TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
Tập 22, Số 2 (2025): 235-246
Website: https://journal.hcmue.edu.vn
ISSN: 2734-9918

HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
Vol. 22, No. 2 (2025): 235-246
https://doi.org/10.54607/hcmue.js.22.2.4320(2025)

Bài báo nghiên cứu

CẢI TIẾN MÔ HÌNH DỊCH MÁY MẠNG NƠ-RON ANH-VIỆT SỬ DỤNG ĐỒ THỊ TRI THỨC

***Lê Công Trí*[1], *Nguyễn Phương Nam*[1], *Nguyễn Hồng Bửu Long*[2], *Trần Thanh Nhã*[1*]**

[1]*Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam*

[2]*Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam*

[*]*Tác giả liên hệ: Trần Thanh Nhã – Email: nhatt@hcmue.edu.vn*

Ngày nhận bài: 06-6-2024; ngày nhận bài sửa: 21-10-2024; ngày duyệt đăng: 23-10-2024

TÓM TẮT

Dịch máy là một vấn đề quan trọng trong lĩnh vực xử lí ngôn ngữ tự nhiên (XLNNTN), với mục tiêu tạo ra các bản dịch từ ngôn ngữ nguồn sang ngôn ngữ đích có ý nghĩa tương đương. Tuy nhiên, các mô hình dịch máy mạng nơ-ron (Neural Machine Translation - NMT) hiện tại gặp khó khăn trong việc xử lí các thực thể, đặc biệt là với các ngôn ngữ thiếu nguồn tài nguyên chất lượng cao như tiếng Việt. Bài báo này đề xuất một phương pháp cải thiện khả năng của các mô hình NMT bằng cách tích hợp thông tin từ đồ thị tri thức (Knowledge Graph - KG) vào mô hình Transformer. Phương pháp này giúp mô hình học được biểu diễn của các thực thể trong quá trình huấn luyện, từ đó tăng cường khả năng dịch tự động khi gặp các thực thể và các yếu tố ngôn ngữ tương tự khác. Kết quả thực nghiệm cho thấy phương pháp đề xuất cải thiện đáng kể chất lượng dịch của mô hình Transformer, đặc biệt trong việc dịch các thực thể. Những kết quả này chứng minh hiệu quả của việc tích hợp đồ thị tri thức vào mô hình NMT và mở ra hướng phát triển mới cho nghiên cứu trong lĩnh vực này.

Từ khóa: BERT; xử lí ngôn ngữ tự nhiên; dịch máy; đồ thị tri thức; dịch máy mạng nơ-ron

1. Giới thiệu

Dịch máy là một vấn đề quan trọng trong lĩnh vực xử lí ngôn ngữ tự nhiên (XLNNTN). Mục tiêu của dịch máy là tạo ra các bản dịch của câu, đoạn văn, hoặc tài liệu từ ngôn ngữ nguồn sang ngôn ngữ đích với ý nghĩa tương đương. Việc này giúp giảm bớt rào cản ngôn ngữ khi tiếp cận các nguồn thông tin được trình bày bằng ngôn ngữ mà chúng ta chưa biết, hoặc chưa có nhiều kiến thức để phân tích.

Dịch máy mạng nơ-ron (Neural Machine Translation – NMT) dựa trên kiến trúc bộ mã hóa-giải mã đã trở thành một phương pháp dịch tự động tiên tiến nhờ khả năng học biểu diễn ngôn ngữ hiệu quả trên các cặp ngôn ngữ khác nhau. Các mô hình Seq2seq (Sutskever et al., 2014), Conv2seq (Gehring et al., 2017) và Transformer (Vaswani et al., 2017) có thể

Cite this article as: Le Cong Tri, Nguyen Phuong Nam, Nguyen Hong Buu Long, & Tran Thanh Nha (2025). Enhancing neural machine translation with knowledge graph integration. *Ho Chi Minh City University of Education Journal of Science, 22*(2), 235-246.

tạo ra các bản dịch sánh ngang với bản dịch của con người khi dịch giữa các ngôn ngữ có nguồn tài nguyên phong phú trong các điều kiện cụ thể. Tuy nhiên, vẫn còn những thách thức, đặc biệt là với các ngôn ngữ có ít ngữ liệu chất lượng cao sẽ làm cho hệ thống NMT giảm đi hiệu năng đáng kể như đối với tiếng Việt.

Một trong những vấn đề nghiêm trọng khi huấn luyện các mô hình dịch máy NMT trong điều kiện thiếu ngữ liệu đó là việc dịch các thực thể. Các thực thể trong câu đóng vai trò quan trọng và việc dịch chính xác các thực thể ảnh hưởng lớn đến toàn bộ chất lượng dịch của câu. Tuy nhiên, các thực thể thường không xuất hiện nhiều lần trong ngữ liệu huấn luyện, dẫn đến việc các mô hình NMT thường xem các thực thể như các từ có tần suất xuất hiện thấp (Out of Vocabulary - OOV). Do đó, các mô hình này sẽ loại bỏ các từ này ra khỏi quá trình huấn luyện, dẫn đến việc không thể dịch được các từ này. Cụ thể, các mô hình NMT chuyển tất cả các từ có tần suất xuất hiện thấp thành kí hiệu "unk" (viết tắt của "unknown") trong bước tiền xử lí trước khi tiến hành huấn luyện. Vì vậy, mô hình NMT không thể dịch được các từ này khi triển khai thực tế.

