



## XẾP HẠNG CÂU TRẢ LỜI TRONG CÁC TRANG WEB HỎI ĐÁP CỘNG ĐỒNG

### Ranking related answers in communityquestion answer sites

Nguyễn Văn Tú<sup>1</sup>, Trần Thị Quyên<sup>2</sup>

<sup>1</sup>tuspttb@gmail.com, <sup>2</sup>quyencdsl@gmail.com

<sup>1</sup>Trường Đại học Tây Bắc, Sơn La, Việt Nam

<sup>2</sup>Trường Cao đẳng Sơn La, Sơn La, Việt Nam

Đến tòa soạn: 23/05/2017; Chấp nhận đăng: 17/08/2017

**Tóm tắt.** Các trang web hỏi đáp cộng đồng có chứa một lượng lớn thông tin hỏi-đáp có giá trị sinh ra bởi những người sử dụng. Trong các trang web hỏi đáp cộng đồng, người dùng có thể gửi các câu hỏi, trả lời các câu hỏi của người khác và cung cấp thông tin phản hồi cho những câu hỏi/câu trả lời. Trong nghiên cứu này chúng tôi tập trung giải quyết bài toán xếp hạng lại các câu trả lời của người dùng trong các trang web hỏi đáp cộng đồng. Chúng tôi thực hiện trích rút nhiều loại đặc trưng quan trọng từ mỗi cặp câu hỏi – câu trả lời nhằm đánh giá chính xác sự liên quan giữa chúng, sau đó chúng tôi xây dựng mô hình để phân loại và xếp hạng các câu trả lời theo độ liên quan của chúng với câu hỏi. Các kết quả thực nghiệm trên bộ dữ liệu cung cấp bởi SemEval 2016 cho thấy những đề xuất của chúng tôi cho kết quả cao hơn so với các nghiên cứu trước đó.

**Từ khóa:** Hỏi đáp cộng đồng; Hệ thống hỏi đáp tự động; Xếp hạng câu trả lời; Trích rút đặc trưng

**Abstract.** Community question - answer sites contain large amounts of valuable question - answer information generated by users. In community question answer sites, users can submit questions, answer other people's questions, and provide feedback on their questions/answers. In this study, we focused on solving the problem of ranking answers in community question answer sites. We extracted a variety of important features from each question - answer pair to accurately assess the relevance of them, then we built the model to classify and ranking answers according to their relevance to the question. The experimental results on the dataset provided by SemEval 2016 shows that our proposed give higher results than previous studies

**Keywords:** Community question answer; Automatic question answer system; Ranking answer; Feature extraction

### 1. TỔNG QUAN

Trong lĩnh vực xử lý ngôn ngữ tự nhiên và truy xuất thông tin, vấn đề hỏi-đáp đã thu hút nhiều sự chú ý trong những năm qua. Tuy nhiên, các nghiên cứu về hỏi-đáp chủ yếu tập trung vào việc tìm câu trả lời chính xác cho câu hỏi factoid được trích rút từ các tài liệu liên quan. Các đánh giá nổi tiếng nhất về nhiệm vụ hỏi-đáp factoid là hội nghị truy hồi văn bản (Text REtrieval Conference-TREC<sup>1</sup>). Các câu hỏi và câu trả lời được phát hành bởi TREC đã trở thành nguồn dữ liệu quan trọng cho các nhà nghiên cứu trong việc nghiên cứu xây dựng các hệ thống hỏi đáp tự động [9]. Tuy nhiên, khi phải đối mặt với các câu hỏi non-factoid như các câu hỏi về lý do tại sao, như thế nào, hoặc những gì về, ... hầu như không có hệ thống hỏi đáp tự động nào làm việc tốt.

Các cặp câu hỏi-câu trả lời do người dùng tạo ra chắc chắn sẽ rất quan trọng để giải quyết vấn đề trả lời các câu hỏi non-factoid. Rõ ràng, những cặp câu hỏi-câu trả lời tự nhiên thường được tạo ra trong quá trình giao tiếp của con người thông qua phương tiện truyền thông xã hội Internet, trong đó chúng tôi đặc biệt quan tâm tới các trang web hỏi đáp dựa vào cộng đồng. Các trang web hỏi đáp dựa vào cộng đồng cung cấp nền tảng mà ở đó người dùng có thể tự do đặt câu hỏi, cung cấp câu trả lời và các thông tin phản hồi (ví dụ, bằng cách biểu quyết hoặc cho ý kiến) cho những câu hỏi/câu trả lời và những câu trả lời tốt nhất sẽ được lựa chọn và xếp hạng để hiển thị cho người dùng.

