# THÔNG TIN CHUNG CỦA NHÓM

* Link YouTube video của báo cáo:  
  <https://www.youtube.com/watch?v=MSrPnkgcSws>
* Link slides:  
  <https://github.com/HoMinhHao/CS519.N11/blob/main/CS519.N11.DeCuong.Nhom-HaoQuySon.FinalReport.AIO.pptx>

| * Họ và Tên: Đỗ Phi Sơn * MSSV: 20521839   Mở ảnh | * Lớp: CS519.N11 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 1 * Link Github: <https://github.com/dophison/CS519.N11> * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Lên ý tưởng cho đồ án   + Làm video YouTube   + Thu thập dữ liệu   + Tìm hiểu về đề tài |
| --- | --- |
| * Họ và Tên: Trần Ngọc Quý * MSSV: 20521817 | * Lớp: CS519.N11 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 1 * Link Github: <https://github.com/oFLSUo/CS519.N11> * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Làm slide thuyết trình   + Thu thập bộ dữ liệu   + Đóng góp ý tưởng cho đồ án   + Tìm hiểu về đề tài |
| * Họ và Tên: Hồ Minh Hào * MSSV: 20521290 | * Lớp: CS519.N11 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 1 * Link Github: <https://github.com/HoMinhHao/CS519.N11> * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Đóng góp ý tưởng cho đồ án   + Làm bài báo cáo   + Thu thập dữ liệu   + Tìm hiểu về đề tài |