Các thực thể thông thường có thể được chia làm hai nhóm: danh từ riêng và danh từ chung. Danh từ riêng bao gồm tên người, tên địa phương, và tên tổ chức (ví dụ: Hà Nội, Tôn Đức Thắng...). Danh từ chung dùng để chỉ các nhóm đối tượng mang tính tổng quát (ví dụ: cổ phần, cổ đông...). Các thực thể này thông thường được định nghĩa trong các đồ thị tri thức (Knowledge Graph - KG) dưới dạng các bộ ba (triplet) bao gồm: một chủ thể (là một thực thể), mối quan hệ (là một liên kết) và một đối tượng (là một thực thể khác). Ví dụ, bộ ba <Hà Nội, is-a, thủ đô> có biểu diễn ngôn ngữ tự nhiên là "Hà Nội là thủ đô". Như vậy, nếu được tích hợp thông tin từ đồ thị tri thức, mô hình NMT có thể mô hình hóa các thực thể và có khả năng xử lí được các thực thể trong quá trình dịch.

Mục tiêu của bài báo là cải thiện khả năng của các mô hình NMT khi xử lí các thực thể, bằng cách tích hợp thông tin từ bên ngoài. Cụ thể, đề tài nghiên cứu các phương pháp tích hợp thông tin từ đồ thị tri thức vào mô hình dịch máy tiên tiến nhất hiện nay là Transformer (Vaswani et al., 2017). Điều này nhằm giúp mô hình có thể học được biểu diễn của các thực thể trong quá trình huấn luyện, từ đó tăng cường khả năng dịch tự động khi gặp các trường hợp như thực thể, thuật ngữ và các yếu tố ngôn ngữ tương tự khác.

Trong nội dung bài báo này, các đóng góp chính bao gồm:

Nghiên cứu và đề xuất phương pháp sử dụng đồ thị tri thức để cải thiện việc dịch các thực thể của mô hình Transformer;

Đánh giá và phân tích các kết quả thực nghiệm nhằm chứng minh hiệu quả của phương pháp đề xuất và đưa ra các kết luận về hướng phát triển trong tương lai.

2. Đối tượng và phương pháp nghiên cứu

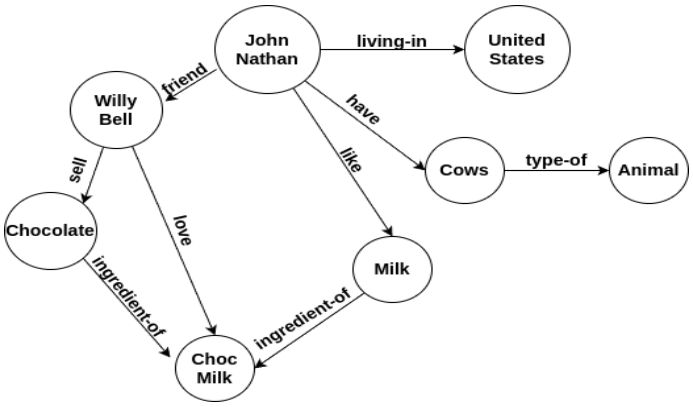
Chương này trình bày cụ thể nội dung về: 1) các đối tượng nghiên cứu gồm đồ thị tri thức (KG), mô hình dịch máy mạng nơ-ron (NMT) và một số công trình tích hợp đồ thị tri

thức vào dịch máy mạng nơ-ron; và 2) phương pháp nghiên cứu đề xuất để tích hợp đồ thị tri thức vào dịch máy mạng nơ-ron.

2.1. Đối tượng nghiên cứu

2.1.1. Đồ thị tri thức

Trong biểu diễn và suy luận tri thức, đồ thị tri thức là một cơ sở tri thức sử dụng mô hình dữ liệu hoặc cấu trúc liên kết có dạng đồ thị để biểu diễn và thao tác trên dữ liệu. Đồ thị tri thức thường được sử dụng để lưu trữ các mô tả liên kết của các thực thể – đối tượng, sự kiện, tình huống hoặc khái niệm trừu tượng – đồng thời mã hóa ngữ nghĩa hoặc mối quan hệ làm cơ sở cho các thực thể này (Ehrlinger et al., 2016).



Hình 1. *Minh họa đồ thị tri thức*

Đồ thị tri thức (KG) là một biểu đồ đa giác không đồng nhất có hướng mà các loại nút và quan hệ có ngữ nghĩa theo miền cụ thể. KG cho phép người dùng mã hóa kiến thức thành dạng mà con người có thể hiểu được, phân tích và suy luận. KG đang trở thành một cách tiếp cận phổ biến để thể hiện các loại thông tin đa dạng dưới dạng các loại thực thể khác nhau được kết nối thông qua các loại quan hệ khác nhau. Các đỉnh của đồ thị tri thức thường được gọi là các thực thể và các cạnh có hướng thường được gọi là bộ ba và được biểu diễn dưới dạng một bộ (*h, r, t*), trong đó *h* là thực thể đầu, *t* là thực thể đuôi và *r* là quan hệ liên kết phần đầu với các thực thể đuôi. Lưu ý rằng thuật ngữ quan hệ ở đây đề cập đến loại quan hệ (ví dụ: trong Hình 1, quan hệ gồm "sell", "love", "like"...).

Những phát triển gần đây trong khoa học dữ liệu và máy học, đặc biệt là trong mạng nơ-ron đồ thị và biểu diễn dữ liệu trong không gian vec-tơ, đã mở rộng phạm vi của đồ thị tri thức trong các công cụ tìm kiếm ngữ nghĩa và hệ thống khuyến nghị với những ứng dụng đáng chú ý trong các lĩnh vực như y sinh, pháp luật… (Wang et al., 2018; Yao et al., 2020).