Tuy nhiên do người dùng có thể tự do gửi câu trả lời cho mỗi câu hỏi nên mỗi câu hỏi có thể nhận được nhiều câu trả lời, trong đó chất lượng của các câu trả lời này có nhiều sự

khác nhau. Vì vậy trước khi hiển thị các câu trả lời cho người dùng thì các trang web hỏi đáp cộng đồng cần phải sắp xếp lại các câu trả lời này theo mức độ từ các câu trả lời tốt nhất cho câu hỏi.

Đã có một số công trình nghiên cứu về vấn đề xếp hạng lại các câu trả lời trong các trang web hỏi đáp cộng đồng. Trong nghiên cứu của Daniel Balchev và các tác giả khác [1], các tác giả đã sử dụng nhiều loại đặc trưng từ thông tin cung cấp bởi người dùng (số lượng bình chọn cho mỗi câu trả lời) đến các đặc trưng từ vệt trích rút từ các cặp câu hỏi – câu trả lời; sau đó sử dụng bộ phân loại SVM để sinh ra điểm số tương tự giữa câu hỏi và câu trả lời và dùng điểm số tương tự này để xếp hạng lại các câu trả lời theo độ tương tự của chúng với câu hỏi. Trong nghiên cứu của Marc Franco-Salvador và các tác giả khác [4], các tác giả sử dụng mạng ngữ nghĩa đa ngôn ngữ lớn nhất BabelNet để sinh ra các đồ thị tri thức cho các câu hỏi và các câu trả lời, sau đó tính toán độ tương tự của câu hỏi và câu trả lời dựa trên các đồ thị tri thức của chúng. Trong khi đó, nghiên cứu của Chang'e Jia và các tác giả khác [5], các tác giả lại sử dụng mô hình Latent Semantic Analysis để xác định chủ đề cho mỗi câu hỏi, câu trả lời và tính toán độ tương tự giữa các chủ đề này. Nghiên cứu của Xiaoqiang Zhou và các tác giả khác [10], các tác giả sử dụng mô hình mạng nơ ron để mô hình hóa mối quan hệ giữa câu hỏi và câu trả lời.

Trong bài báo này, chúng tôi sử dụng tiếp cận học máy nhằm xếp hạng lại các câu trả lời cho mỗi câu hỏi trích rút từ các trang web hỏi đáp cộng đồng. Chúng tôi đề xuất sử dụng sự kết hợp của nhiều loại đặc trưng quan trọng trích rút từ mỗi

<sup>1</sup> <http://trec.nist.gov/>

cặp câu hỏi - câu trả lời và xây dựng mô hình để phân loại và xếp hạng các câu trả lời để đạt được kết quả tốt nhất.

Để thực hiện những đề xuất của mình, chúng tôi đã sử dụng tập dữ liệu cung cấp bởi SemEval 2016 trong các thực nghiệm. Chúng tôi tiến hành đánh giá thử nghiệm rộng rãi để chứng minh tính hiệu quả của phương pháp tiếp cận của chúng tôi. Các kết quả thực nghiệm của chúng tôi đã cho thấy phương pháp tiếp cận mà chúng tôi đề xuất cho kết quả phân loại và xếp hạng cao hơn so với các nghiên cứu trước đó trên cùng tập dữ liệu và các độ đo đánh giá.

### 2. BÀI TOÁN

Bài toán xếp hạng các câu trả lời trong các trang web hỏi đáp cộng đồng được định nghĩa như sau: Cho một tập Q các câu hỏi, mỗi câu hỏi  $q_i \in Q$  gắn với một tập các câu trả lời  $A_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$ . Yêu cầu xếp hạng lại các câu trả lời  $\{a_{i1}, a_{i2}, \dots, a_{in}\}$  theo độ liên quan của chúng với câu hỏi  $q_i$ .

