# Vietnamese Emotion Classification with PhoBERT

Trần Nguyễn Phúc Minh Quân

20521797

Nguyễn Phương Duy 20521242

**TÓM TẮT**

Thể hiện cảm xúc là nhu cầu cơ bản của con người và chúng ta sử dụng ngôn ngữ không chỉ để truyền đạt sự thật, mà còn là cảm xúc của chúng ta. Cảm xúc quyết định chất lượng cuộc sống và ta tổ chức cuộc sống của mình để tối đa hóa trải nghiệm những cảm xúc tích cực và giảm thiểu trải nghiệm những cảm xúc tiêu cực. Nó còn đóng một số vai trò quan trọng trong việc đánh giá giá trị thương hiệu của sản phẩm bằng cách nhận diện những cảm xúc cụ thể trong những nhận xét, feedback, comment của khách hàng về sản phẩm… Sáu cảm xúc cơ bản của con người bao gồm thích thú, buồn bã, tức giận, ngạc nhiên, sợ hãi và ghê tởm qua nét mặt. Tuy nhiên, ngoài nét mặt, có thể sử dụng nhiều nguồn thông tin khác nhau để phân tích cảm xúc.

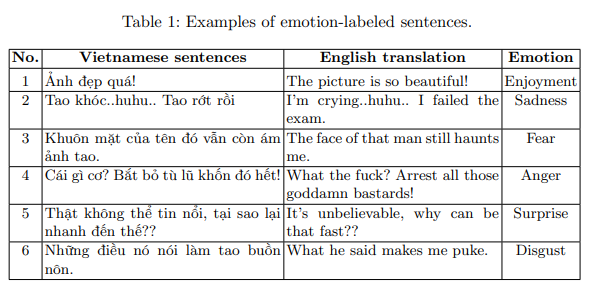
**GIỚI THIỆU**

Phân loại văn bản (Text classification) là một bài toán xử lý văn bản cổ điển phổ biến trong xử lý ngôn ngữ tự nhiên (Nature language processing). Nó giải quyết nhiều vấn đề thực tế như phân tích ngữ nghĩa, lọc spam, phân loại tin tức, email routing, sentiment analysis hay ứng dụng vào tạo chatbot.... của các tài liệu (bài báo, tạp chí, trang web, hay là cả những status, comment trên MXH).

Ở project này nhóm mình sẽ tập trung vào lớp nhận diện cảm xúc (Emotion recognition). Nhận dạng cảm xúc đã nổi lên và đóng một số vai trò quan trọng trong khu vực nghiên cứu. Trong những năm gần đây, nhận dạng cảm xúc trong văn bản đã trở nên phổ biến hơn do các ứng dụng tiềm năng to lớn của nó trong tiếp thị, an ninh, tâm lý học, tương tác giữa người và máy tính, trí tuệ nhân tạo, v.v…

Emotion recognition là 1 cách tiếp cận đặc biệt của phân tích cảm xúc (sentiment analysis). Ở cách tiếp cận này không dựa vào đánh giá tích cực hay tiêu cực dựa theo ngữ nghĩa của câu mà nó sẽ biểu diễn và phân tích 1 cách chi tiết hơn và mô tả bằng nhiều biểu cảm hơn như buồn bã, thích thú, tức giận, ghê tởm, sợ hãi hay ngạc nhiên.

Ở đây nhóm tập trung vào vấn đề nhận diện cảm xúc cho ngôn ngữ tiếng Việt. Cụ thể hơn, đầu vào của vấn đề là một bình luận tiếng Việt từ mạng xã hội, đầu ra là dự đoán cảm xúc của bình luận đó được gắn nhãn với một trong những điều sau: thích thú, buồn bã, tức giận, ngạc nhiên, sợ hãi, ghê tởm và vài cảm xúc khác.



**NGHIÊN CỨU LIÊN QUAN**

[1] đã phát triển bộ dữ liệu để nhận dạng cảm xúc với sáu lớp cảm xúc (thích thú, tức giận, ghê tởm, buồn bã, sợ hãi và ngạc nhiên) bao gồm 1.250 câu.