2.1.2. Mô hình dịch máy mạng nơ-ron

Trong khuôn khổ bài báo này, chúng tôi nghiên cứu mô hình dịch máy Transformer, vốn là mô hình dịch máy tốt nhất hiện nay. Tranformer bao gồm hai thành phần chính: bộ mã hóa và bộ giải mã. Cả hai đều bao gồm các lớp tương tự được xếp chồng lên nhau.

Bộ mã hóa gồm N lớp với mỗi lớp bao gồm 02 thành phần chính:

Cơ chế tự chú ý đa góc độ (Multi-Head Self-Attention) là sự kết hợp giữa cơ chế tự chú ý (self-attention) và sự kết hợp ở các góc độ khác nhau (multi-head):

Cơ chế tự chú ý: tính trọng số giữa mỗi từ với các từ còn lại trong một câu:

$Attention(Q,K,V)= softmax\left(\frac{QK^T}{Z}\right)V$

Kết hợp các góc độ: tính toán cơ chế chú ý nhiều lần và kết hợp lại với nhau:

$MultiHead(Q, K, V)=Concat(head_1, ..., head_h)W^O$

$head_i=Attention(QW^Q_i, KW^K_i, VW^V_i)$

Mạng lan truyền thẳng theo vị trí (Position-wise Feed-Forward Network): áp dụng hai phép biến đổi tuyến tính với hàm kích hoạt ReLU:

$FFN(x)= max(0, xW_1+b_1)W_2+b_2$

Bộ giải mã tương tự như bộ mã hóa nhưng bao gồm thêm một lớp con để nhận kết quả từ cơ chế chú ý của bộ giải mã trả về bao gồm các vector biểu diễn từ đầu vào sẽ được xử lí qua nhiều lớp để tạo ra dự đoán cho từ tiếp theo trong câu dịch. Quá trình này diễn ra với các bước như sau:

Đầu tiên, tại lớp Masked Multi-Head Self-Attention các vector biểu diễn từ đầu vào $(\overrightarrow{y_1}, \overrightarrow{y_2}, ... \overrightarrow{y_m})$ với m là số lượng từ đã được dịch;

Tiếp theo, tại lớp Multi-Head Cross-Attention các vector $\vec{z}$ sẽ được dùng làm Query (Q), trong khi các vector $(\overrightarrow{x_1}, \overrightarrow{x_2}, ... \overrightarrow{x_m})$ từ bộ mã hóa sẽ được sử dụng làm Key (K) và Value (V) cho Multihead-Attention;

Sau đó, các vector sẽ đi qua lớp Feed-Forward gồm hai tầng ẩn với hàm kích hoạt ReLU tạo ra một tập hợp mới các vector. Các vector này mang thông tin ngữ nghĩa và ngữ cảnh phong phú hơn cho mỗi từ trong câu đầu ra;

Cuối cùng, tại mỗi bước dịch các vector tương ứng với vị trí từ tiếp theo sẽ được sử dụng để tính xác suất cho các từ có thể xuất hiện ở vị trí đó trong câu dịch, thông qua một lớp softmax. Từ có xác suất cao nhất sẽ được chọn làm dự đoán cho bước dịch hiện tại.

2.1.3. Các công trình liên quan

Trên thế giới, các công trình nghiên cứu đáng chú ý liên quan đến việc tích hợp đồ thị tri thức vào dịch máy mạng nơ-ron chủ yếu tập trung vào tiếng Anh. Du và cộng sự (Du et al., 2016) đã đề xuất phương pháp cải tiến chất lượng dịch máy bằng cách xử lí các từ ngoài bộ từ vựng (OOV) bằng BabelNet (từ điển ngữ nghĩa đa ngôn ngữ) (Navigli et al., 2010). Shi và cộng sự (Shi et al., 2016) trình bày phương pháp nhúng ngữ nghĩa dựa trên tri thức (KBSE) cho dịch máy, sử dụng cơ sở tri thức để tạo ra không gian ngữ nghĩa liên kết ngôn ngữ nguồn và ngôn ngữ đích. Phương pháp này bao gồm hai thành phần chính: hiểu câu nguồn (ánh xạ các câu nguồn tới một không gian ngữ nghĩa bằng cách sử dụng các bộ dữ liệu ngữ nghĩa) và phát sinh câu đích (xây dựng câu đích từ các bộ dữ liệu này). Moussallem và cộng sự (Moussallem et al., 2019) giới thiệu phương pháp KG-NMT giúp cải tiến mô hình NMT bằng cách kết hợp nhúng đồ thị tri thức (KG Embedding) để cải thiện độ chính xác của bản dịch, đặc biệt đối với các thực thể và biểu thức thuật ngữ. Lu và cộng sự (Lu et

al., 2018) sử dụng các mối quan hệ thực thể trong biểu đồ tri thức làm các ràng buộc để tăng cường kết nối giữa các từ nguồn và từ đích. Cụ thể, tác giả đề xuất 02 loại ràng buộc gồm: ràng buộc đơn ngữ sử dụng các quan hệ thực thể trong KG để tăng cường biểu diễn ngữ nghĩa của các từ trong câu nguồn và ràng buộc song ngữ nhằm tìm ra các mối quan hệ thực thể giữa các từ trong câu nguồn được liên kết trong bản dịch trong câu đích.

Đối với tiếng Việt, chưa có bất kì công trình nghiên cứu nào liên quan đến việc sử dụng đồ thị tri thức trong dịch máy mạng nơ-ron.