Việc xếp hạng các câu trả lời có thể mô hình hóa bởi hàm  $r: Q \times A \rightarrow R$ , trong đó Q là tập các câu hỏi, A là tập các câu trả lời. r được mô hình hóa như một hàm tuyến tính  $r(q_i, a_{ij}) = \vec{w} \cdot \phi(q_i, a_{ij})$ , trong đó  $\vec{w}$  là một mô hình và  $\phi(q_i, a_{ij})$  cung cấp một sự biểu diễn véc tơ của cặp  $(q_i, a_{ij})$ . Sự biểu diễn véc tơ của cặp  $(q_i, a_{ij})$  sẽ được tính toán dựa trên các độ đo sự tương tự giữa câu hỏi  $q_i$  và câu trả lời  $a_{ij}$  như được trình bày trong phần 3.

Trong nghiên cứu này, chúng tôi sử dụng bộ phân loại Support Vector Machine<sup>2</sup> (SVMs) để học r và sinh ra điểm số phân loại giữa câu hỏi và câu trả lời, việc xếp hạng lại các câu trả lời sẽ dựa trên điểm số phân loại này.

### 3. TRÍCH RÚT CÁC ĐẶC TRƯNG

Để đánh giá sự liên quan giữa câu hỏi và câu trả lời chúng tôi đã thực hiện trích rút các loại đặc trưng quan trọng như được trình bày dưới đây.

#### 3.1 Các đặc trưng từ vựng

##### Đặc trưng n-gram

Các đặc trưng n-gram của một cặp câu hỏi-câu trả lời được trích rút dựa trên ngữ cảnh của các từ của câu, nghĩa là, các từ đó xuất hiện trong một cặp câu hỏi-câu trả lời. Mỗi cặp câu hỏi-câu trả lời được biểu diễn giống như sự biểu diễn tài liệu trong mô hình không gian véc tơ như sau:

$$x = (x_1, x_2, \dots, x_N) \tag{1}$$

Trong đó  $x_i$  là tần số xuất hiện của từ i trong x và N là tổng số các từ trong x. Do tính thưa thớt của các đặc trưng, chỉ các đặc trưng có giá trị khác không mới được giữ lại trong véc tơ đặc trưng. Bởi vậy mỗi cặp câu hỏi-câu trả lời cũng được biểu diễn dưới hình thức sau:

$$x = \{(t_1, f_1), \dots, (t_p, f_p)\} \tag{2}$$

trong đó  $t_i$  là từ thứ i trong x và  $f_i$  là tần số xuất hiện của  $t_i$  trong x. Để trích rút các đặc trưng n-gram, bất kì n từ liên tiếp nào trong một cặp câu hỏi - câu trả lời đều được coi là một đặc trưng. Bảng 1 là danh sách một số đặc trưng n-gram của câu hỏi "How many Grammys did Michael Jackson win in 1983?".

##### Tỉ lệ giữa số lượng từ trong câu hỏi và câu trả lời

Để đánh giá sự liên quan giữa câu hỏi và câu trả lời, trong nghiên cứu này chúng tôi sử dụng đặc trưng là tỉ lệ giữa số lượng các từ trong câu hỏi và số lượng các từ trong câu trả lời.

Bảng 1. Ví dụ về một số đặc trưng n-gram

Tên đặc trưng	Đặc trưng
Unigram	{(How, 1) (many, 1) (Grammys, 1) (did, 1) (Michael, 1) (Jackson, 1) (win, 1) (in, 1) (1983, 1) (? , 1)}
Bigram	{(How-many, 1) (many-Grammys, 1) (Grammys-did, 1) (did-Michael, 1) (Michael-Jackson, 1) ... (1983-?, 1)}
Trigram	{(How-many-Grammys, 1) (many-Grammys-did, 1) ... (in-1983-?, 1)}

##### Tỉ lệ giữa số lượng câu (sentence) trong câu câu hỏi và trả lời

Để đánh giá sự liên quan giữa câu hỏi và câu trả lời, trong nghiên cứu này chúng tôi sử dụng đặc trưng là tỉ lệ giữa số lượng các câu (sentence) trong câu hỏi và số lượng các câu trong câu trả lời.