# ĐỀ CƯƠNG NGHIÊN CỨU

| **TÊN ĐỀ TÀI**  NHẬN DIỆN CUNG BẬC CẢM XÚC CỦA BÌNH LUẬN TIẾNG VIỆT TRÊN MẠNG XÃ HỘI SỬ DỤNG PHƯƠNG PHÁP CNN |
| --- |
| **TÊN ĐỀ TÀI TIẾNG ANH**  EMOTION RECOGNITION FOR VIETNAMESE COMMENTS ON SOCIAL NETWORK USING CNN |
| **TÓM TẮT**  Với đề tài này, nhóm sẽ nghiên cứu về việc xây dựng mô hình nhận diện cảm xúc thông qua các bình luận tiêng Việt trên mạng xã hội thông qua các mô hình học máy và học sâu CNN có thể phân loại bình luận thành các loiaj cảm xúc khác nhau. Để thực hiện, đề tài sẽ tiến hành thu thập dữ liệu bình luận trên các mạng xã hội khác nhau và tiền xử lý dữ liệu để chuẩn bị cho việc huấn luyện mô hình. Sau đó, đề tài sẽ xây dựng một mô hình CNN với kiến trúc phù hợp để giúp tăng độ chính xác của mô hình. Cuối cùng, đề tài sẽ đánh giá hiệu suất của mô hình thông qua các độ đo như accuracy, precision, recall, F1-score để đánh giá độ chính xác và độ hiệu quả của mô hình. CNN (Convolutional Neural Network) là một trong những kiến trúc mạng nơ-ron tiên tiến nhất hiện nay được sử dụng rộng rãi trong các bài toán liên quan đến xử lý ảnh và âm thanh, bao gồm cả bài toán nhận diện cảm xúc. CNN được thiết kế để có thể tự động học các đặc trưng đặc biệt trong dữ liệu bằng cách sử dụng các lớp convolution và pooling. Vì vậy, việc sử dụng CNN trong bài toán nhận diện cảm xúc của bình luận trên mạng xã hội được coi là một lựa chọn phù hợp để giải quyết bài toán này. Đặc biệt, khi sử dụng CNN, chúng ta có thể tự động học các đặc trưng phức tạp trong dữ liệu như tần số từ, cụm từ đặc trưng của ngôn ngữ, từ đó giúp cải thiện độ chính xác và độ hiệu quả của mô hình phân loại cảm xúc. |
| **GIỚI THIỆU**  Phân tích cảm xúc (sentiment analysis) là chủ đề nghiên cứu “nóng” những năm gần đây trong lĩnh vực xử lý ngôn ngữ tự nghiên (NLP). Nghiên cứu này có nhiều hữu ích trong việc xây dựng các ứng dụng phục vụ cho việc giám sát và lắng nghe người dùng qua các phương tiện truyền thông qua các mạng xã hội như Twitter, Facebook, Instagram, Reddit,… hay các trang website về thương mại điện tử như Amazon, Tiki, Lazada, Shopee,… thậm chí là các kênh hỗ trợ tìm kiếm các món ăn như Foody, đánh giá nhà hàng, khách sạn như TripAdvisor và đặt nơi ở cho chuyến đi mới như Booking.com, vì nó cho ta cái nhìn tổng quan về dư luận, về một chủ đề nhất định theo hướng tích cực (positive) hay tiêu cực (negative). Nhưng ngày nay để đánh giá một sản phẩm, thái độ của một người thông qua lời nói, lời nhận xét của người đó không chỉ dừng lại ở mức tích cực hay tiêu cực. Mà cần phải chi tiết hơn bằng đánh giá cung bậc cảm xúc của người đưa ra lời nói, lời nhận xét về một vấn đề nào đó. Qua các cung bậc cảm xúc thì việc lắng nghe và giám sát người dùng qua các phương tiện truyền thông sẽ trở nên cụ thể hơn. Chính vì thế chúng tôi đã tiến hành nghiên cứu về cảm xúc của người dùng thông qua mạng xã hội Facebook, là một mạng xã hội mà ở Việt Nam đông đảo người dùng nhất với 61 triệu người dùng và đứng thứ 7 trên thế giới với:  **Input**: một bình luận từ mạng xã hội Facebook.  **Output**: Cung bậc cảm xúc bình bình luận đó theo thuyết cảm xúc của Paul Ekman bao gồm sáu cảm xúc chính là thích thú (enjoyment), buồn (sadness), ngạc nhiên (surprise), ghê tởm (disgust), sợ (fear) và tức giận (anger) |
| **MỤC TIÊU**   * Xây dựng mô hình CNN để phân loại văn bản * Xây dựng bộ ngữ liệu chuẩn cho bài toán nhận diện cung bậc cảm xúc của bình luận tiếng Việt trên mạng xã hội * Triển khai ứng dụng trên nền tảng Messenger. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  **a. NỘI DUNG:**   * Nghiên cứu về thuyết cảm xúc của giáo sư Paul Ekman và ứng dụng vào bài toán phân tích cung bậc cảm xúc cho tiếng Việt.     *Năm loại cảm xúc được đề cập trong “Atlas of Emotion” của Paul Ekman*   * Xây dựng bộ ngữ liệu chuẩn cho bài toán nhận diện cung bậc cảm xúc, cụ thể là xây dựng, thiết kế bộ hướng dẫn gán nhãn (annotation guidelines) và đánh giá bộ dữ liệu. * Nghiên cứu, cài đặt và đánh giá phương pháp học máy SVM (Support Vector Machine) và phương pháp học sâu CNN (Convolutional Neural Network) kết hợp với phương pháp biểu diễn từ trên máy tính (word embedding).   **b. PHƯƠNG PHÁP:**   * Thu thập dữ liệu: Thu thập dữ liệu từ mạng xã hội như Facebook, Twitter, hoặc các trang web có chứa các bình luận tiếng Việt. * Biểu diễn từ với Word2vec và fastText: Sử dụng phương pháp Word2vec hoặc fastText để biểu diễn các từ trong bình luận thành các vector số. Các vector số này sẽ có kích thước cố định và thể hiện đặc trưng của từ đó.     *Ví dụ về phương pháp biểu diễn từ word2vec*   * Xây dựng mô hình SVM: Sử dụng các vector biểu diễn từ để huấn luyện mô hình SVM. Ta có thể sử dụng các thư viện như Scikit-learn để thực hiện việc này. Mô hình SVM sẽ học cách phân loại các bình luận thành các lớp cảm xúc khác nhau.   Machine Learning cơ bản  *Tổng quan mô hình SVM*   * Xây dựng mô hình CNN: Sử dụng các vector biểu diễn từ để xây dựng mô hình CNN. Mô hình này sẽ học cách rút trích các đặc trưng của từ và các kết hợp của chúng để phân loại các bình luận thành các lớp cảm xúc khác nhau.     *Tổng quan mô hình CNN sử dụng trong bài toán* |
| **KẾT QUẢ MONG ĐỢI**   * Mô hình phân loại được các bình luận thành những cảm xúc khác nhau với độ chính xác cao (khoảng 65%) * Ứng dụng được vào messenger |
| **TÀI LIỆU THAM KHẢO**  [1] “statisca,” April 2019. [Online]. Available: https://www.statista.com/statistics/268136/top-15-countries-based-on-numberof-facebookusers/?fbclid=IwAR0s\_2qsyRlht3MAvKC2TB9ZB2SVqBSYC4U0gUGr18bFh uhviQaM1G0a47w.  [2] F. B.-M. Saif M.Mohammad, “WASSA-2017 Shared Task on Emotion Intensity,” in Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Copenhagen, Denmark, 2017.  [3] O. A. RomanKlinger, “IEST:WASSA-2018 Implicit Emotions Shared Task,” in Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Brussels, Belgium, 2018.  [4] A. ,. N. K. N. J. M. A. P. Chatterjee, “SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text,” in Proceedings of the 13th International Workshop on Semantic Evaluation, USA, Association for Computational Linguistics, 2019, pp. 39-48.  [5] S. K. Saif M.Mohammad, “Using hastags to capture fine emotion categories from tweets,” Computational Intelligence, vol. 0, p. 0, 2014.  [6] M. Z. e. al., “A Multi-emotion Classification Method Based,” NLPCC, pp. 190- 199, 2018.  [7] M. Z. e. al., “Overview of NLPCC 2018 Shared Task 1: Emotion Detection in Code-Switching Text,” pp. 492-433, 2018.  [8] “Paul Ekman,” [Online]. Available: <https://en.wikipedia.org/wiki/Paul_Ekman>.  [9] “Atlas Of Emotions,” [Online]. Available: http://atlasofemotions.org. 61  [10] “Charles Darwin,” [Online]. Available: <https://en.wikipedia.org/wiki/Charles_Darwin>.  [11] “The Expression of the Emotion in Man and Animals,” [Online]. Available: https://en.wikipedia.org/wiki/The\_Expression\_of\_the\_Emotions\_in\_Man\_and\_ Animals.  [12] “Eve Ekman,” [Online]. Available: <https://www.paulekman.com/blog/eveekman-cultivating-emotional-balance/>.  [13] “Dalai\_Lama,” [Online]. Available: <https://en.wikipedia.org/wiki/Dalai_Lama>.  [14] “Cohen's kappa,” [Online]. Available: https://en.wikipedia.org/wiki/Cohen%27s\_kappa |