

BIÊN BẢN
HỌP HỘI ĐỒNG CHẤM ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Căn cứ quyết định số Quyết định số 289/QĐ-HV ngày 06 tháng 03 năm 2024 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi 18 giờ 30 phút, ngày 20 tháng 03 năm 2024 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: **Nguyễn Đình Tuấn**

Tên đề án tốt nghiệp: **Tóm tắt tin tức Tiếng Việt sử dụng mô hình BERT**

Chuyên ngành: **Khoa học máy tính** Mã số: **8.48.01.01**

Các thành viên của Hội đồng chấm đề án tốt nghiệp có mặt: 5 / 05

| TT | HỌ VÀ TÊN | TRÁCH NHIỆM TRONG HĐ | GHI CHÚ |
|----|----------------------------------|----------------------|---------|
| 1 | GS.TS. Từ Minh Phương | Chủ tịch | |
| 2 | TS. Trần Tiến Công | Thư ký | |
| 3 | PGS.TS. Nguyễn Long Giang | Phản biện 1 | |
| 4 | PGS.TS. Phan Xuân Hiếu | Phản biện 2 | |
| 5 | PGS.TS. Hoàng Xuân Dậu | Ủy viên | |

Các nội dung thực hiện:

- Chủ tịch Hội đồng điều khiển buổi họp. Công bố quyết định của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ.
- Người hướng dẫn khoa học hoặc thư ký đọc lý lịch khoa học và các điều kiện bảo vệ đề án tốt nghiệp của học viên (Có bản lý lịch khoa học và kết quả các môn học cao học của học viên kèm theo).
- Học viên trình bày tóm tắt đề án tốt nghiệp.
- Phản biện 1 đọc nhận xét (có văn bản kèm theo)
- Phản biện 2 đọc nhận xét (có văn bản kèm theo)
- Các câu hỏi của thành viên Hội đồng:

Nêu rõ vai trò của các khối sử dụng trong mô hình bài toán
sử dụng để đo khoảng cách: cosine để so sánh giữa 2 vector hoặc
2 phân bố
Khi nào có sử dụng mô hình ngôn ngữ lớn hay không?
Có so sánh hay không?

7. Trả lời của học viên:

Học viên giải thích chi tiết về các khối số đang
đang mô hình đồ xuất
Học viên tiếp thu khuyến nghị

8. Thư ký đọc nhận xét về quá trình thực hiện đề án tốt nghiệp của học viên (có văn bản kèm theo).

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:

1. Trưởng Ban kiểm phiếu: PGS.TS. Hoàng Xuân Dân

2. Ủy viên Ban kiểm phiếu: TS. Trần Tiến Công

3. Ủy viên Ban kiểm phiếu: PGS.TS. Phạm Xuân Hoàng

- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.

- Ban kiểm phiếu làm việc:

- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu (có Biên bản họp Ban kiểm phiếu kèm theo)

- Điểm trung bình của đề án tốt nghiệp: 8.9

Kết luận:

1. Các nội dung cần chỉnh sửa, hoàn thiện sau bảo vệ đề án tốt nghiệp:

Theo nhận xét của PGS.TS. Nguyễn Long Giang, học viên cần
sửa và kiểm tra lại chất lượng luận văn, chú ý đã căn cứ
lập ra đúng góp ý của đề án
chỉnh sửa luận văn từ "chính lý"

2. Đề nghị Học viện công nhận (hoặc không) và cấp bằng (hoặc không) thạc sĩ cho học viên:

3. Đề án tốt nghiệp có thể phát triển thành đề tài nghiên cứu cho

NCS.....

Buổi làm việc kết thúc vào..... cùng ngày.

Chủ tịch



GS.TS. Từ Minh Phương

Thư ký



TS. Trần Tiến Công

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập - Tự do - Hạnh phúc

=====

BẢN NHẬN XÉT LUẬN VĂN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài luận văn: Tóm tắt tin tức tiếng Việt sử dụng mô hình BERT.

Chuyên ngành: Khoa học máy tính

Mã số: 8.48.01.01

Họ và tên học viên: Nguyễn Đình Tuấn

Họ và tên người nhận xét: PGS.TS. Nguyễn Long Giang

Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn Lâm KH&CN Việt Nam

Số điện thoại: 0904739189

Email: nlgiang@ioit.ac.vn

NỘI DUNG NHẬN XÉT

1. Cơ sở khoa học và thực tiễn, sự cần thiết lựa chọn đề tài

Tóm tắt văn bản tiếng Việt là bài toán trong lĩnh vực NLP với nhiều ứng dụng trong thực tiễn. Luận văn nghiên cứu, giải quyết bài toán văn bản tóm tắt tin tức tiếng Việt theo tiếp cận tóm tắt tóm lược sử dụng các mô hình BERT và các biến thể. Chủ đề nghiên cứu của luận văn có ý nghĩa khoa học và thực tiễn.

2. Nội dung của luận văn, các kết quả đã đạt được

Luận văn đạt được kết quả chính như sau:

- 1) Tìm hiểu bài toán tóm tắt tiếng Việt và xây dựng mô hình giải quyết bài toán bằng mạng học sâu BERT và LSTM
- 2) Thử nghiệm, đánh giá mô hình trên bộ dữ liệu mẫu được thu thập và tiền xử lý từ 3 nguồn dữ liệu [2, 28, 29] nhằm đánh giá tính hiệu quả của các mô hình.

Kết quả đạt được của luận văn đáp ứng yêu cầu đối với luận văn thạc sĩ chuyên ngành khoa học máy tính.