##### Chồng chéo n-gram từ giữa câu hỏi và câu trả lời

Khi trả lời một câu hỏi nào đó trên các trang web hỏi đáp cộng đồng, người sử dụng thường có xu hướng sử dụng lại một số từ ở câu hỏi trong câu trả lời của họ. Vì vậy nếu trong câu trả lời có chứa từ hoặc cụm từ của câu hỏi thì câu trả lời đó có khả năng là một câu trả lời tốt, liên quan nhiều đến câu hỏi. Để tính toán sự chồng chéo từ giữa câu hỏi và câu trả lời, chúng tôi thực hiện loại bỏ các stopwords trong mỗi câu hỏi, câu trả lời sau đó tính toán sự chồng chéo từ sử dụng n-gram từ (n=1, 2, 3).

##### Bag-of-word

Để xây dựng các đặc trưng này, chúng tôi thực hiện loại bỏ các từ stopwords trong mỗi câu hỏi và câu trả lời. Các câu hỏi và câu trả lời sau đó được biểu diễn dưới dạng véc tơ (bag-of-word). Để tính toán sự giống nhau giữa câu hỏi và câu trả lời chúng tôi tính toán độ tương tự giữa hai véc tơ. Các độ đo sử dụng tính toán độ tương tự bao gồm: euclidean, manhattan, minkowski, cosine, jaccard. Bảng 2 là một ví dụ về việc tính toán các đặc trưng đo sự giống nhau này.

Bảng 2. Ví dụ về các đặc trưng Bag-of-word

Câu hỏi	Câu trả lời	Các độ đo	Các giá trị độ đo
Massage oil. Where I can buy good oil for massage?	You might be able to find Body Massage Oil in Body Shop at Landmark or City Centre, and if they do have it there, ...	euclidean	5.196152
		manhattan	25
		minkowski	3.141
		cosine	0.405062
		jaccard	1.0

#### 3.2 Các đặc trưng dựa trên sự biểu diễn véc tơ từ

Chúng tôi sử dụng sự biểu diễn véc tơ từ để mô hình hóa mối quan hệ ngữ nghĩa giữa câu hỏi và câu trả lời của chúng. Chúng tôi chọn mô hình word2vec<sup>3</sup> đề xuất bởi Mikolov [6, 7] để tính toán độ tương tự ngữ nghĩa giữa câu hỏi và câu trả lời. Word2vec biểu diễn các từ dưới dạng một phân bố quan hệ với các từ còn lại. Giả sử ta có một véc tơ có số chiều 100. Khi đó, mỗi từ được biểu diễn bằng một véc tơ có các phần tử mang giá trị là phân bố quan hệ của từ này đối với các từ khác trong từ điển. Trong bài báo này chúng tôi sử dụng tập

<sup>2</sup> <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

<sup>3</sup> <https://code.google.com/p/word2vec>

dữ liệu từ Qatar Living (English)<sup>4</sup> để huấn luyện lại mô hình word2vec với các véc tơ có số chiều là 200, window = 5.

**Độ tương tự ngữ nghĩa giữa câu hỏi và câu trả lời**

Để tính toán độ tương tự ngữ nghĩa giữa câu hỏi và câu trả lời chúng tôi thực hiện như sau:

Bước 1: Tất cả các câu trong câu hỏi và câu trả lời được phân tích thành các từ tổ và các từ này biểu diễn dưới dạng các véc tơ từ sử dụng mô hình huấn luyện word2vec.

Bước 2: Đối với mỗi câu hỏi (câu trả lời), chúng tôi lấy giá trị trung bình của tất cả các véc tơ từ của các từ trong câu hỏi (câu trả lời) để có được sự biểu diễn véc tơ cho câu hỏi (câu trả lời). Việc lấy giá trị trung bình của tất cả các véc tơ từ đã được chứng minh tính hiệu quả trong các nghiên cứu [2, 3].

Bước 3: Độ tương tự giữa hai véc tơ được tính như công thức dưới đây:

$$\text{cosine}_{sim}(u,v) = \frac{\sum_{i=1}^n u_i \times v_i}{\sqrt{\sum_{i=1}^n (u_i)^2} \times \sqrt{\sum_{i=1}^n (v_i)^2}} \quad (3)$$

Trong đó  $u$  và  $v$  là hai véc tơ  $n$  chiều,  $u_i$  là thành phần thứ  $i$  của véc tơ  $u$ .

Đối với việc tính toán độ tương tự chúng tôi sử dụng tính toán độ tương tự giữa các thành phần của câu hỏi với câu trả lời: giữa tiêu đề của câu hỏi (QSubject) với câu trả lời, giữa phần mô tả của câu hỏi (QBody) với câu trả lời, giữa câu hỏi (Qsubject+ QBody) với câu trả lời. Bảng 3 là một ví dụ về việc tính toán độ tương tự ngữ nghĩa giữa câu hỏi và câu trả lời.