Năm 2012, Mohammad [2] đã xuất bản một kho cảm xúc với 21.052 bình luận từ các Tweet cũng được chú thích bởi sáu nhãn của cảm xúc trên.

Năm 2017 Mohammad [3] lại xuất bản một tập văn bản được chú thích chỉ với bốn nhãn cảm xúc (giận dữ, sợ hãi, thích thú và buồn bã) cho 7.079 bình luận từ Tweets.

Năm 2018 Wang [4] đã đưa ra một kho ngữ liệu song ngữ bằng tiếng Trung và tiếng Anh để nhận dạng cảm xúc bao gồm 6.382 câu được gắn thẻ theo năm cảm xúc khác nhau (thích thú, buồn, sợ hãi, tức giận và ngạc nhiên). Nói chung, ngữ liệu để nhận dạng cảm xúc tác vụ sử dụng một số trong sáu cảm xúc cơ bản của con người (vui, buồn, giận, ghê tởm, sợ hãi và ngạc nhiên) dựa trên lý thuyết cảm xúc của Ekman. [5]

Về thuật toán, Kratzwald [6] đã thử nghiệm hiệu quả của thuật toán học máy Random Forest và SVM, Long Short-Term Memory - Bộ nhớ ngắn hạn (LSTM) và bộ nhớ ngắn hạn dài hai chiều (BiLSTM) kết hợp với các từ nhúng được đào tạo trước trên nhiều tập hợp cảm xúc. Tương tự như vậy, Wang [7] đã đề xuất Bộ nhớ ngắn hạn dài hai chiều.

Mô hình bộ phân loại (BLSTM-MC) trên kho ngữ liệu song ngữ bằng tiếng Trung và tiếng Anh đạt điểm F1 là 46,7%, đứng thứ 3 trong nhiệm vụ chia sẻ NLPCC2018. Ở Việt Nam, có khá nhiều công trình nghiên cứu về NLP khác bao gồm các tác vụ như phân tích cú pháp [8] [9], một phần lời nói [10] [11], nhận dạng thực thể được đặt tên [15,16] [12] [13], phân tích tình cảm [14] [15] [16] và trả lời câu hỏi [17] [18].

*Ở đây nhóm sẽ tìm hiểu 3 mô hình đó là Long Short-Term Memory(LSTM), Transformer và BERT(Bidirectional Encoder Representations from Transformers) và sử dụng PhoBERT cho bài toán.*

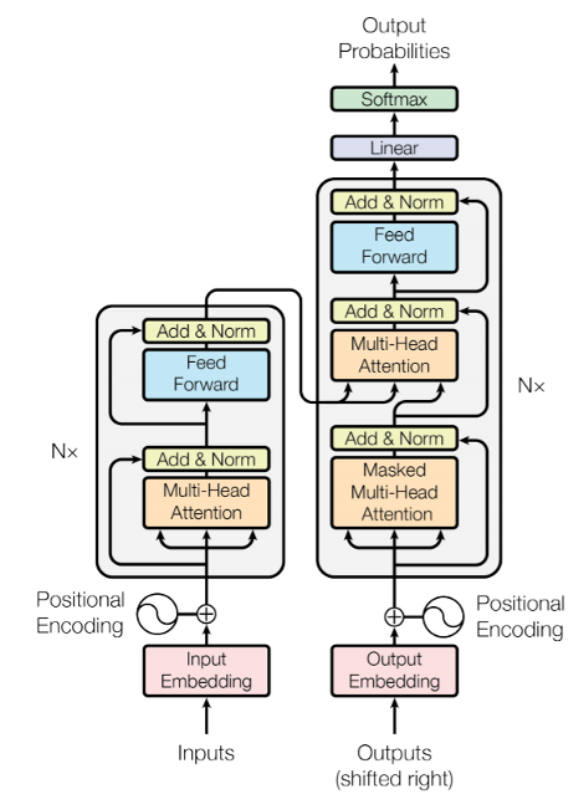
**TRANSFORMER**

Transformer là một mô hình học sâu được thiết kế để phục vụ giải quyết nhiều bài toán trong xử lý ngôn ngữ và tiếng nói. Tuy nhiên, khác với RNNs, Transformer không xử lý các phần tử trong một chuỗi một cách tuần tự. Nếu đầu vào là một câu, Transformer không cần xử lý phần đầu câu trước rồi mới tới phần cuối câu. Do đó Transformer có thể tận dụng khả năng tính toán song song của GPU và giảm thời gian xử lý đáng kể.

