ĐẠI HỌC QUỐC GIA HÀ NỘI

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

NGUYỄN ĐẮC NAM

HỆ THỐNG TỰ ĐỘNG PHÂN LUỒNG CÂU HỎI VÀ GIẢI ĐÁP YÊU CẦU TRỰC TUYẾN

Ngành: Công nghệ thông tin

Chuyên ngành: Kỹ thuật phần mềm

Mã số: 60480103

TÓM TẮT LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Hà Nội – 2017

CHƯƠNG 1: TỔNG QUAN VỀ HỆ THỐNG TRẢ LỜI TỰ ĐỘNG 4
1.1 Hệ thống trả lời tự động.... 4
1.2 Tình hình nghiên cứu trong và ngoài nước.... 4
1.3 Phân loại các mô hình trả lời tự động.... 4
CHƯƠNG 2: CƠ SỞ MẠNG NƠ-RON NHÂN TẠO.....4
2.1 Kiến trúc mạng nơ-ron nhân tạo.... 4
2.2 Hoạt động của mạng nơ-ron nhân tạo.... 4
2.3 Mạng nơ-ron tái phát và ứng dụng.... 4
CHƯƠNG 3: ỨNG DỤNG MÔ HÌNH MẠNG NƠ-RON VÀO TRẢ LỜI TỰ ĐỘNG.....5
3.1 Phát sinh ngôn ngữ trả lời tự động.....5
3.2 Mô hình chuỗi tuần tự liên tiếp.....5
3.3 Mô hình trả lời tự động.....5
3.4 Một số đặc điểm khi xây dựng hệ thống trả lời tự động6
3.4.1. Phụ thuộc bối cảnh.... 6
3.4.2. Kết hợp tính cách.... 6
3.5 Các vấn đề khó khăn khi trả lời tự động bằng Tiếng Việt 6
3.5.1 Đặc điểm ngữ âm.... 7
3.5.2 Đặc điểm từ vựng:.... 7
3.5.3 Đặc điểm ngữ pháp.... 8
CHƯƠNG 4: XÂY DỰNG HỆ THỐNG TRAO ĐỔI THÔNG TIN TRỰC TUYẾN GIỮA SINH VIÊN VỚI NHÀ TRƯỜNG TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI.....9
4.1 Lựa chọn bài toán.....9
4.2 Quy trình trao đổi thông tin (hỏi đáp trực tuyến) giữa HSSV với Nhà trường tại Trường Đại học Công nghiệp Hà Nội.....9
4.2.1 Quy trình áp dụng.... 9
4.2.2 Mô tả quy trình áp dụng.... 9
4.3 Kiến trúc ứng dụng.....10
4.4 Cài đặt hệ thống.....12
4.4.1 Mô hình cài đặt.... 12
4.4.2 Môi trường cài đặt.... 13
4.4.3 Công cụ cài đặt.... 13
4.5 Kết quả đạt được.....13
4.5.1 Một số kết quả.... 13
4.5.2 Hiệu năng.... 13
KẾT LUẬN.....13
TÀI LIỆU THAM KHẢO.....14

TÓM TẮT LUẬN VĂN

Hiện tại việc tiếp nhận, giải quyết và trả lời câu hỏi thắc mắc hoặc yêu cầu của người dùng như (Hệ thống hỏi đáp Q&A và giải quyết thắc mắc): của khách hàng trong hoạt động thương mại, của người dân trong thủ tục hành chính, của học sinh - sinh viên trong hoạt động đào tạo của các trường đại học - cao đẳng ... là rất lớn. Các hoạt động tiếp nhận câu hỏi và trả lời câu hỏi hiện nay đều là hoạt động mang tính thủ công mà chưa có công cụ nào trợ giúp. Việc tiếp nhận và xử lý còn chậm, thiếu chính xác và chưa công khai minh bạch. Các câu hỏi và yêu cầu của người dùng thì đi vào nhiều lĩnh vực và thuộc nhiều đối tượng trả lời khác nhau, việc lựa chọn đúng đối tượng trả lời gây khó khăn và hiểu nhầm cho người dùng dẫn đến các câu hỏi và yêu cầu thường không được trả lời thỏa đáng.

Cho đến nay các hệ thống trực tuyến đã giải quyết được những yêu cầu tiện lợi hơn. Ví dụ như mua sắm trên mạng: người sử dụng có thể truy cập vào một địa chỉ và có thể mua sắm được nhiều mặt hàng của nhiều đơn vị sản xuất (Ví dụ amazon, lazada). Yêu cầu của người mua hàng được các website này phân tích và đưa ra các đề nghị sản phẩm hợp lý với người mua hàng nhờ vào các hệ thống trí tuệ nhân tạo (AI) và học máy (ML) giúp cải thiện doanh thu bán hàng đáng kể và là thành phần không thể thiếu trong các website bán hàng ngày nay.

Do vậy hệ thống trả lời tự động không thể thiếu trong bối cảnh hiện nay.

1. Tính cấp thiết của bài toán trả lời tự động

Trong bối cảnh mạng xã hội và các website mua sắm đang ngày càng trở nên rất phổ biến như hiện nay, con người cũng tăng nhu cầu kết nối với con người thông qua mạng xã hội, vào bất kỳ thời gian nào và ở bất cứ nơi đâu. Sẽ thật tốt hơn nếu có một hệ thống tự động thông minh hỗ trợ con người bằng cách trò chuyện, có khả năng nhắc nhở, có thể giải đáp mọi thắc mắc chỉ trong thời gian ngắn nhất.