**Bảng 3.** Ví dụ về tính toán độ tương tự ngữ nghĩa giữa câu hỏi và câu trả lời

	Câu hỏi	Câu trả lời	Độ tương tự
Qsubject	Massage oil.	You might be able to find	0.2692716
QBody	Where I can buy good oil for massage?	Body Massage Oil in Body Shop at	0.7076797
Qsubject +QBody	Massage oil. Where I can buy good oil for massage?	Landmark or City Centre, and if they do have it there, ...	0.6686702

**Giống từ giữa câu hỏi và câu trả lời**

Các câu hỏi và câu trả lời được phân tích thành các từ tổ và biểu diễn dưới dạng các véc tơ từ sử dụng mô hình huấn luyện word2vec. Mỗi từ  $t_k$  trong câu hỏi sau đó sẽ được giống với tất cả các từ trong câu trả lời và lựa chọn từ có độ tương tự lớn nhất như công thức dưới đây:

$$\text{score}(t_k) = \max_{1 \leq h \leq m} (\text{word2vec}_{sim}(t_k, b_h)) \quad (4)$$

Trong đó:

$m$ : số từ trong câu hỏi.

$t_k$ : sự biểu diễn véc tơ của từ thứ  $k$  trong câu hỏi

$b_h$ : sự biểu diễn véc tơ của từ thứ  $h$  trong câu trả lời

$\text{word2vec}_{sim}(t_k, b_h)$ : độ tương tự cosin giữa hai sự biểu diễn véc tơ từ của  $t_k$  và  $b_h$ .

Điểm số tương tự giữa câu hỏi và câu trả lời được tính toán như sau:

$$\text{score}(q_i a_{ij}) = \frac{\sum_{k=1}^n \text{score}(t_k)}{n} \quad (5)$$

Trong đó  $n$  là số lượng các từ trong câu hỏi.

**Độ tương tự ngữ nghĩa giữa câu trả lời và loại của câu hỏi (QCategory)**

Chúng tôi cũng sử dụng độ tương tự ngữ nghĩa giữa mỗi câu trả lời với loại của câu hỏi (question category) tương ứng của nó. Trong tập dữ liệu làm thực nghiệm ở phần IV, các câu hỏi trong tập dữ liệu SemEval 2016 đã được phân vào một trong 27 loại khác nhau. Bảng 4 là một ví dụ về việc tính toán độ tương tự ngữ nghĩa giữa câu trả lời và các loại của câu hỏi.

**Bảng 4.** Ví dụ về tính toán độ tương tự ngữ nghĩa giữa loại câu hỏi và câu trả lời

Câu trả lời	Loại câu hỏi (QCategory)	Độ tương tự ngữ nghĩa
You might be able to find Body Massage Oil in Body Shop at Landmark or City Centre, and if they do have it there, ...	Beauty and Style	0.1182937
	Electronics	0.2048591
	Doha Shopping	0.3174826
	Cars	0.0705854

**4. CÁC THỰC NGHIỆM VÀ ĐÁNH GIÁ**

**4.1 Tập dữ liệu và các độ đo**

Trong các trang web hỏi đáp cộng đồng, mỗi câu hỏi thường chứa một tiêu đề hỏi và một đoạn văn bản ngắn mô tả về nội dung hỏi được đưa ra bởi người hỏi. Phần tiêu đề hỏi và phần mô tả có thể coi như là một câu hỏi duy nhất gồm nhiều câu [8].

Để thực hiện các thực nghiệm của mình, chúng tôi đã sử dụng tập dữ liệu từ SemEval 2016<sup>5</sup>. Tập dữ liệu này được trích rút từ các trang web hỏi đáp cộng đồng<sup>6</sup>, bao gồm các câu hỏi và mỗi câu hỏi gồm một tập các câu trả lời tương ứng. Tất cả các cặp câu hỏi-câu trả lời đều được trình bày bằng ngôn ngữ tiếng Anh. Tập dữ liệu này bao gồm 2 tập con: train – tập dữ liệu dùng để huấn luyện mô hình phân loại, test – tập dữ liệu dùng để kiểm tra tính hiệu quả của mô hình phân loại. Bảng 5 trình bày một số thống kê trên tập dữ liệu này.