*Nguyên lý hoạt động*

Transformer sử dụng self-attention trong kiến trúc của mình, Transformer chứa 6 encoder và 6 decoder.Mỗi encoder chứa hai lớp: Self-attention và mạng truyền thẳng (FNN).

Self-Attention là cơ chế giúp encoder nhìn vào các từ khác trong lúc mã hóa một từ cụ thể, vì vậy, Transformers có thể hiểu được sự liên quan giữa các từ trong một câu, kể cả khi chúng có khoảng cách xa. Các decoder cũng có kiến trúc giống như vậy nhưng giữa chúng có một lớp attention để nó có thể tập trung vào các phần liên quan của đầu vào.



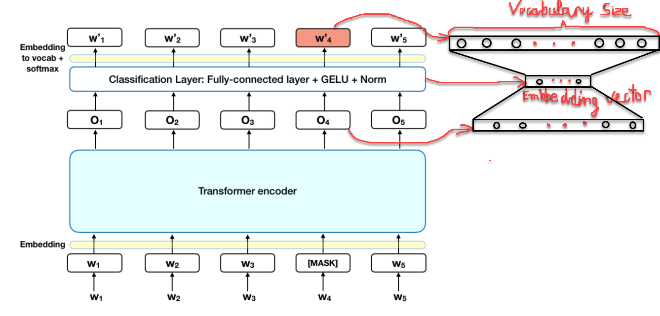
*Áp dụng*

Transformer là nền tảng cho các mô hình nổi tiếng khác như Bidirectional Encoder Representations from Transformer – BERT (2019), mô hình mà Google đã áp dụng cho cỗ máy tìm kiếm của mình hay trong AlphaStar của DeepMind – một chương trình có khả năng đánh bại những người chơi đỉnh cao của tựa game Starcraft.

Phục vụ giải quyết nhiều bài toán trong xử lý ngôn ngữ và tiếng nói, ví dụ như bài toán dịch tự động, bài toán sinh ngôn ngữ, phân loại, nhận dạng thực thể, nhận dạng tiếng nói, chuyển văn bản thành tiếng nói.

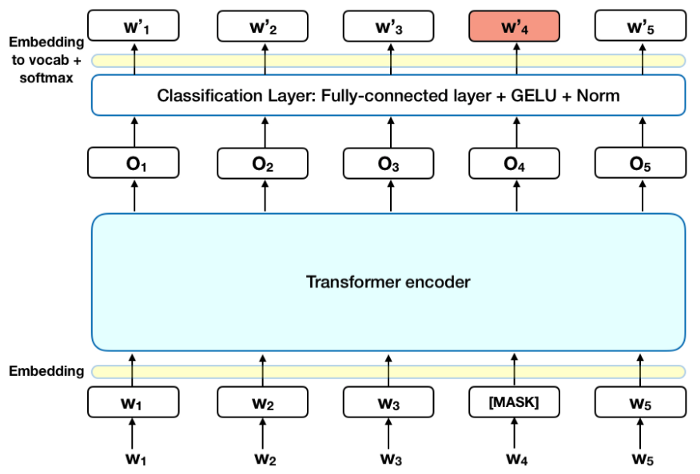
Thêm dấu cho Tiếng Việt bằng Transformer [21].

**BERT**

BERT (Bidirectional Encoder Representations from Transformers) là một mô hình học sẵn (pre-train model), học ra các vector đại diện theo ngữ cảnh 2 chiều của từ, được sử dụng để transfer sang các bài toán khác trong NLP. BERT đã thành công trong việc cải thiện những công việc gần đây trong việc tìm ra đại diện của từ trong không gian số thông qua ngữ cảnh của nó.