Khái niệm về trợ lý ảo, chatbot, hay hệ thống trả lời tự động đang là chủ đề nóng, khi các công ty lớn như Microsoft (Cortana), Google (Google Assistant), Facebook (M), Apple (Siri), Samsung (Viv) đã giới thiệu các trợ lý ảo của mình, là các hệ thống trả lời tự động. Chính thức vào cuộc chơi chatbot, với mong muốn tạo ra một trợ lý ảo thực sự thông minh tồn tại trong

hệ sinh thái trong các sản phẩm của mình. Gần đây nhất Microsoft đã tạo ra Microsoft Chat Framework cho phép các nhà phát triển tạo ra các chatbot trên nền tảng Web và Skype, hay Facebook cũng phát hành F8 SDK cho phép nhà phát triển tích hợp vào Messenger.

Như vậy, hệ thống trả lời tự động có những nhiệm vụ và vai trò quan trọng, có thể trợ giúp được con người rất nhiều trong rất nhiều lĩnh vực: y tế, giáo dục, thương mại điện tử, …, xứng đáng để nghiên cứu và đưa ra các sản phẩm phù hợp với thực tế. Với sự ra đời của framework sequence-to-sequence [10] gần đây, nhiều hệ thống huấn luyện đã sử dụng các mạng nơ-ron để sinh ra các câu trả lời mới khi đưa vào mạng một câu hỏi hoặc một thông điệp. Đây là một hướng tiếp cận mới có nhiều triển vọng trong việc xây dựng một hệ thống trả lời tự động. Qua đó, chúng tôi đã nghiên cứu dựa trên khung làm việc sequence-to-sequence, để xây dựng mô hình trả lời tự động cho tiếng Việt, từ có có thể áp dụng được vào các bài toán thực tế [1].

2. Mục tiêu của luận văn

Phân luồng câu hỏi (phân tích câu hỏi) là pha đầu tiên trong kiến trúc chung của một hệ thống hỏi đáp, có nhiệm vụ tìm ra các thông tin cần thiết làm đầu vào cho quá trình xử lý của các pha sau (trích chọn tài liệu, trích xuất câu trả lời, …). Vì vậy phân tích câu hỏi có vai trò hết sức quan trọng, ảnh hưởng trực tiếp đến hoạt động của toàn bộ hệ thống. Nếu phân tích câu hỏi không tốt thì sẽ không thể tìm ra được câu trả lời. Chính vì lý do này mà tác giả chọn và nghiên cứu đề tài "**Hệ thống tự động phân luồng câu hỏi và giải đáp yêu cầu trực tuyến**".

Luận văn đặt ra mục tiêu nghiên cứu các mô hình có thể phát sinh văn bản, sử dụng các mạng học sâu Deep Neural Networks, dựa trên khung làm việc sequence-to-sequence, để huấn luyện trên tập dữ liệu câu hỏi và trả lời tại trường Đại học Công nghiệp Hà Nội. Từ đó xây dựng, cài đặt và vận hành một mô hình trả lời tự động với mục tiêu của đề tài là tiết kiệm được nhân lực và thời gian trong quá trình tiếp nhận, và giải quyết các yêu cầu của học sinh - sinh viên trong trường.

3. Cấu trúc của luận văn

Để mô tả kết quả nghiên cứu, luận văn được chia thành 4 chương với các nội dung như sau:

CHƯƠNG 1: Tổng quan về hệ thống trả lời tự động

CHƯƠNG 2: Cơ sở mạng nơ-ron nhân tạo CHƯƠNG 3: Ứng dụng mô hình mạng nơ-ron vào trả lời tự động

CHƯƠNG 4: Xây dựng hệ thống trao đổi thông tin trực tuyến giữa sinh viên với nhà trường tại trường đại học công nghiệp hà nội

CHƯƠNG 1: TỔNG QUAN VỀ HỆ THỐNG TRẢ LỜI TỰ ĐỘNG

Bài toán xây dựng hệ thống hỏi đáp là một bài toán khó thuộc lĩnh vực xử lý ngôn ngữ tự nhiên. Chúng ta biết rằng ngôn ngữ tự nhiên vốn nhập nhằng, đa nghĩa, việc xác định được ngữ nghĩa của câu hỏi cũng như phát hiện ra câu trả lời là một thách thức không nhỏ. Không những vậy, giữa câu hỏi và câu trả lời còn tồn tại các quan hệ "ngầm" hay phụ thuộc vào ngữ cảnh. Bài toán đặt ra nhiều thách thức để phát hiện ra được câu trả lời phù hợp nhất, thông tin hữu ích nhất. Chương này sẽ giới thiệu tổng quan về hệ thống trả lời tự động, tìm hiểu các nghiên cứu ở trong và ngoài nước để thấy được tình hình nghiên cứu và các phương pháp tiếp cận của các nghiên cứu trước đây.

1.1 Hệ thống trả lời tự động

1.2 Tình hình nghiên cứu trong và ngoài nước

1.3 Phân loại các mô hình trả lời tự động

1.3.1 Phân loại theo miền ứng dụng

1.3.4 Phân loại theo hướng tiếp cận

1.4. Các bước chung của hệ thống hỏi đáp tự động

CHƯƠNG 2: CƠ SỞ MẠNG NƠ-RON NHÂN TẠO

Chương này giới thiệu về cơ sở lý thuyết về mạng nơ rơn nhân tạo (ANN), cách thức hoạt động của mạng nơ-ron, phiên bản mở rộng của mạng nơ-ron nhân tạo RNN - Recurrent Neural Network (Mạng nơ-ron tái phát). Mạng nơ-ron tái phát RNN là một trong những mô hình Deep learning được đánh giá có nhiều ưu điểm trong các tác vụ xử lý ngôn ngữ tự nhiên. Đây cũng là cơ sở chính để thực hiện xây dựng mô hình trả lời tự động trong đề tài luận văn.