**Bảng 5.** Thống kê tập dữ liệu được sử dụng

Tập dữ liệu	Số câu hỏi	Số câu trả lời
Train	2669	17900
Test	327	3270

Để đánh giá hiệu suất của mô hình, chúng tôi sử dụng các độ đo phân loại và xếp hạng. Các độ đo phân loại bao gồm: Accuracy (Acc), Precision (P), Recall (R), và  $F_1$ -measure ( $F_1$ ). Các độ đo xếp hạng bao gồm: Mean Average Precision (MAP), Average Recall (AvgRec) và Mean Reciprocal Rank (MRR).

**4.2 Các thực nghiệm**

**Thực nghiệm 1:**

Trong thực nghiệm này chúng tôi muốn kiểm tra tính hiệu quả của việc sử dụng các đặc trưng từ vựng như được trình bày trong mục 3.1. Các đặc trưng từ vựng này bao gồm: đặc trưng Unigram, tỉ lệ giữa số từ của câu trả lời và câu hỏi, tỉ lệ giữa số câu của câu trả lời và câu hỏi, chồng chéo n-gram từ giữa câu hỏi và câu trả lời, bag-of-word. Bảng 6 trình bày các kết quả của thực nghiệm 1.

**Thực nghiệm 2:**

Thực nghiệm thứ 2 này chúng tôi sử dụng các đặc trưng dựa trên sự biểu diễn véc tơ từ. Để tính toán được sự giống

<sup>4</sup> <http://alt.qcri.org/semeval2016/task3/index.php?id=data-and-tools>

<sup>5</sup> <http://alt.qcri.org/semeval2016/task3/>

<sup>6</sup> <http://www.qatarliving.com/>

nhau giữa câu hỏi và câu trả lời, chúng tôi thực hiện: (1) loại bỏ các từ stopwords trong mỗi câu hỏi và câu trả lời, (2) biểu diễn mỗi câu hỏi và câu trả lời sử dụng mô hình word2vec, (3) sử dụng độ đo cosine để tính toán độ tương tự giữa hai véc tơ. Kết quả của thực nghiệm 2 được trình bày trong Bảng 7.

**Thực nghiệm 3:**

Trong thực nghiệm 3 chúng tôi thực hiện phân loại và xếp hạng các cặp câu hỏi - câu trả lời bằng cách kết hợp tất cả các loại đặc trưng đã được thực hiện trong các thực nghiệm 1 và 2. Các kết của thực nghiệm 3 được trình bày trong Bảng 8.

Từ các kết quả của các thực nghiệm trên chúng tôi nhận thấy rằng việc phân loại và xếp hạng các cặp câu hỏi - câu trả lời trong các hệ thống hỏi đáp cộng đồng cần sự kết hợp của nhiều loại đặc trưng khác nhau để cho kết quả tốt hơn. Các đặc trưng về từ vựng đóng một vai trò quan trọng trong nhiệm vụ này. Điều này là do các câu trả lời của người dùng

thường được viết một cách tự do, không theo một cấu trúc nhất định, có nhiều câu trả lời trình bày sai cấu trúc cú pháp hoặc chứa những từ không liên quan đến câu hỏi. Các kết quả từ thực nghiệm 2 cho thấy việc trích rút các đặc trưng dựa trên sự biểu diễn véc tơ từ (ở đây là word2vec) cũng có ý nghĩa quan trọng trong việc phân loại và xếp hạng các cặp câu hỏi - câu trả lời. Việc huấn luyện lại mô hình word2vec và sử dụng nó trong việc tính toán độ tương tự ngữ nghĩa giữa các thành phần của câu hỏi với câu trả lời, giữa câu trả lời với các loại của câu hỏi đã cho kết quả phân loại cao.

Trong thực nghiệm 3 chúng tôi đã thực hiện việc kết hợp của nhiều loại đặc trưng khác nhau và đã đạt được kết quả phân loại và xếp hạng cao nhất trong tất cả các độ đo mà chúng tôi sử dụng. Điều này cũng chứng minh rằng vấn đề phân loại và xếp hạng các cặp câu hỏi - câu trả lời trong các trang web hỏi đáp cộng đồng cần sự kết hợp của nhiều loại đặc trưng khác nhau.