*Nguyên lý hoạt động*

BERT sử dụng Transformer là một mô hình attention (attention mechanism) học mối tương quan giữa các từ (hoặc 1 phần của từ) trong một văn bản. Transformer gồm có 2 phần chính: Encoder và Decoder, encoder thực hiện đọc dữ liệu đầu vào và decoder đưa ra dự đoán. Ở đây, BERT chỉ sử dụng Encoder.



Theo đó, input đầu vào là một chuỗi các token w1, w2,...được biểu diễn thành chuỗi các vector trước khi đưa vào trong mạng neural. Output của mô hình là chuỗi ccs vector có kích thước đúng bằng kích thước input. Trong khi huấn luyện mô hình, một thách thức gặp phải là các mô hình directional truyền thống gặp giới hạn khi học ngữ cảnh của từ. Để khắc phục nhược điểm của các mô hình cũ, BERT sử dụng chiến lược training như sau:

Masked ML(MLM) là một tác vụ cho phép chúng ta fine-tuning lại các biểu diễn từ trên các bộ dữ liệu unsupervised-text bất kỳ. Chúng ta có thể áp dụng Masked ML cho những ngôn ngữ khác nhau để tạo ra biểu diễn embedding cho chúng. Các bộ dữ liệu của tiếng anh có kích thước lên tới vài vài trăm tới vài nghìn GB được huấn luyện trên BERT đã tạo ra những kết quả khá ấn tượng.

**Sơ đồ kiến trúc BERT cho tác vụ Masked ML**

*Theo đó*

Khoảng 15% các token của câu input được thay thế bởi [MASK] token trước khi truyền vào model đại diện cho những từ bị che dấu (masked). Mô hình sẽ dựa trên các từ không được che (non-masked) dấu xung quanh [MASK] và đồng thời là bối cảnh của [MASK] để dự báo giá trị gốc của từ được che d0ấu. Số lượng từ được che dấu được lựa chọn là một số ít (15%) để tỷ lệ bối cảnh chiếm nhiều hơn (85%).

Bản chất của kiến trúc BERT vẫn là một mô hình seq2seq gồm 2 phase encoder giúp embedding các từ input và decoder giúp tìm ra phân phối xác suất của các từ ở output. Kiến trúc Transfomer encoder được giữ lại trong tácvụMasked ML. Sau khi thực hiện self-attention và feed forward ta sẽ thu được các vector embedding ở output.

Để tính toán phân phối xác suất cho từ output, chúng ta thêm một Fully connect layer ngay sau Transformer Encoder. Hàm softmax có tác dụng tính toán phân phối xác suất. Số lượng units của fully connected layer phải bằng với kích thước của từ điển.

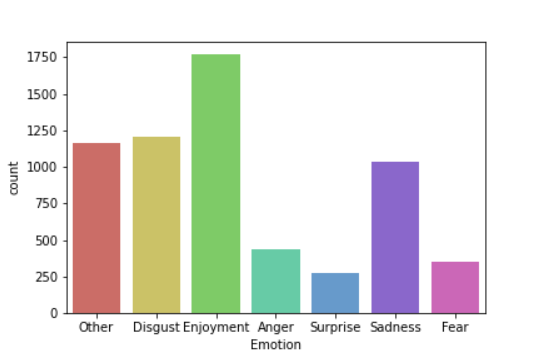
Cuối cùng ta thu được vector nhúng của mỗi một từ tại vị trí MASK sẽ là embedding vector giảm chiều của vector sau khi đi qua fully connected layer như mô tả trên hình vẽ.

Hàm loss function của BERT sẽ bỏ qua mất mát từ những từ không bị che dấu và chỉ đưa vào mất mát của những từ bị che dấu. Do đó mô hình sẽ hội tụ lâu hơn nhưng đây là đặc tính bù trừ cho sự gia tăng ý thức về bối cảnh. Việc lựa chọn ngẫu nhiên 15% số lượng các từ bị che dấu cũng tạo ra vô số các kịch bản input cho mô hình huấn luyện nên mô hình sẽ cần phải huấn luyện rất lâu mới học được toàn diện các khả năng.