2.1 Kiến trúc mạng nơ-ron nhân tạo

2.2 Hoạt động của mạng nơ-ron nhân tạo

2.3 Mạng nơ-ron tái phát và ứng dụng

Mạng nơ-ron tái phát Recurrent Neural Network (RNN) là một trong những mô hình Deep learning được đánh giá có nhiều ưu điểm trong các tác

vụ xử lý ngôn ngữ tự nhiên (NLP). Trong phần này, tôi sẽ trình bày các khái niệm, các đặc điểm cũng như những ứng dụng của RNNs trong các bài toán thực tế.

2.3.1 Mạng nơ-ron tái phát

2.3.2 Các ứng dụng của RNN

2.3.3 Huấn luyện mạng

2.3.4 Các phiên bản mở rộng của RNN

CHƯƠNG 3: ỨNG DỤNG MÔ HÌNH MẠNG NƠ-RON VÀO TRẢ LỜI TỰ ĐỘNG

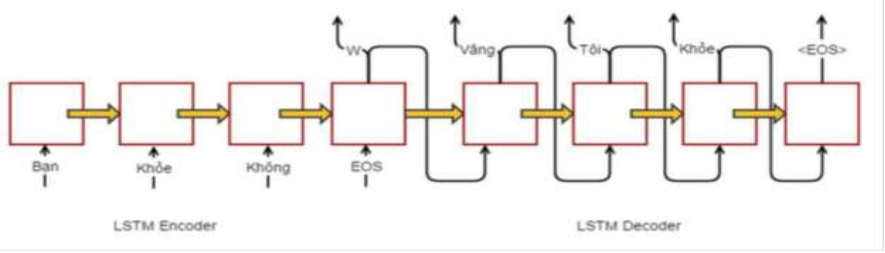
Chương này tác giả tập trung giới thiệu về mô hình mạng nơ ron có thể sản sinh ra văn bản sau khi được huấn luyện, đồng thời đề cập đến mô hình chuỗi tuần tự liên tiếp sequence to sequence. Từ đó đưa ra cách thức ứng dụng mạng nơ ron để xây dựng được một hệ thống trả lời tự động.

3.1 Phát sinh ngôn ngữ trả lời tự động

3.2 Mô hình chuỗi tuần tự liên tiếp

3.3 Mô hình trả lời tự động

Bản thân mô hình seq2seq nó bao gồm hai mạng RNN: Một cho bộ mã hóa, và một cho bộ giải mã. Bộ mã hóa nhận một chuỗi (câu) đầu vào và xử lý một phần tử (từ trong câu) tại mỗi bước. Mục tiêu của nó là chuyển đổi một chuỗi các phần tử vào một vectơ đặc trưng có kích thước cố định mà nó chỉ mã hóa thông tin quan trọng trong chuỗi và bỏ qua các thông tin không cần thiết. Có thể hình dung luồng dữ liệu trong bộ mã hóa dọc theo trục thời gian, giống như dòng chảy thông tin cục bộ từ một phần tử kết thúc của chuỗi sang chuỗi khác.



Hình 3.4: Mô hình đối thoại seq2seq.

Mỗi trạng thái ẩn ảnh hưởng đến trạng thái ẩn tiếp theo và trạng thái ẩn cuối cùng được xem như tích lũy tóm tắt về chuỗi. Trạng thái này được

gọi là bối cảnh hay vevtơ suy diễn, vì nó đại diện cho ý định của chuỗi. Từ bối cảnh đó, các bộ giải mã tạo ra một chuỗi, một phần tử (word) tại một thời điểm. Ở đây, tại mỗi bước, các bộ giải mã bị ảnh hưởng bởi bối cảnh và các phần tử được sinh ra trước đó.

3.4 Một số đặc điểm khi xây dựng hệ thống trả lời tự động

Có một số thách thức thể hiện một cách rõ ràng hoặc không thể thấy rõ khi xây dựng một mô hình đối thoại nói chung đang là tâm điểm được chú ý bởi nhiều nhà nghiên cứu.

3.4.1. Phụ thuộc bối cảnh

Để sinh ra các câu trả lời hợp lý, các hệ thống đối thoại cần phải kết hợp với cả hai bối cảnh ngôn ngữ và bối cảnh vật lý. Trong các hội thoại dài, người nói cần theo dõi và nhớ được những gì đã được nói và những thông tin gì đã được trao đổi. Đó là một ví dụ về bối cảnh ngôn ngữ. Phương pháp tiếp cận phổ biến nhất là nhúng cuộc hội thoại vào một Vector, nhưng việc làm này đối với đoạn hội thoại dài là một thách thức lớn. Các thử nghiệm trong nghiên cứu [3], [15] đều đi theo hướng này. Hướng nghiên cứu này cần kết hợp các loại bối cảnh như: Ngày/ giờ, địa điểm, hoặc thông tin về một người.

3.4.2. Kết hợp tính cách

Khi phát sinh các câu trả lời, các hệ thống trợ lý ảo lý tưởng là tạo ra câu trả lời phù hợp với ngữ nghĩa đầu vào cần nhất quán giống nhau. Ví dụ, chúng ta muốn nhận được câu trả lời với mẫu hỏi "*Bạn bao nhiêu tuổi*" hay "*Tuổi của bạn là mấy*". Điều này nghe có vẻ đơn giản, nhưng việc tổng hợp, tích hợp các kiến thức nhất quán hay "có tính cách" vào trong các mô hình đối thoại là một vấn đề rất khó để nghiên cứu.