2.2. Phương pháp nghiên cứu

Phương pháp đề xuất của chúng tôi bao gồm: trích xuất đồ thị tri thức, tiền huấn luyện mô hình BERT (Devlin et al., 2019) với đồ thị tri thức và tích hợp mô hình BERT vào Transformer. Do song ngữ Anh-Việt thuộc nhóm ngôn ngữ ít tài nguyên, việc tích hợp KG vào NMT thông qua BERT có những ưu điểm sau:

Cải thiện khả năng học chuyển giao: việc tích hợp Đồ thị tri thức (KG) vào BERT trước khi truyền sang hệ thống dịch máy (NMT) giúp tận dụng sức mạnh học chuyển giao của BERT. BERT được huấn luyện trên một lượng lớn dữ liệu đa dạng, do đó có khả năng học các biểu diễn ngữ nghĩa của ngôn ngữ. Khi được bổ sung kiến thức có cấu trúc từ KG, các biểu diễn này trở nên phong phú hơn, giúp hệ thống NMT học và áp dụng các thông tin hữu ích vào nhiều ngữ cảnh khác nhau, đặc biệt là trong các trường hợp có ít dữ liệu huấn luyện (ngôn ngữ hoặc lĩnh vực ít tài nguyên).

Cải thiện khả năng biểu diễn thông tin: BERT, nhờ được huấn luyện trên lượng dữ liệu khổng lồ, có khả năng nắm bắt ngữ nghĩa và các mối quan hệ giữa từ ngữ một cách tổng quát và chính xác. Khi được tích hợp thêm với KG, BERT có thể tiếp cận các tri thức chuyên ngành, giúp tăng cường khả năng hiểu thông tin và phân tích ngữ cảnh phức tạp. Nhờ đó, NMT có thể tận dụng những biểu diễn ngữ nghĩa này để tạo ra bản dịch không chỉ đúng về ngữ pháp mà còn chính xác về ngữ nghĩa và thực tế, đặc biệt trong các lĩnh vực chuyên sâu.

2.2.1. Trích xuất đồ thị tri thức

Chúng tôi sử dụng phương pháp phát sinh đồ thị tri thức Grapher được đề xuất bởi Melnyk (Melnyk et al., 2022). Phương pháp Grapher được minh họa trong Hình 2, gồm các thành phần cụ thể như sau:

Văn bản: văn bản đầu vào chứa các thực thể cần trích xuất;

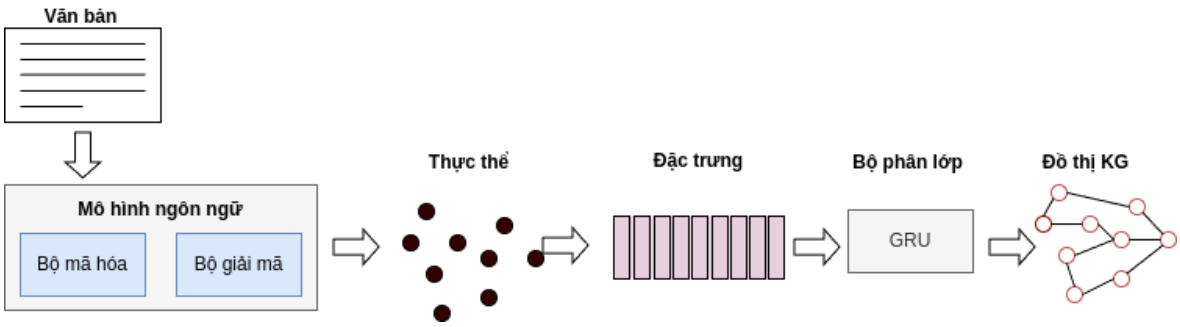
Mô hình ngôn ngữ: sử dụng mô hình ngôn ngữ Seq2Seq được huấn luyện trước để chuyển văn bản đầu vào thành một danh sách các thực thể;

Thực thể: các thực thể được trích xuất bởi mô hình ngôn ngữ;

Đặc trưng nút: các thực thể được xem là các nút (node) trong đồ thị tri thức được sinh ra bằng phương pháp DETR (Carion et al., 2020);

Bộ phân lớp: sử dụng bộ phân lớp được xây dựng từ mạng LSTM hoặc GRU để phát sinh các cạnh (như chuỗi các từ có trật tự);

Đồ thị KG: đồ thị tri thức kết quả được biểu diễn dưới dạng các bộ ba.



Hình 2. *Phương pháp xây dựng đồ thị tri thức Grapher*

Quá trình xây dựng KG trong Hình 2 được miêu tả qua các bước như sau:

a) Quá trình phát sinh thực thể: Hệ thống trước tiên sinh các nút đồ thị (thực thể) từ văn bản đầu vào bằng cách sử dụng mô hình encoder-decoder đã được huấn luyện trước như T5. Mô hình này đã được huấn luyện trên các tập dữ liệu lớn, giúp mô hình có khả năng hiểu và tạo ra các thực thể từ văn bản một cách hiệu quả.

Cụ thể, mỗi thực thể trong văn bản sẽ được mô hình trích xuất và biểu diễn dưới dạng chuỗi các token văn bản. Token là những đơn vị nhỏ nhất của văn bản, thường là từ hoặc cụm từ. Ví dụ, trong câu "Trường Đại học Quốc gia Thành phố Hồ Chí Minh", các token có thể là ["Trường", "Đại", "học", "Quốc", "gia", "Thành", "phố", "Hồ", "Chí", "Minh"]. Hệ thống sẽ nhận biết và phân biệt được các thực thể khác nhau trong văn bản, sau đó gom chúng lại thành các nút đồ thị tương ứng.