3. Những vấn đề học viên cần giải thích thêm

- Rà soát, chỉnh sửa luận văn về trình bày (hiện tại văn dịch từ tài liệu tiếng Anh và các nguồn trên mạng khá nhiều chưa chuẩn hoá), rất nhiều chỗ vẫn đề “đề án” (trang 9, 39)., rà soát các đoạn văn, tiêu đề, trích dẫn tài liệu tham khảo, nhiều thuật ngữ tiếng Anh chưa giải thích (BLLA – trang 10)

- Phần 1.3 giải pháp đề xuất nên chuyển sang chương 3

- Sửa lại tên Chương 2, chỉ là cơ sở lý thuyết về các mô hình, chỉnh sửa lại cách trình bày ở các tiêu đề, nội dung (dịch từ tiếng Anh). Nên giới thiệu thêm một số mô hình mở rộng của BERT như phoBERT...mà luận văn sử dụng.

- Chương 3: Cần có sơ đồ tổng thể về mô hình BERT-LSTM giải quyết bài toán với hình vẽ các khối.

- Về phần thử nghiệm đánh giá, cần mô tả chi tiết hơn quá trình thu thập, tiền xử lý dữ liệu (với các công cụ) để được tập dữ liệu cuối cùng cho huấn luyện.

- Nên đưa ra thời gian thực hiện huấn luyện mô hình (pre-train) và thời gian inference mô hình (thực thi) cho dữ liệu đầu vào.

4. Kết luận

Luận văn đáp ứng các yêu cầu đối với một luận văn Thạc sĩ chuyên ngành Khoa học máy tính. Đồng ý cho phép học viên bảo vệ luận văn tốt nghiệp.

Câu hỏi: 1) Mô hình xây dựng có thể áp dụng cho văn bản bất kỳ không? Nếu thực hiện trên văn bản bất kỳ thì có gì khác biệt ?

Hà Nội, ngày 14 tháng 03 năm 2020

XÁC NHẬN CỦA CƠ QUAN

NGƯỜI NHẬN XÉT



Nguyễn Long Giang

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài đề án tốt nghiệp: Tóm tắt tin tức tiếng Việt sử dụng mô hình BERT

Chuyên ngành: Khoa học máy tính

Mã chuyên ngành: 8.48.01.01

Họ và tên học viên: Nguyễn Đình Tuấn

Họ và tên người nhận xét: Phan Xuân Hiếu

Học hàm, học vị: Phó Giáo sư, Tiến sĩ

Chuyên ngành: Khoa học thông tin và máy tính

Cơ quan công tác: Trường ĐH Công nghệ, ĐHQG Hà Nội

Số điện thoại: 0948855916

E-mail: hieupx@vnu.edu.vn

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:

Đề án này tập trung giải quyết bài toán tóm tắt tiếng Việt dựa mô hình mạng học sâu với sự kết hợp của mô hình BERT và LSTM và cơ chế attention. Học viên tiến hành thực nghiệm trên các môi trường phù hợp và sử dụng các mô hình huấn luyện trước cho tiếng Việt như PhoBERT và các công cụ xử lý tiếng Việt như VnCoreNLP ... Tuy vấn đề tóm tắt tiếng Việt không còn mới, nhưng việc sinh viên ứng dụng các tiến bộ gần đây của học sâu và NLP vào giải quyết một vấn đề cụ thể cho tiếng Việt cho thấy học viên chịu khó cập nhật kiến thức, có khả năng thực hành tốt. Đề tài của đề án vì thế có tính hiện đại và có tính thực tiễn tốt.

II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:

Đề án đề xuất một kiến trúc dựa trên các mô hình BERT, cơ chế chú ý và kiến trúc LSTM. Sự kết hợp này học viên gọi là BLLA (BERT-LSTM-LSTM). Với sự kết hợp này, học viên đã tận dụng được những ưu điểm về mã hoá có ngữ cảnh của BERT (dùng cho phía mã hoá) cũng như khả năng học cho dữ liệu chuỗi (seq2seq). LSTM được dùng cho cả phía mã hoá và phía giải mã với phía mã hoá là Bi-LSTM và phía mã hoá là LSTM.

III/ Những vấn đề cần giải thích thêm:

1/ Tại sao trong phần đầu học viên nói là huấn luyện kiến trúc với việc đóng băng BERT (trang 31) nhưng ở phần sau (Chương 3, trang 33), học viên lại đề cập “Điều chỉnh để huấn luyện BERT”?

2/ Học viên có thử đánh giá đo đạc để so sánh hiệu quả của giải pháp học viên đề xuất và của ChatGPT, Gemini hay không? Nếu có thì kết quả như thế nào?

IV/ Kết luận:

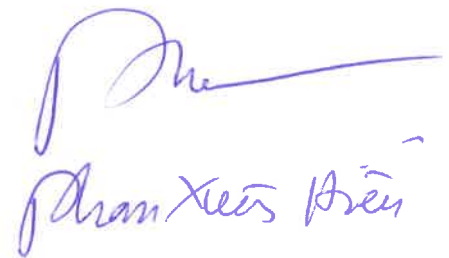
Đề án cho thấy học viên nắm bắt vấn đề tốt, và khả năng thực hành, thi hành của học viên. Đề án gồm 3 chương và mô tả đầy đủ các kiến thức lý thuyết, diễn giải và thực nghiệm.

Đồng ý cho phép học viên bảo vệ đề án tốt nghiệp.

Đồng ý (hoặc không đồng ý) cho phép học viên bảo vệ đề án tốt nghiệp.

Ngày 18 tháng 03 năm 2024

NGƯỜI NHẬN XÉT



Phan Xương Biên