Rất nhiều các hệ thống được huấn luyện để trả lời câu hỏi thỏa đáng với ngôn ngữ, nhưng chúng không được huấn luyện để sinh ra các câu trả lời nhất quán về ngữ nghĩa. Mô hình như thế đang được nghiên cứu trong [10], tạo ra những bước đầu tiên tập trung vào hướng mô hình hóa tính cách.

3.5 Các vấn đề khó khăn khi trả lời tự động bằng Tiếng Việt

Theo tác giả vấn đề khó khăn nhất khi xây dựng một hệ thống trả lời tự động đó là vấn đề xử lý Tiếng Việt. Tiếng Việt thuộc ngôn ngữ đơn lập, tức là mỗi một tiếng (âm tiết) được phát âm tách rời nhau và được thể hiện bằng một chữ viết. Đặc điểm này thể hiện rõ rệt ở tất cả các mặt ngữ âm, từ

vựng, ngữ pháp. Dưới đây trình bày một số đặc điểm của tiếng Việt theo các tác giả ở Trung tâm ngôn ngữ học Việt Nam đã trình bày [30].

3.5.1 Đặc điểm ngữ âm

Tiếng Việt có một loại đơn vị đặc biệt gọi là "tiếng", về mặt ngữ âm, mỗi tiếng là một âm tiết. Hệ thống âm vị tiếng Việt phong phú và có tính cân đối, tạo ra tiềm năng của ngữ âm tiếng Việt trong việc thể hiện các đơn vị có nghĩa. Nhiều từ tượng hình, tượng thanh có giá trị gợi tả đặc sắc. Khi tạo câu, tạo lời, người Việt rất chú ý đến sự hài hoà về ngữ âm, đến nhạc điệu của câu văn.

3.5.2 Đặc điểm từ vựng:

Mỗi tiếng nói chung là một yếu tố có nghĩa. Tiếng là đơn vị cơ sở của hệ thống các đơn vị có nghĩa của tiếng Việt. Từ tiếng, người ta tạo ra các đơn vị từ vựng khác để định danh sự vật, hiện tượng..., chủ yếu nhờ phương thức ghép và phương thức láy.

Việc tạo ra các đơn vị từ vựng ở phương thức ghép luôn chịu sự chi phối của quy luật kết hợp ngữ nghĩa, ví dụ: sinh viên, đất nước, máy bay, nhà lầu xe hơi, nhà tan cửa nát... Hiện nay, đây là phương thức chủ yếu để sản sinh ra các đơn vị từ vựng. Theo phương thức này, tiếng Việt triệt để sử dụng các yếu tố cấu tạo từ thuần Việt hay vay mượn từ các ngôn ngữ khác để tạo ra các từ, ngữ mới, ví dụ như *tiếp thị*, *karaoke*, *thư điện tử* (e-mail), *thư thoại* (voice mail), *phiên bản* (version), *xa lộ thông tin*, *siêu liên kết văn bản*, *truy cập ngẫu nhiên*, v.v.

Việc tạo ra các đơn vị từ vựng ở phương thức láy thì quy luật phối hợp ngữ âm chi phối chủ yếu việc tạo ra các đơn vị từ vựng, chẳng hạn như *chôm chỉa*, *chỏng chơ*, *đỏng đa đỏng đảnh*, *thơ thẩn*, *lúng lá lúng liếng*, v.v.

Vốn từ vựng tối thiểu của tiếng Việt phần lớn là các từ đơn tiết (một âm tiết, một tiếng). Sự linh hoạt trong sử dụng, việc tạo ra các từ ngữ mới một cách dễ dàng đã tạo điều kiện thuận lợi cho sự phát triển vốn từ, vừa phong phú về số lượng, vừa đa dạng trong hoạt động. Cùng một sự vật, hiện tượng, một hoạt động hay một đặc trưng, có thể có nhiều từ ngữ khác nhau biểu thị. Tiềm năng của vốn từ ngữ tiếng Việt được phát huy cao độ trong các phong cách chức năng ngôn ngữ, đặc biệt là trong phong cách ngôn ngữ nghệ thuật. Hiện nay, do sự phát triển vượt bậc của khoa học-kĩ thuật, đặc

biệt là công nghệ thông tin, thì tiềm năng đó còn được phát huy mạnh mẽ hơn.

3.5.3 Đặc điểm ngữ pháp

Từ của tiếng Việt không biến đổi hình thái. Đặc điểm này sẽ chi phối các đặc điểm ngữ pháp khác. Khi từ kết hợp từ thành các kết cấu như ngữ, câu, tiếng Việt rất coi trọng phương thức trật tự từ và hư từ.

Việc sắp xếp các từ theo một trật tự nhất định là cách chủ yếu để biểu thị các quan hệ cú pháp. Trong tiếng Việt khi nói "Sinh viên học giỏi" là khác với "Học giỏi sinh viên". Khi các từ cùng loại kết hợp với nhau theo quan hệ chính phụ thì từ đứng trước giữ vai trò chính, từ đứng sau giữ vai trò phụ. Nhờ trật tự kết hợp của từ mà "củ cải" khác với "cải củ", "tình cảm" khác với "cảm tình". Trật tự chủ ngữ đứng trước, vị ngữ đứng sau là trật tự phổ biến của kết cấu câu tiếng Việt.

Phương thức hư từ cũng là phương thức ngữ pháp chủ yếu của tiếng Việt. Nhờ hư từ mà tổ hợp "anh của em" khác với tổ hợp "anh và em", "anh vì em". Hư từ cùng với trật tự từ cho phép tiếng Việt tạo ra nhiều câu cùng có nội dung thông báo cơ bản như nhau nhưng khác nhau về sắc thái biểu cảm. Ví dụ, so sánh các câu sau đây:

- Tôi đang học bài.
- Bài, tôi đang học.
- Bài, tôi cũng đang học.