Điểm mạnh của cách tiếp cận này là sử dụng một mô hình ngôn ngữ mạnh mẽ đã được huấn luyện trên nhiều tập dữ liệu lớn. Do đó, mô hình có thể hiểu được ngữ cảnh của các từ trong văn bản và phân tích ngữ nghĩa để sinh ra các thực thể chính xác, làm nền tảng cho quá trình phát sinh cạnh sau này.

b) Quá trình phát sinh cạnh: Sau khi sinh các nút, hệ thống sinh các cạnh (quan hệ) giữa các nút. Bộ phân lớp GRU dự đoán sự tồn tại của các cạnh giữa các cặp nút dựa trên các đặc trưng của chúng.

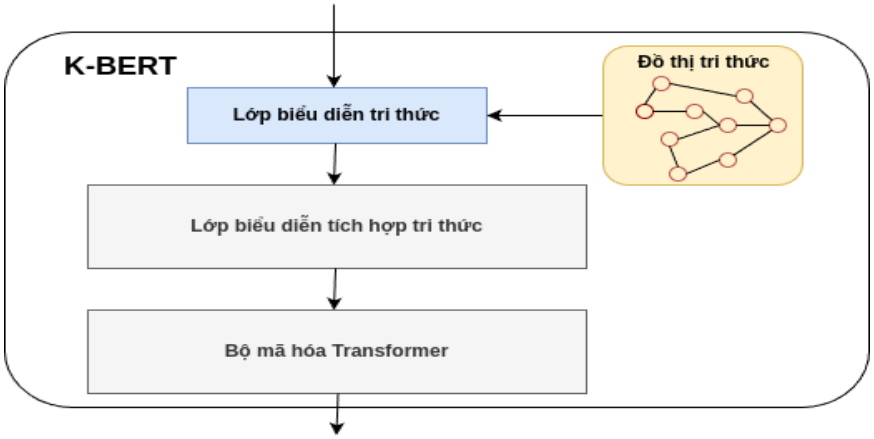
Bộ phân lớp GRU hoạt động bằng cách dự đoán sự tồn tại của các cạnh giữa từng cặp nút dựa trên các đặc trưng của các nút đó. Các đặc trưng này được mô hình trích xuất từ văn bản đầu vào và có thể bao gồm các thông tin như ngữ nghĩa của từ, vị trí trong câu, hoặc các mối liên hệ ngữ pháp. Bộ phân lớp GRU sẽ kiểm tra từng cặp nút và xác định xem có quan hệ nào giữa chúng hay không.

Trong các trường hợp mà các quan hệ thuộc một tập hợp loại cố định, chẳng hạn như “là cha của”, “làm việc tại”, hoặc “nằm ở”, phương pháp này tỏ ra rất hiệu quả. GRU sẽ tìm kiếm những mối quan hệ này trong văn bản và gán chúng cho các cặp nút tương ứng, từ đó xây dựng hoàn chỉnh đồ thị tri thức.

2.2.2. Tiền huấn luyện mô hình BERT với đồ thị tri thức

Các mô hình ngôn ngữ được tiền huấn luyện như BERT mô hình hóa các biểu diễn ngôn ngữ tổng quát từ các tập ngữ liệu lớn, nhưng lại thiếu tri thức về các thực thể. Khi xử

lí các văn bản có chứa thực thể, các chuyên gia sử dụng kiến thức liên quan để đưa ra suy luận. Để các mô hình có thể làm được điều này, chúng tôi sử dụng một mô hình biểu diễn ngôn ngữ tích hợp tri thức (K-BERT) kết hợp các đồ thị tri thức (KGs) (xem mô hình ở Hình 3) bằng cách đưa các bộ ba vào văn bản (Liu et al., 2020). Tuy nhiên, việc tích hợp quá nhiều tri thức có thể làm sai lệch ý nghĩa ban đầu của câu, một vấn đề được gọi là nhiễu kiến thức (Knowledge Noise - KN). Để khắc phục điều này, K-BERT sử dụng kĩ thuật vị trí mềm và ma trận quan sát để giảm thiểu tác động của tri thức bổ sung. K-BERT có thể dễ dàng tích hợp tri thức bên ngoài vào các mô hình bằng cách sử dụng KG và tinh chỉnh các tham số từ BERT đã được tiền huấn luyện.



Hình 3. *Minh họa K-BERT*

Nghiên cứu của tác giả Liu (Liu et al., 2020) cho thấy kết quả hứa hẹn trong các tác vụ XLNNTN, với K-BERT vượt trội hơn hẳn so với BERT trong các tác vụ cần bổ sung tri thức như các tác vụ chuyên ngành (như tài chính, luật và y học).

2.2.3. Tích hợp mô hình BERT vào Transformer

BERT là một kiến trúc mô hình ngôn ngữ tiền huấn luyện dựa trên Transformer đã tạo nên một bước tiến lớn trong lĩnh vực xử lí ngôn ngữ tự nhiên (NLP). Tích hợp BERT vào Transformer đã mang lại nhiều ưu điểm đáng kể trong việc xử lí và hiểu ngôn ngữ tự nhiên. Đặt biệt là trong tác vụ dịch máy việc tích hợp BERT vào Transformer nhằm giải quyết vấn đề khó khăn trong việc nắm bắt ngữ cảnh và ý nghĩa sâu sắc của câu nguồn. Một trong những nghiên cứu tiêu biểu (Guo et al 2021). Phương pháp này bao gồm việc tích hợp mô hình đào tạo trước BERT vào 2 khối encoder và decoder của mô hình Transformer (xem mô hình ở Hình 4).