*Áp dụng*

Tại thời điểm công bố, BERT đã nhanh chóng trở nên đáng chú ý trong mảng NLP bởi những cải tiến chưa từng có ở những model trước đó.

Trong bài báo “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [22]” các tác giả đã nêu ra những cải tiến của model BERT trong các tác vụ:

Tăng GLUE score (General Language Understanding Evaluation score), một chỉ số tổng quát đánh giá mức độ hiểu ngôn ngữ lên 80.5%. [22]

Tăng accuracy trên bộ dữ liệu MultiNLI [23] đánh giá tác vụ quan hệ văn bản (text entailment) lên 86.7%.

Tăng accuracy F1 score trên bộ dữ liệu SQuAD v1.1 [24] đánh giá tác vụ question and answering lên 93.2%.

Ở thời điểm hiện tại, BERT đã được ứng dụng cho Tiếng Việt. Ta có thể tham khảo dự án PhoBERT [25] của VinAI về huấn luyện trước biểu diễn từ (pre-train word embedding) sử dụng model BERT.

**PhoBERT**

PhoBERT là một mô hình tiền huấn luyện cho tiếng Việt dựa trên kiến trúc RoBERTa do 2 tác giả D. Q. Nguyen và A. T. Nguyen thuộc viện nghiên cứu VinAI Việt Nam được giới thiệu vào tháng 03/2020. PhoBERT cũng có 2 phiên bản PhoBERT\_base với 12 transformers block và PhoBERT\_large với 24 trainsformers block. PhoBERT được huấn luyện trên khoảng 20GB dữ liệu bao gồm 1GB dữ liệu Vietnamese Wikipedia corpus và 19 GB được thu thập và xử lý từ bộ dữ liệu thô 50 GB2F. PhoBERT sử dụng RDRSegmenter của VncoreNLP để tách từ cho dữ liệu đầu vào trước khi đi qua BPE encoder. PhoBERT dựa trên kiến trúc RoBERTa bỏ đi task Next Sentence Prediction mà chỉ sự dụng Masked Language Model để train.

**DATASET**

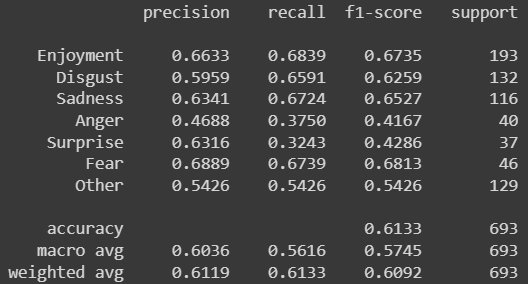
Nhóm đã chọn và sử dụng tập dữ liệu cảm xúc trên mạng xã hội tiếng Việt (UIT-VSMEC, Vietnamese Social Media Emotion Corpus) chuẩn với khoảng 6.927 câu do con người chú thích với bảy nhãn cảm xúc là enjoyment(vui vẻ), disgust(ghê tởm), sadness(buồn), anger(giận dữ), surprise(ngạc nhiên), fear(sợ hãi) và những cảm xúc khác góp phần vào nghiên cứu nhận dạng cảm xúc bằng tiếng Việt, một ngôn ngữ sử dụng ít tài nguyên trong Xử lý ngôn ngữ tự nhiên (NLP).

Link dataset: <https://sites.google.com/uit.edu.vn/uit-nlp/datasets-projects?pli=1>