Ngoài trật tự từ và hư từ, tiếng Việt còn sử dụng phương thức ngữ điệu. Ngữ điệu giữ vai trò trong việc biểu hiện quan hệ cú pháp của các yếu tố trong câu, nhờ đó nhằm đưa ra nội dung muốn thông báo. Trên văn bản, ngữ điệu thường được biểu hiện bằng dấu câu. Sự khác nhau trong nội dung thông báo được nhận biệt khi so sánh hai câu sau:

- Đêm hôm qua, cầu gãy.
- Đêm hôm, qua cầu gãy.

Kết luận: Qua một số đặc điểm nổi bật vừa nêu trên đây, chúng ta có thể hình dung được phần nào bản sắc và tiềm năng của tiếng Việt.

CHƯƠNG 4: XÂY DỰNG HỆ THỐNG TRAO ĐỔI THÔNG TIN TRỰC TUYẾN GIỮA SINH VIÊN VỚI NHÀ TRƯỜNG TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI

4.1 Lựa chọn bài toán

Trường Đại học Công nghiệp Hà Nội (ĐHCNHN) hiện tại đang đào tạo trên 60000 sinh viên với nhiều ngành nghề đào tạo (Tiến sĩ, Thạc sĩ, Đại học chính quy, Cao đẳng , Trung cấp chuyên nghiệp, Đào tạo Sau đại học, Đào tạo nghề), với 3 cơ sở chính đào tạo có vị trí cách xa nhau Cơ sở 1 (Số 298 đường Cầu Diễn, quận Bắc Từ Liêm, thành phố Hà Nội), Cơ sở 2 (Phường Tây Tựu, quận Bắc Từ Liêm, thành phố Hà Nội), Cơ sở 3 (Phường Lê Hồng Phong và xã Phù Vân, thành phố Phủ Lý, tỉnh Hà Nam) và có hơn 30 cơ sở liên kết đào tạo ngoài trường.

Để nâng cao chất lượng giảng dạy của cán bộ, giáo viên cũng như kết quả học tập của học sinh, sinh viên trong trường nhà trường đã đầu tư xây dựng một cổng thông tin điện tử nhằm giúp sinh viên tra cứu thông tin và gửi thắc mắc liên quan đến quá trình học tập và rèn luyện quả sinh viên qua mạng internet. Tuy nhiên việc giải đáp thắc mắc của toàn bộ sinh viên gặp phải một số khó khăn do hiện tại bộ phận trả lời được nằm tại nhiều cơ sở, nhiều phòng ban, sinh viên chủ yếu sử dụng các kênh thông tin không chính thức như Facebook, gây nên hiện tượng không tìm được câu trả lời thỏa đáng. Nhu cầu giải đáp phục vụ cho quá trình nghiên cứu và học tập của sinh viên còn gặp nhiều khó khăn nên trường ĐHCNHN đã xây dựng hệ thống giải đáp trực tuyến nhằm giúp giải đáp sinh viên một cách nhanh chóng và thiết thực.

Việc tin học hóa cổng hỏi đáp đã giúp việc quản lý việc học tập và trao đổi trong nhà trường trở nên thuận tiện hơn, giúp cán bộ, giáo viên, học sinh, sinh viên trong trường giải quyết được những thắc mắc giúp học tập đạt kết quả tốt hơn do đó yêu cầu đặt ra cần phải xây dựng một hệ thống trao đổi thông tin trực tuyến trong nhà trường có thể tự động phân luồng câu hỏi một cách chính xác từ người hỏi đến đúng người có khả năng trả lời là cấp thiết

4.2 Quy trình trao đổi thông tin (hỏi đáp trực tuyến) giữa HSSV với Nhà trường tại Trường Đại học Công nghiệp Hà Nội

4.2.1 Quy trình áp dụng

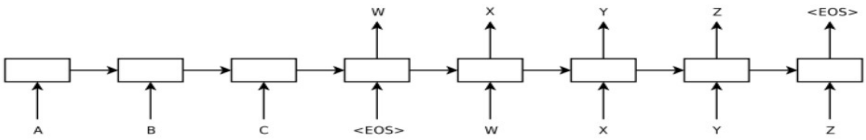
4.2.2 Mô tả quy trình áp dụng

4.3 Kiến trúc ứng dụng

Mạng học sâu DNN - Deep Neural Networks là cách tiếp cận hiện đại của các thuật toán học máy. Các mô hình học sâu có kiến trúc tương tự mạng Nơ-ron nhưng khác về cách tiếp cận vấn đề, với ý tưởng cơ bản là dữ liệu tại mỗi lớp sẽ có mức độ trừu tượng hóa (khái quát) cao hơn bằng cách tổ hợp các dữ liệu có mức trừu tượng hóa thấp ở lớp trước.

DNN rất mạnh bởi vì chúng có thể thực hiện tính toán song song tùy ý với một số lượng rất ít các bước. Hơn nữa, Mạng DNN lớn có thể được huấn luyện với lan truyền ngược giám sát bất cứ khi nào tập huấn luyện được dán nhãn có đủ thông tin để xác định các thông số của mạng. Do đó, nếu có tồn tại một thiết lập thông số của một DNN lớn mà đạt được kết quả tốt, giám sát lan truyền ngược sẽ tìm thấy những thông số và giải quyết vấn đề.