Bộ mã hóa tương tự như trong kiến trúc Transformer, tuy nhiên có một số sự khác biệt nhỏ sau:

Đầu vào của bộ mã hóa là các vector biểu diễn từ (word representations) được trích xuất từ mô hình ngôn ngữ BERT đã được tiền huấn luyện;

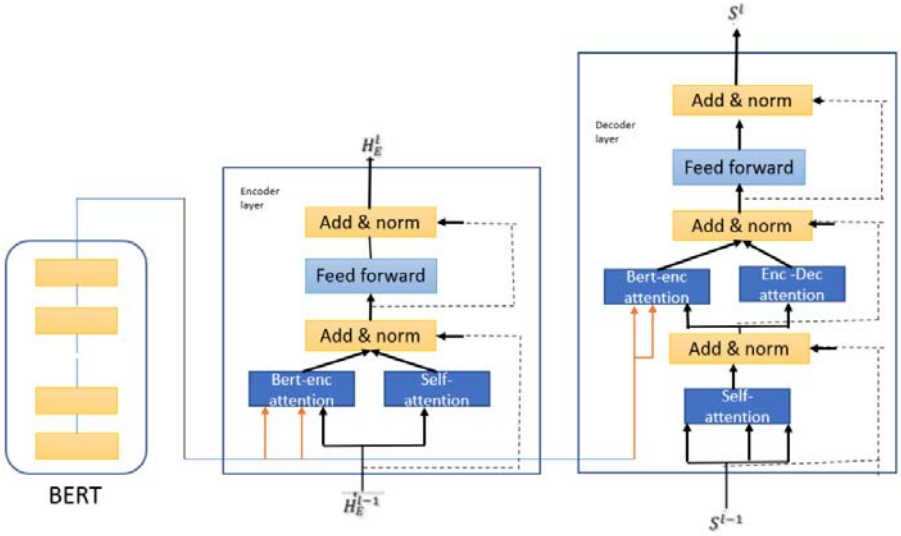
Các vector biểu diễn từ của BERT được kết hợp với biểu diễn từ của bộ mã hóa thông qua một lớp tích hợp "BERT-Enc Attention". Lớp này hoạt động tương tự như lớp Multi-

Head Attention trong Transformer, nhưng thay vì tính toán attention giữa các từ, nó tính toán attention giữa biểu diễn từ của BERT và biểu diễn từ của bộ mã hóa;

Bộ giải mã tương tự như bộ mã hóa, đầu vào của bộ giải mã là biểu diễn từ của câu đích được trích xuất thông qua mô hình BERT. Tại bộ giải mã:

Vector trạng thái ẩn S^l được tính toán dựa trên biểu diễn của lớp trước đó cùng đầu ra của bộ mã hóa H_E^lvà biểu diễn ngữ cảnh từ BERT thông qua cơ chế chú ý.

Các lớp trong bộ giải mã được ánh xạ thông qua phép biến đổi tuyến tính và softmax để dự đoán từ tiếp theo trong chuỗi đầu ra. Quá trình giải mã tiếp tục cho đến khi gặp token cuối câu.



Hình 4. *Mô hình BERT kết hợp Transformer*

3. Kết quả và thảo luận

3.1. Kết quả

3.1.1. Thống kê các tập ngữ liệu

Bộ ngữ liệu phục vụ cho việc đánh giá mô hình là IWSLT15 Anh - Việt, gồm khoảng 130.000 câu song ngữ được lấy từ các bài thuyết trình ở TED và TEDx cho tập huấn luyện. Tập tst2012 (gồm 1553 cặp song ngữ) được sử dụng để điều chỉnh tham số. Sau đó, mô hình được đánh giá dựa trên hai tập kiểm thử chính thức là tst2013 và tst2015, lần lượt gồm 1268 và 1080 cặp song ngữ Anh-Việt. Thống kê chi tiết tập ngữ liệu được mô tả trong Bảng 1.

Bảng 1. *Thống kê tập ngữ liệu Anh-Việt*

Tập ngữ liệu	Số câu	Số từ	
		Tiếng Anh	Tiếng Việt
Huấn luyện	133K	2,44M	2,87M
Tinh chỉnh	1553	28K	34K
Kiểm thử 1	1268	26,7K	33,6K
Kiểm thử 2	1080	21,0K	26,2K

3.1.2. Các thiết lập trong quá trình huấn luyện

Các cấu hình của các mô hình cơ sở như sau:

Transformer: Sử dụng N = 6 cho cả bộ mã hóa và bộ giải mã, với 512 chiều cho vector biểu diễn từ, 2048 chiều cho mạng biến đổi tuyến tính, và H = 8 cho lớp tự chú ý.

BERT: Sử dụng mô hình BERT 12 lớp.

Các mô hình đề xuất có cấu hình tương tự như mô hình cơ sở, ngoài ra, có thêm cấu hình cho bộ mã hóa đồ thị với 128 chiều cho vector biểu diễn thông tin về cạnh và node, 2 lớp mã hóa và kết hợp thông tin từ các node lân cận với phép kết trung bình cho LSTM và max pooling đối với các mô hình còn lại.