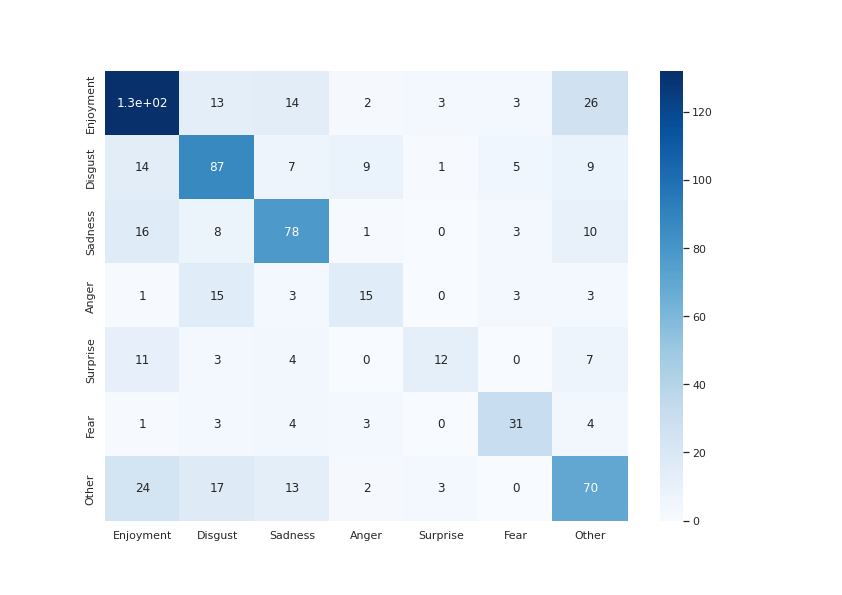
**THỰC NGHIỆM**

Trong bài báo cáo này, nhóm sẽ thực hiện huấn luyện tập dữ liệu trên với PhoBERT\_base, trình bày kết quả và so sánh kết quả với các phương pháp đi trước [29].

Kết quả trên tập Test:



Confusion Matrix:



Kết quả so sánh:

| Methods | Accuracy  (%) | Weighted  F1-score (%) |
| --- | --- | --- |
| RandomForest  +BoW | 50.64 | 40.11 |
| SVM+BoW | 58.00 | 56.87 |
| LSTM+word2Vec | 53.39 | 53.30 |
| LSTM+fastTest | 54.25 | 53.77 |
| CNN+word2Vec | 59.74 | 59.74 |
| CNN+fastText | 56.85 | 56.79 |
| **PhoBERT (Our)** | **61.33** | **60.92** |

**TÀI LIỆU THAM KHẢO**

[1] "Strapparava, C., Mihalcea, R.: SemEval-2007 Task 14: Affective Text. In: Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007). pp. 70–74. Association for Computational Linguistics, Prague (2007)".

[2] "Mohammad, S.M.: #Emotional Tweets. In: First Joint Conference on Lexical and Computational Semantics (\*SEM). pp. 246–255. Association for Computational Linguistics, Montreal, Canada (2012)".

[3] "Mohammad, S., Bravo-Marquez, F.: Emotion Intensities in Tweets. In: Proceedings of the Sixth Joint Conference on Lexical and Computational Semantics (\*SEM). pp. 65–77. Association for Computational Linguistics, Vancouver, Canada (2017)".

[4] "Ekman, P.: In: Facial expression and emotion. vol. 48, pp. 384–392. American Psychologist (1993)".

[5] "Kratzwald, B., Ilic, S., Kraus, M., S. Feuerriegel, H.P.: Decision support with textbased emotion recognition: Deep learning for affective computing. pp. 24 – 35. Decision Support Systems (2018)".

[6] "Nguyen, K.V., Nguyen, V.D., Nguyen, P., Truong, T., Nguyen, N.L.T.: UIT-VSFC: Vietnamese Students’ Feedback Corpus for Sentiment Analysis. In: 2018 10th International Conference on Knowledge and Systems Engineering (KSE). pp. 19–24. IEEE, Ho Chi Minh City".

[7] "Nguyen, P.X.V., Truong, T.T.H., Nguyen, K.V., Nguyen, N.L.T.: Deep Learning versus Traditional Classifiers on Vietnamese Students’ Feedback Corpus. In: 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). pp. 75–80. Ho Chi Minh City, V".

[8] "Nguyen, V.D., Nguyen, K.V., Nguyen, N.L.T.: Variants of Long Short-Term Memory for Sentiment Analysis on Vietnamese Students’ Feedback Corpus. In: 2018 10th International Conference on Knowledge and Systems Engineering (KSE). pp. 306–311. IEEE, Ho Chi Min".