Trong luận văn này, chúng tôi cho thiết kế một ứng dụng đơn giản dựa vào kiến trúc Long ShortTerm Memory (LSTM) [33] có thể giải quyết các vấn đề chuỗi tuần tự liên tiếp sequence-to-sequence. Ý tưởng là sử dụng một mạng LSTM để đọc chuỗi đầu vào, một bước thời gian tại một thời điểm, để có được biểu diễn vector kích thước cố định, và sau đó sử dụng một mạng LSTM để trích xuất các trình tự đầu ra từ vector đó (hình 4.2). Mạng LSTM thứ hai về cơ bản là một mạng nơ-ron tái phát dựa trên mô hình ngôn ngữ [34, 35], ngoại trừ việc nó được bổ sung thêm các điều kiện trên các chuỗi đầu vào. LSTM có khả năng học thành công trên dữ liệu phụ thuộc thời gian tầm xa, làm cho nó trở thành một sự lựa chọn tự nhiên cho ứng dụng này do độ có trễ thời gian đáng kể giữa các đầu vào và đầu ra tương ứng của chúng (hình 4.2).



Hình 4.2: Kiến trúc mô hình đối thoại cho tiếng Việt

Kiến trúc mô hình trên chúng tôi dựa vào kết quả nghiên cứu của Lê Viết Quốc cho bài toán hỏi đáp bằng tiếng Anh, trong [11], chúng tôi cũng sử dụng mô hình này sẽ đọc một câu đầu vào tiếng Việt, ví dụ: “A B C” và sinh ra ra một câu tiếng Việt đầu ra “W X Y Z”. Mô hình sẽ dừng dự đoán sau khi sản xuất ra một mã hiệu kết thúc câu <EOS>. Lưu ý, mạng LSTM

đọc câu đầu vào theo hướng ngược lại, bởi vì làm như vậy sẽ đưa ra nhiều các phụ thuộc ngắn hạn trong các dữ liệu mà làm cho các vấn đề được tối ưu hơn nhiều.

Tiếp cận của chúng tôi sử dụng một khung làm việc sequence-to-sequence (seq2seq) được mô tả trong [10]. Mô hình này dựa trên một mạng nơ-ron tái phát, mà sẽ đọc chuỗi đầu vào tuần tự, một dấu hiệu (token) tại mỗi thời điểm, và dự đoán chuỗi đầu ra, cũng một dấu hiệu tại một thời điểm. Trong suốt thời gian huấn luyện, chuỗi tuần tự đầu ra được đưa vào mô hình, và việc học có thể hoàn tất bởi quá trình lan truyền ngược. Mô hình này được huấn luyện để cực đại hóa cross entropy theo đúng tuần tự cho bối cảnh của nó. Trong quá trình suy luận, mô hình cho chuỗi đầu ra đúng mà không quan sát được, bằng cách đơn giản chúng tôi nạp vào dấu hiệu token đã được dự đoán làm đầu vào để dự đoán dấu hiệu đầu ra tiếp theo. Đây là một phương pháp suy luận "tham lam". Một cách tiếp cận ít tham lam sẽ được sử dụng tìm kiếm Beam Search, đây là thuật toán tìm kiếm mà có thể phát hiện ra một đồ thị bằng việc mở rộng các nút tiềm năng trong một tập có giới hạn, bằng cách nạp một vài ứng cử viên ở các bước trước vào bước tiếp theo. Một chuỗi được dự đoán có thể được chọn dựa trên xác suất của chuỗi.

Cụ thể, giả sử rằng chúng ta quan sát một cuộc trò chuyện với hai lượt: người đầu tiên thốt ra “A B C”, và người thứ hai trả lời “W X Y Z”. Chúng tôi có thể sử dụng một mạng nơ-ron tái phát, và huấn luyện để ánh xạ “ABC” sang “WXYZ” như trên hình 4.2 ở trên. Các trạng thái ẩn của mô hình khi đó nhận được ký tự kết thúc chuỗi <EOS>, có thể được xem như là vector ngưỡng uy nghĩ vì nó lưu trữ các thông tin của câu, hoặc nghĩ, “A B C”.

Thế mạnh của mô hình này nằm ở sự đơn giản và tính tổng quát của nó. Chúng ta có thể sử dụng mô hình này cho Máy dịch, Hỏi đáp, và các cuộc trò chuyện mà không cần thay đổi nhiều trong kiến trúc. Việc áp dụng kỹ thuật này để mô hình hóa cuộc đối thoại cũng rất đơn giản: các chuỗi đầu vào có thể được nối bối cảnh đã được trò chuyện với chuỗi đầu ra là câu trả lời.

Không giống như các nhiệm vụ đơn giản hơn như dịch thuật, tuy nhiên, một mô hình như sequence-to-sequence sẽ không thể “giải quyết” thành công vấn đề của việc mô hình hóa đối thoại do: các hàm mục tiêu được

tối ưu hóa không nắm bắt được mục tiêu thực tế cần đạt được thông qua giao tiếp của con người, mà thường là thông tin dài hạn và dựa trên trao đổi thông tin chứ không phải là dự đoán bước tiếp theo. Việc thiếu một mô hình để đảm bảo tính thống nhất và kiến thức nói chung cũng là một hạn chế rõ ràng của một mô hình hoàn toàn không có giám sát.