Trong quá trình huấn luyện, các tham số được thiết lập như sau:

Dropout = 0.5

Optimizer: Adam với learning rate = 0.0005

Label smoothing = 0.1

Dropout = 0.3

Max-token = 4000

Min-lr = $1x10^{-9}$

Adam-betas = (0.9, 0.98)

Toàn bộ quá trình thử nghiệm được thực hiện trên Google Colab, nơi cung cấp GPU miễn phí cho mỗi phiên làm việc kéo dài tối đa 12 giờ. Cấu hình Google Colab sử dụng bao gồm RAM 12.72GB và GPU NVIDIA Tesla P100 16GB. Thời gian huấn luyện cho cả mô hình cơ sở và mô hình đề xuất là khoảng dưới 120 phút, với số epoch là 10.

3.1.3. Kết quả thực nghiệm

Kết quả dịch Anh-Việt (số liệu trong Bảng 2) cho thấy kết quả khi sử dụng BERT (không sử dụng đồ thị tri thức) và BERT (có sử dụng thêm đồ thị tri thức) đều có kết quả BLEU tốt hơn so với mô hình Transformer. Kết quả BLEU (Kishore Papineni et al., 2002) tốt hơn chính là nhờ BERT và KBERT có biểu diễn ngôn ngữ tốt hơn.

Bảng 2. *Kết quả dịch Anh-Việt*

Mô hình	**Kết quả (BLEU)**
Transformer	21,42
Transformer + BERT	22,21 (60,5/33,9/20,0/12,0)
Transformer + KBERT	**22,71** (60,1/33,7/29,9/12,0)

Sự khác nhau giữa BERT và KBERT có thể được nhận thấy qua kết quả độ đo BLEU. Khi sử dụng đồ thị tri thức, tổng điểm BLEU sẽ tăng khi bốn kết quả BLEU ứng với từng n-grams cũng tăng theo. Mỗi BLEU được tính dựa trên số lượng từ dịch.

Dưới đây là một số phân tích chi tiết cho các điểm BLEU:

Đối với BLEU1, BLEU2, và BLEU3, kết quả sẽ thay đổi do đồ thị tri thức chứa các cụm từ gồm 3 từ và biểu diễn ngữ nghĩa của chúng. Do đó, khi dịch ba từ, độ chính xác sẽ cao hơn.

Mặc dù BLEU1 và BLEU2 có thể giảm so với khi sử dụng kết hợp với BERT (không sử dụng đồ thị tri thức), vì BERT đã biểu diễn ngữ nghĩa tốt cho các cụm từ gồm một hoặc hai từ, kết quả tổng thể của BLEU vẫn không thay đổi lớn.

Đối với BLEU4, kết quả không thay đổi vì đồ thị tri thức có rất ít cụm từ bốn từ, do đó không bổ sung thêm ngữ nghĩa so với BERT.

Tóm lại, độ đo BLEU khi sử dụng đồ thị tri thức sẽ tăng lên, đặc biệt là ở BLEU3, nhờ vào khả năng biểu diễn ngữ nghĩa của các cụm từ ba từ trong đồ thị tri thức. Trong khi đó, BLEU1 và BLEU2 có thể giảm nhẹ và BLEU4 giữ nguyên do sự thiếu hụt các cụm từ bốn từ trong đồ thị tri thức.

3.2. Thảo luận

Sự áp dụng đồ thị tri thức vào dịch máy NMT giúp cải thiện kết quả độ đo BLEU, như đã được minh chứng ở phần kết quả ở trên. Việc khai thác yếu tố tri thức ngôn ngữ vào bài toán dịch máy, cùng với việc chú trọng đến sự quan trọng của các thực thể trong câu, ảnh hưởng tích cực đến chất lượng dịch. Đặc biệt, mối quan hệ giữa các thực thể, đối tượng, hành vi, sự kiện và tình huống đã được tích hợp vào mô hình dịch máy. Điều này được minh chứng qua Bảng 3, số lượng thực thể được dịch đúng khi có thông tin từ đồ thị tri thức tăng từ 161 lên 172.

Bảng 3. *So sánh khả năng dịch thực thể của các mô hình*

Mô hình	#dịch đúng	#dịch sai
Transformer	161	35
Transformer + KBERT	172	24

Ngoài ra, việc áp dụng đồ thị tri thức vào dịch máy giúp các câu trở nên giàu kiến thức hơn, từ đó tránh được các tình trạng dịch sai và giảm thiểu sự xuất hiện của từ <unk> (unknown) trong quá trình dịch. Điều này làm cho mô hình dịch máy trở nên hiệu quả và chính xác hơn. Số liệu trong Bảng 4 cũng thể hiện được số từ OOV (unk) trong kết quả dịch giữa mô hình Transformer gốc và khi có tích hợp thông tin từ đồ thị tri thức.

Bảng 4. *So sánh hiệu quả mô hình đối với từ OOV*

Mô hình	#OOV trong kết quả dịch
Transformer	286
Transformer + KBERT	260

4. Kết luận

Bài nghiên cứu đã cải thiện chất lượng dịch của các mô hình dịch máy mạng nơ-ron bằng cách tích hợp thông tin từ đồ thị tri thức vào mô hình Transformer. Kết quả nghiên cứu đã chứng minh rằng phương pháp này giúp nâng cao đáng kể khả năng dịch các thực thể, vốn là một thách thức lớn đối với các mô hình NMT truyền thống. Nhờ tích hợp đồ thị tri thức, mô hình NMT có thể tạo ra các bản dịch chính xác hơn, đặc biệt khi xử lí các thực thể và các yếu tố ngôn ngữ phức tạp. Ngoài ra, việc sử dụng đồ thị tri thức không chỉ cải thiện hiệu quả của các mô hình NMT mà còn mở ra nhiều hướng nghiên cứu mới, như tích hợp

đồ thị tri thức vào mô hình NMT đa ngữ và tối ưu hóa các phương pháp biểu diễn đồ thị tri thức. Một số hướng phát triển tương lai cho mô hình tích hợp đồ thị tri thức này bao gồm: tích hợp đồ thị tri thức vào mô hình đa ngữ, tăng cường biểu diễn thông tin của các phương pháp vec-tơ hóa đồ thị tri thức.