[9] "T. Wang and X. Yang and C. Ouyang: A Multi-emotion Classification Method Based on BLSTM-MC in Code-Switching Text: 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26–30, 2018, Proceedings, Part II. pp. 190– 199. Natural Language Proces".

[10] "Wang, Z., Li, S.: Overview of NLPCC 2018 Shared Task 1: Emotion Detection in Code-Switching Text: 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26–30, 2018, Proceedings, Part II. pp. 429–433. Natural Language Processing and Chinese C".

[11] "Nguyen, D.Q., Pham, S.B., Nguyen, P.T., Le Nguyen, M., et al.: From treebankconversion to automatic dependency parsing for vietnamese. In: International Con-ference on Applications of Natural Language to Data Bases/Information Systems.pp. 196–207. Springe".

[12] ". Nguyen, K.V., Nguyen, N.L.T.: Vietnamese transition-based dependency parsingwith supertag features. In: 2016 Eighth International Conference on Knowledgeand Systems Engineering (KSE). pp. 175–180. IEEE (2016)".

[13] "Nguyen, D.Q., Pham, S.B., Nguyen, P.T., Le Nguyen, M., et al.: From treebankconversion to automatic dependency parsing for vietnamese. In: International Con-ference on Applications of Natural Language to Data Bases/Information Systems.pp. 196–207. Springe".

[14] "Bach, N.X., Linh, N.D., Phuong, T.M.: An empirical study on pos tagging forvietnamese social media text. Computer Speech & Language50, 1–15 (2018)".

[15] "Nguyen, D.Q., Vu, T., Nguyen, D.Q., Dras, M., Johnson, M.: From word segmentation to pos tagging for vietnamese. arXiv preprint arXiv:1711.04951 (2017)".

[16] "Thao, P.T.X., Tri, T.Q., Dien, D., Collier, N.: Named entity recognition in vietnamese using classifier voting. ACM Transactions on Asian Language InformationProcessing (TALIP)6(4), 1–18 (2007)".

[17] "Nguyen, L.H., Dinh, D., Tran, P.: An approach to construct a named entity annotated english-vietnamese bilingual corpus. ACM Transactions on Asian and LowResource Language Information Processing (TALLIP)16(2), 1–17 (2016)".

[18] "Nguyen, D.Q., Nguyen, D.Q., Pham, S.B.: A vietnamese question answeringsystem. 2009 International Conference on Knowledge and Systems Engineering(Oct 2009).".

[19] "Nguyen, D.Q., Nguyen, D.Q., Pham, S.B.: A vietnamese question answeringsystem. 2009 International Conference on Knowledge and Systems Engineering(Oct 2009). https://doi.org/10.1109/kse.2009.42,http://dx.doi.org/10.1109/KSE.2009.42".

[20] "Le-Hong, P., Bui, D.T.: A factoid question answering system for vietnamese. In:Companion Proceedings of the The Web Conference 2018. pp. 1049–1055 (2018)".

[21] "https://www.researchgate.net/publication/339374672\_UNG\_DUNG\_MANG\_NO-RON\_LSTM\_LONG\_SHORT-TERM\_MEMORY\_NEURAL\_NETWORKS\_DU\_BAO\_MUC\_NUOC\_TAI\_CONG\_-AU\_THUYEN\_CAU\_CAT\_TRONG\_HE\_THONG\_BAC\_HUNG\_HAI.".

[22] "http://tapchikttv.vn/data/article/3471/9.%20Proofreading.pdf".

[23] "https://duongnt.com/restore-vietnamese-diacritics-vie/".

[24] "https://arxiv.org/abs/1810.04805".

[25] "https://cims.nyu.edu/~sbowman/multinli/".

[26] "https://rajpurkar.github.io/SQuAD-explorer/".

[27] "https://github.com/VinAIResearch/PhoBERT".

[28] "<https://github.com/suicao/PhoBert-Sentiment-Classification/>".

[29] “Ho, Vong & Nguyen, Duong & Nguyen, Danh & Pham, Linh & Nguyen, Duc-Vu & Nguyen, Kiet & Nguyen, Ngan. (2019). Emotion Recognition for Vietnamese Social Media Text.”