4.4 Cài đặt hệ thống

4.4.1 Mô hình cài đặt

Mạng nơ-ron tái phát RNN [36, 37] là một mạng tổng quát của các mạng nơ-ron truyền thẳng cho các chuỗi tuần tự. Với mỗi chuỗi đầu vào $(x_1, \dots, x_T)$, là một mạng RNN chuẩn sẽ tính toán một chuỗi các kết quả đầu ra $(y_1, \dots, y_T)$, bằng cách duyệt phương trình sau:

$$h_t = sigm(W^{hx}x_t + W^{hh}h_{t-1})$$

$$y_t = W^{yh}h_t$$

Mạng RNN có thể dễ dàng ánh xạ tuần tự chuỗi bất cứ khi nào sự liên kết giữa đầu vào và đầu ra được biết đến trước khi hết hạn. Tuy nhiên, nó không là cách rõ ràng để áp dụng một mạng RNN cho các vấn đề mà đầu vào và đầu ra có độ dài khác nhau với các mối quan hệ phức tạp và không đơn điệu (thay đổi). Cách làm đơn giản nhất cho việc học chuỗi nói chung là ánh xạ chuỗi đầu vào thành một vector có kích thước cố định sử dụng một mạng RNN và sau đó, ánh xạ vector vào chuỗi đích sử dụng một mạng RNN khác (cách làm này được thực hiện bởi Kyunghyun Cho và cộng sự [36]). Trong khi nó có thể hoạt động trên nguyên tắc kể từ khi RNN được cung cấp với tất cả các thông tin liên quan, nó sẽ gặp khó khăn trong việc huấn luyện do sự phụ thuộc thời gian dài [33, 38]. Tuy nhiên, mạng LSTM [33] có thể học các vấn đề phụ thuộc thời gian dài, vì vậy, sử dụng mạng LSTM có thể thành công trong trường hợp này.

Mục tiêu của LSTM là để ước lượng xác suất có điều kiện $p(y_1, \dots, y_T, |x_1, \dots, x_T)$ trong đó $(x_1, \dots, x_T)$ là một chuỗi đầu vào và $(y_1, \dots, y_{T'})$ là chuỗi đầu ra tương ứng của nó có chiều dài T'có thể khác nhau từ T. Mạng LSTM tính xác suất có điều kiện này bằng cách có được thông tin đại diện mà số chiều cố định ū của chuỗi đầu vào $(x_1, \dots, x_T)$ được tính bởi các trạng thái ẩn cuối cùng của mạng LSTM, và sau đó tính toán xác suất của $(y_1, \dots, y_{T'})$ ới một công thức LSTM-LM tiêu chuẩn mà ban đầu trạng thái ẩn được thiết lập để đại diện ū của $(x_1, \dots, x_T)$:

$$p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \dots, y_{t-1})$$

Trong phương trình này, mỗi phân phối xác suất $p(y_1, \dots, y_T, | x_1, \dots, x_T)$ được biểu diễn bởi một hàm softmax trên tất cả từ trong từ vựng. Sử dụng công thức LSTM của Graves, trong [37]. Chú ý là mỗi câu kết thúc với một ký hiệu đặc biệt end-of-sentence "<EOS>", cho phép mô hình để xác định một phân phối các chuỗi của tất cả các độ dài có thể. Xem lược đồ tổng quát trong hình 4.2, trong đó LSTM tính xác suất đại diện của "A", "B", "C", "<EOS>" và sau đó sử dụng đại diện này để tính xác xuất của "W", "X", "Y", "Z", "<EOS>".

4.4.2 Môi trường cài đặt

4.4.3 Công cụ cài đặt

4.5 Kết quả đạt được

4.5.1 Một số kết quả

4.5.2 Hiệu năng

KẾT LUẬN

1. Những kết quả chính của luận văn:

Luận văn đã đạt được các kết quả chính sau đây:

- Đưa ra được các lý thuyết và vấn đề gặp phải trong việc xây dựng hệ thống trả lời tự động trực tuyến.
- Ứng dụng mạng học sâu vào giải quyết bài toán phân luồng câu hỏi và trả lời câu hỏi tự động trực tuyến
- Cài đặt hệ thống trả lời câu hỏi tự động trên cở sở mô hình mạng học sâu đã lựa chọn với kết quả thực nghiệm đạt trên 50% hài lòng.
- Phần mềm đưa vào ứng dụng giúp tiết kiệm chi phí cho nguồn nhân lực trong quá trình tiếp nhận và trả lời câu hỏi.
- Tổng hợp các kết quả và hướng nghiên cứu về bài toán đã có thể đưa ra được trợ lý ảo tiếp nhận và hiểu được nhu cầu của sinh viên.
- Có khả năng áp dụng vào các hệ thống tự động hỏi đáp khác như tư vấn bán hàng, tư vấn sức khỏe, ...

2. Hướng phát triển của luận văn:

- Tiếp tục triển khai mở rộng và thu thập nhiều câu hỏi hơn ở nhiều trường Đại học để có thể gia tăng sự huấn luyện, tăng độ chính xác.

- Tiếp tục nghiên cứu các mô hình mạng giải quyết bài toán phân luồng câu hỏi và trả lời yêu cầu trực tuyến.
- Tìm hiểu nhu cầu thực tế, cũng như tham khảo các ý kiến của chuyên gia để xây dựng chương trình áp dụng kỹ thuật đã nghiên cứu, bổ sung một số yếu tố khác để hoàn thiện hệ thống trả lời tự động đạt hiệu quả cao.

TÀI LIỆU THAM KHẢO

[1] Nhữ Bảo Vũ, Nguyễn Văn Nam. XÂY DỰNG MÔ HÌNH ĐỐI THOẠI CHO TIẾNG VIỆT TRÊN MIỀN MỞ DỰA VÀO PHƯƠNG PHÁP HỌC CHUỖI LIÊN TIẾP. Khóa luận tốt nghiệp thạc sỹ CNTT 2016

[2] Hồ Tú Bảo, Lương Chi Mai. Về xử lý tiếng Việt trong công nghệ thông tin, Viện Công nghệ Thông tin, Viện Khoa học và Công nghệ Tiên tiến Nhật bản.

