2. Hạn chế và hướng phát triển của luận án

Luận án đã nghiên cứu, xây dựng một số mô hình mới để tích hợp nhiều nguồn tri thức vào trong các mô hình học sâu nhằm cải thiện chất lượng của các hệ thống cQA. Những kết quả đạt được bước đầu sẽ là động lực cho những nghiên cứu tiếp theo. Trong quá trình nghiên cứu và thực nghiệm, luận án vẫn còn một số hạn chế và hướng phát triển, cụ thể như sau:

- Các thực nghiệm mới chỉ xây dựng trên tập dữ liệu hỏi đáp với ngôn ngữ tiếng Anh. Trong tương lai chúng tôi sẽ nghiên cứu để có thể xây dựng và áp dụng các mô hình mới trên các bộ dữ liệu tiếng Việt.
- Nghiên cứu để sử dụng, tích hợp nhiều nguồn tri thức hơn trong việc đánh giá chất lượng của các câu trả lời cũng như đánh giá sự tương đồng giữa các câu hỏi.
- Nghiên cứu để có thể tổng hợp các câu trả lời khác nhau thành câu trả lời đầy đủ nhất cho câu hỏi.

# Danh mục công trình khoa học của tác giả liên quan đến luận án

- [1] **Nguyễn Văn Tú**, Lê Anh Cường, Nguyễn Hà Nam. (2015). Phân loại câu hỏi sử dụng sự kết hợp của nhiều đặc trưng. *Tạp chí Khoa học và Kỹ thuật - Học viện KTQS*, Số 172, pages 5-14.
- [2] **Van-Tu Nguyen**, Anh-Cuong Le. (2016). Improving Question Classification by Feature Extraction and Selection. *Indian Journal of Science and Technology*, Vol 9(17), DOI: 10.17485/ijst/2016/v9i17/93160. Scopus.
- [3] **Van-Tu Nguyen**, Anh-Cuong Le. (2016). Answer Validation For Question Answering Systems By Using External Resources. In *Proceedings of International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*, Springer, pages 305-316. Scopus, DBLP.
- [4] **Nguyễn Văn Tú**, Lê Anh Cường, Nguyễn Hà Nam. (2017). Xây dựng các cặp câu hỏi-câu trả lời chất lượng cao từ các trang web hỏi đáp cộng đồng. *Tạp chí khoa học công nghệ thông tin và truyền thông, học viện Công nghệ BCVT*, pages 25-33.
- [5] **Van-Tu Nguyen**, Anh-Cuong Le, Dinh-Hong Vu. (2017). An Efficient Model for Finding and Ranking Related Questions in community Question Answering Systems. In *Proceedings of 4th International conference on Information system Design and Intelligent Applications*, Springer, pages 776-786. Scopus.
- [6] **Van-Tu Nguyen**, Anh-Cuong Le. (2018). Deep Neural Network-based Models for Ranking Question - Answering Pairs in Community Question Answering Systems. In *Proceedings of International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*, Springer, pages 179-190. Scopus, DBLP.
- [7] **Van-Tu Nguyen**, Anh-Cuong Le, Ha-Nam Nguyen. (2020). A Model of Convolutional Neural Network Combined with External Knowledge to Measure the Question Similarity for Community Question Answering Systems. *International Journal of Machine Learning and Computing*, vol. 11, no. 3, pages 194-201. DOI: 10.18178/ijmlc.2021.11.3.1035