**Bảng 6.** Kết quả phân loại và xếp hạng sử dụng các đặc trưng từ vựng

Các đặc trưng sử dụng	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F <sub>1</sub>	MAP	AvgRec	MRR
Unigram	49.48	23.04	10.38	14.32	57.24	68.19	61.22
Tỉ lệ giữa số từ của câu trả lời và câu hỏi	61.04	56.69	17.53	26.78	61.69	74.40	69.05
Tỉ lệ giữa số câu của câu trả lời và câu hỏi	61.07	56.86	17.46	26.71	61.83	74.52	69.25
Chồng chéo n-gram từ giữa câu hỏi và câu trả lời	62.78	64.74	18.51	28.79	63.85	76.14	71.68
Bag-of-word	63.67	70.09	18.51	29.29	65.58	77.70	74.03
Tất cả các đặc trưng trên	67.00	82.55	23.85	37.01	67.90	79.30	76.18

**Bảng 7.** Kết quả phân loại và xếp hạng sử dụng các đặc trưng dựa trên sự biểu diễn véc tơ từ

Các đặc trưng sử dụng	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F <sub>1</sub>	MAP	AvgRec	MRR
Độ tương tự ngữ nghĩa giữa câu hỏi và câu trả lời	69.08	85.18	28.97	43.23	69.42	80.48	77.49
Giống từ giữa câu hỏi và câu trả lời	67.83	78.09	28.97	42.26	68.40	79.37	76.14
Độ tương tự ngữ nghĩa giữa câu trả lời và loại câu hỏi	63.21	60.68	26.94	37.31	65.32	75.37	70.91
Tất cả các đặc trưng trên	72.35	86.96	37.62	52.52	71.82	81.93	78.52

**Bảng 8.** Kết quả phân loại và xếp hạng sử dụng sự kết hợp của nhiều loại đặc trưng

Các đặc trưng sử dụng	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F <sub>1</sub>	MAP	AvgRec	MRR
Các đặc trưng từ vựng	67.00	82.55	23.85	37.01	67.90	79.30	76.18
Các đặc trưng dựa trên sự biểu diễn véc tơ từ	72.35	86.96	37.62	52.52	71.82	81.93	78.52
Cả hai loại đặc trưng trên	72.75	88.97	37.62	52.88	72.38	82.36	79.19

**5. SO SÁNH VỚI CÁC NGHIÊN CỨU KHÁC**

**Bảng 9.** So sánh với các kết quả nghiên cứu khác

Nghiên cứu của các tác giả	Các độ đo phân loại				Các độ đo xếp hạng		
	Acc	P	R	F <sub>1</sub>	MAP	AvgRec	MRR
Chang'e Jia và các tác giả [5]	64.43	73.18	19.71	31.06	<b>71.52</b>	82.67	80.26
Xiaoqiang Zhou và các tác giả [10]	<b>69.51</b>	62.48	62.53	62.50	70.90	83.36	77.38
Daniel Balchev và các tác giả [1]	56.73	47.81	70.58	57.00	68.79	79.94	80.00
Marc Franco-Salvador và các tác giả [4]	63.21	55.64	46.80	50.84	67.42	79.38	76.97
Nghiên cứu của chúng tôi	<b>72.75</b>	88.97	37.62	52.88	<b>72.38</b>	82.36	79.19

Chúng tôi cũng thực hiện so sánh các kết quả nghiên cứu của chúng tôi với các kết quả nghiên cứu của các tác giả khác. Các nghiên cứu mà chúng tôi sử dụng để so sánh ở đây cũng sử dụng tập dữ liệu từ SemEval 2016 và sử dụng cùng các độ đo đánh giá.

Bảng 9 trình bày một số kết quả nghiên cứu của các tác giả khác để so sánh với các kết quả của chúng tôi trong vấn đề xếp hạng các câu trả lời trong các trang web hỏi đáp cộng đồng.

Từ bảng so sánh cho thấy nghiên cứu của chúng tôi cho kết quả cao nhất về cả độ đo phân loại Accuracy và độ đo xếp hạng MAP.