[3] Hà Quang Thụy & nhóm khai phá dữ liệu và ứng dụng. Bài giảng về khai phá dữ liệu, 2007

[4] Walter S. Lasecki, Ece Kamar, Dan Bohus, Conversations in the Crowd: Collecting Data for Task-Oriented Dialog Learning

[5] Alan Ritter, Colin Cherry, and Bill Dolan. 2010. “Unsupervised modeling of twitter conversations”. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT ’10, pages 17

[6] Rafael E. Banchs and Haizhou Li. 2012. “Iris: a chat-oriented dialogue system based on the vector space model”. In Proceedings of the ACL 2012 System Demonstrations, pages 37–42, Jeju Island, Korea, July. Association for Computational Linguistics.

[7] Karthik Narasimhan, Tejas Kulkarni, and Regina Barzilay. 2015. “Language understanding for text-based games using deep reinforcement learning”. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1–11, Lisbon,

[8] T.-H. Wen, D. Vandyke, N. Mrksic, M. Gasic, L. M. Rojas-Barahona, P.-H. Su, S. Ultes, and S. Young. 2016. A Network-based

End-to-End Trainable Task-oriented Dialogue System. ArXiv eprints, April
[9] Heriberto Cuayahuitl. 2016. Simpleds: “A simple deep reinforcement learning dialogue system”. CoRR, abs/1601.04574
[10] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, 14 Dec 2014. “Sequence to Sequence Learning with Neural Networks” pp. 1–9.
[11] Oriol Vinyals, Quoc Le, 22 Jul 2015. “A Neural Conversational Model”
[12] Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, Bill Dolan, 22 Jun 2015. “A Neural Network Approach to Context-Sensitive Generation of Conversational Responses”
[13] Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, Joelle Pineau, 6 Apr 2016. “Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models”.
[14] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversitypromoting objective function for neural conversation models. arXiv preprint arXiv:1510.03055
[15] Lester, J., Branting, K., and Mott, B, 2004. “Conversational agents. In Handbook of Internet Computing. Chapman & Hall”.
[16] Will, T, 2007. “Creating a Dynamic Speech Dialogue”. VDM Verlag Dr.
[17] Russell, S., Dewey, D., Tegmark, M. (2015). “Research Priorities for Robust and Beneficial Artificial Intelligence”. AI Magazine, 36 (4):105–114.
[18] Alan M Turing. 1950. “Computing machinery and intelligence”. Mind, 59(236):433–460.
[19] Joseph Weizenbaum. 1966. “Elizaa computer program for the study of natural language communication between man and machine”. Communications of the ACM, 9(1):36–45.
[20] Roger C Parkinson, Kenneth Mark Colby, and William S Faught. 1977. “Conversational language comprehension using integrated pattern-matching and parsing”. Artificial Intelligence, 9(2):111–134.
[21] Richard S Wallace. 2009. “The anatomy of ALICE”. Springer.

[22] Jurgen Schmidhuber. 2015. “Deep learning in neural networks: An overview. Neural Networks”, 61:85–117.
[23] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature, 521(7553):436–444.
[24] Marilyn Walker, Grace Lin, and Jennifer Sawyer. 2012. “An annotated corpus of film dialogue for learning and characterizing character style”. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Do ˘ gan, Bente Maegaard, Joseph ˘ Mariani, Jan Odijk, and Stelios Piperidis, editors, Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), pages 1373–1378, Istanbul, Turkey, May. European Language Resources Association (ELRA). ACL Anthology Identifier: L12-1657.
[25] Francesca Bonin, Jose San Pedro, and Nuria Oliver. 2014. “A context-aware nlp approach for noteworthiness detection in cellphone conversations”. In COLING, pages 25–36.
[26] Jaime Carbonell, Donna Harman, Eduard Hovy, and Steve Maiorano, John Prange and Karen Sparck-Jones. Vision Statement to Guide Research in Question & Answering (Q&A) and Text Summarization. Final version 1. 2000
[27] P. Werbos, 1990. “Backpropagation through time: what it does and how to do it”. Proceedings of IEEE.
[28] Sanda M. Harabagiu, Marius A. Paşca, Steven J. Maiorano. *Experiments with open-domain textual Question Answering*. International Conference On Computational Linguistics Proceedings of the 18th conference on Computational linguistics – Volume 1, 2000, tr. 292 - 298
[29] Eduard Hovy, Ulf Hermjakob and Lin, C.-Y. *The Use of External Knowledge in Factoid QA*. Paper presented at the Tenth Text REtrieval Conference (TREC 10), Gaithersburg, MD, 2001, November 13-16.
[30] Trung tâm ngôn ngữ học Việt Nam. “Đặc điểm tiếng Việt”, http://www.vietlex.com/vietnamese.htm
[31] S. Hochreiter and J. Schmidhuber, 1997. “Long Short-Term Memory” Neural Computation, vol. 9, pp. 1735–1780.

[32] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, and S. Khudanpur, 2010. “Recurrent neural network based language model”. In INTERSPEECH,
pages 1045–1048.
[33] M. Sundermeyer, R. Schluter, and H. Ney, 2010. “LSTM neural networks for language modeling”. In INTERSPEECH.
[34] D. Rumelhart, G. E. Hinton, and R. J. Williams, 1986. “Learning representations by back-propagating errors”. Nature, 323(6088):533–536
[35] Y. Bengio, P. Simard, and P. Frasconi, 1994. “Learning long-term dependencies with gradient descent is difficult”. IEEE Transactions on Neural Networks, 5(2):157–166.
[36] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, Sep 2014. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”.
[37] A. Graves, 5 Jun 2014. “Generating sequences with recurrent neural networks”. In Arxiv preprint arXiv:1308.0850.