❖ ***Tuyên bố về quyền lợi:*** *Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.*

❖ ***Lời cảm ơn:*** *Nghiên cứu này được tài trợ bởi Nguồn ngân sách khoa học và công nghệ Trường Đại học Sư phạm Thành phố Hồ Chí Minh trong đề tài mã số CS.2023.19.20.*

TÀI LIỆU THAM KHẢO

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In *European Conference on Computer Vision* (pp.213-229). https://doi.org/10.1007/978-3-030-58452-8_13

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *North American Chapter of the Association for Computational Linguistics*, *1*, 4171-4186. https://doi.org/10.18653/v1/n19-1423

Du, J., Way, A., & Zydro'n, A. (2016). Using BabelNet to improve OOV coverage in SMT. *In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp.9-15).

Ehrlinger, L., & Wöß, W. (2016). Towards a definition of knowledge graphs. *International Conference on Semantic Systems*, *48*, 1-4.

Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017, July). Convolutional sequence to sequence learning. *In International conference on machine learning* (pp.1243-1252).

Guo, J., Zhang, Z., Xu, L., Chen, B., & Chen, E. (2021). Adaptive adapters: An efficient way to incorporate BERT into neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 1740-1751. https://doi.org/10.1109/TASLP.2021.3076863

Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., & Wang, P. (2020). K-BERT: Enabling language representation with knowledge graph. *In 2021 International Conference on Computer Communication and Artificial Intelligence (CCAI)*, (pp.91-96). https://doi.org/10.1609/aaai.v34i03.5681

Lu, Y., Zhang, J., & Zong, C. (2018). Exploiting knowledge graph in neural machine translation. *In China Workshop on Machine Translation* (pp.27-38). https://doi.org/10.1007/978-981-13-3083-4_3

Melnyk, I., Dognin, P., & Das, P. (2022). Knowledge graph generation from text. *Findings of the Association for Computational Linguistics: EMNLP 2022* (pp.1610-1622).

Moussallem, D., Ngonga Ngomo, A.-C., Buitelaar, P., & Arčan, M. (2019). Utilizing Knowledge Graphs for Neural Machine Translation Augmentation. *In Proceedings of the 10th international conference on knowledge capture* (pp.139-146). https://doi.org/10.1145/3360901.3364423

Navigli, R., & Ponzetto, S. P. (2010). BabelNet: Building a very large multilingual semantic network. *Annual Meeting of the Association for Computational Linguistics* (pp.216-225).

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. *Annual Meeting of the Association for Computational Linguistics* (pp.311-318). https://doi.org/10.3115/1073083.1073135

Shi, C., Liu, S., Shi, C., Ren, S., Feng, S., Li, M., Zhou, M., & Sun, X. (2016). Knowledge-Based Semantic Embedding for Machine Translation. *Annual Meeting of the Association for Computational Linguistics, 1*, 2245-2254. https://doi.org/10.18653/v1/p16-1212

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Neural Information Processing Systems, 27*, 3104-3112.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30*. https://doi.org/10.48550/arXiv.1706.03762

Yao, S., Wang, R., Sun, S., Bu, D., & Liu, J. (2020). Joint embedding learning of educational knowledge graphs. *In Artificial Intelligence Supported Educational Technologies. Advances in Analytics for Learning and Teaching* (pp.1-12). Springer. https://doi.org/10.1007/978-3-030-41099-5_12

Wang, Z., Chen, T., Ren, J., Yu, W., Cheng, H., & Li, L. (2018). Deep reasoning with knowledge graph for social relationship understanding. *International Joint Conference on Artificial Intelligence,* 1021-1028. https://doi.org/10.24963/ijcai.2018/142

ENHANCING NEURAL MACHINE TRANSLATION WITH KNOWLEDGE GRAPH INTEGRATION

Le Cong Tri[1], Nguyen Phuong Nam[1], Nguyen Hong Buu Long[2], Tran Thanh Nha[1*]

[1]Ho Chi Minh City University of Education, Vietnam

[2]University of Science, Vietnam National University Ho Chi Minh City, Vietnam

[*]Corresponding author: Tran Thanh Nha – Email: nhatt@hcmue.edu.vn

Received: June 06, 2024; Revised: October 21, 2024; Accepted: October 23, 2024

ABSTRACT

Machine translation is a critical issue in the field of natural language processing (NLP), aiming to produce translations from a source language to a target language with equivalent meaning. However, current neural machine translation (NMT) models struggle with handling entities, particularly in low-resource languages like Vietnamese. This paper proposes a method to enhance NMT models by integrating information from knowledge graphs (KGs) into the Transformer model. This method allows the model to learn the representations of entities during the training process, thereby improving automatic translation when encountering entities and similar linguistic elements. Experimental results show that the proposed method significantly improves the Transformer model's translation quality, particularly in entity translation. These findings highlight the effectiveness of integrating knowledge graphs into NMT models and suggest new directions for research.

Keywords: BERT; natural language processing; machine translation; KG Integration; Knowledge Graph Embedding; Neural Machine Translation