## 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã trình bày những đề xuất của chúng tôi trong việc xếp hạng các câu trả lời trong các trang web hỏi đáp cộng đồng. Chúng tôi đã thực hiện trích rút nhiều loại đặc trưng khác nhau từ các đặc trưng từ vựng, các đặc trưng dựa trên sự biểu diễn véc tơ từ (ở đây là word2vec) và sử dụng bộ phân loại Support Vector Machine để phân loại các cặp câu hỏi- câu trả lời, sinh ra điểm số phân loại dùng để xếp hạng các câu trả lời. Các kết quả của thực nghiệm cho thấy đề xuất của chúng tôi đạt kết quả phân loại với độ đo *Accuracy* là 72.75% và kết quả xếp hạng *MAP* là 72.38% khi sử dụng sự kết hợp của nhiều loại đặc trưng. Các nghiên cứu tiếp theo chúng tôi sẽ nghiên cứu bổ sung thêm các loại đặc trưng mới như các thông tin người sử dụng và nghiên cứu cách kết hợp các loại đặc trưng khác nhau nhằm đạt được kết quả cao hơn nữa.

### TÀI LIỆU THAM KHẢO

- [1] Daniel Balchev, Yassen Kiprov, Ivan Koychev, Preslav Nakov, "PMI-cool at SemEval-2016 Task 3: Experiments with PMI and Goodness Polarity Lexicons for Community Question Answering," Proceedings of SemEval-2016, pp. 844-850, 2016.
- [2] Marc Franco-Salvador, Francisco Rangel, Paolo Rosso, Mariona Taule, and M. Antonia Mart, "Language variety identification using distributed representations of words and documents", Proceeding of the 6th International Conference of CLEF on Experimental IR meets Multilinguality, Multimodality, and Interaction (CLEF 2015), pp. 28-40, 2015.
- [3] Marc Franco-Salvador, Paolo Rosso, and Francisco Rangel, "Distributed representations of words and documents for discriminating similar languages", Proceeding of the Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects (LT4VarDial), RANLP, pp.11-16, 2015.
- [4] Marc Franco-Salvador, Sudipta Kar, Thamar Solorio, and Paolo Rosso, "UH-PRHLT at SemEval-2016 Task 3: Combining lexical and semantic-based features for community question answering.", Proceedings of SemEval-2016, pp .814-821, 2016.
- [5] Chang'e Jia, Xinkai Du, Chengjie Sun and Lei Lin, "ITNLP-AiKF at SemEval-2016 Task 3: a question answering system using community QA repository", Proceedings of SemEval-2016, pp. 904-909, 2016.
- [6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality", CoRR, abs/1310.4546, 2013.
- [7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space", CoRR, abs/1301.3781, 2013.
- [8] Vinay Pande, Tanmoy Mukherjee, Vasudeva Varma, "Summarizing answers for community question answer services", The International Conference of the German Society for Computational Linguistics and Language Technology, pp. 151-161, 2013
- [9] Zeyi Wen, Rui Zhang, Kotagiri Ramamohanarao, "Enabling precision/recall preferences for semi-supervised SVM training", CIKM'14, pp. 421-430, 2014.
- [10] Xiaoqiang Zhou, Baotian Hu, Jiaxin Lin, Yang Xiang, Xiaolong Wang, "ICRC-hit: A deep learning based comment sequence labeling system for answer selection challenge", Proceedings of semeval-2016, pp. 210-214, 2016.

### TIỂU SỬ TÁC GIẢ



*Nguyễn Văn Tú*

Năm sinh 1982, Thái Bình. Tốt nghiệp cử nhân tại Trường Đại học Sư phạm Thái Nguyên ngành Sư phạm Tin năm 2005, tốt nghiệp Thạc sĩ tại Trường Đại học Sư phạm Hà Nội năm 2009. Hiện đang làm nghiên cứu sinh tại Trường Đại học Công nghệ và làm việc tại khoa Toán-Lý-Tin Trường Đại học Tây Bắc. Hướng nghiên cứu bao gồm: Các kỹ thuật học máy, xử lý ngôn ngữ tự nhiên, v.v...



*Trần Thị Quyên*

Năm sinh 1985, Sơn La. Tốt nghiệp cử nhân Toán-Lý-Tin tại trường Đại học Tây Bắc năm 2009, tốt nghiệp Thạc sĩ tại Trường Đại học Sư phạm Hà Nội năm 2011. Hiện đang làm việc tại Trường Cao đẳng Sơn La. Hướng nghiên cứu bao gồm: Các kỹ thuật học máy, khai phá dữ liệu, v.v...