**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**KHÓA LUẬN TỐT NGHIỆP**

**TÌM HIỂU CƠ CHẾ ATTENTION**

**VÀ ÁP DỤNG CHO BÀI TOÁN**

**PHÂN TÍCH CẢM XÚC**

**GIẢNG VIÊN HƯỚNG DẪN: ThS. Trần Khải Thiện**

**SINH VIÊN THỰC HIỆN:**

**Nhan Kim Thành – 15DH110281**

**Trần Đức Khang – 15DH110240**

**TP.HỒ CHÍ MINH - 2019**

LỜI CẢM ƠN

🙞🕮🙜

Để hoàn thành khóa luận đề tài nghiên cứu khoa học này, bên cạnh sự nỗ lực của bản thân đã vận dụng và tìm tòi học hỏi, chúng em đã nhận được sự giúp đỡ, hỗ trợ từ nhiều cơ quan, tổ chức, và cá nhân. Nghiên cứu khoa học được hoàn thành dựa trên nhiều tham khảo, kinh nghiệm từ các kết quả nghiên cứu liên quan, các bài báo chuyên ngành của nhiều tác giả ở các tổ chức nghiên cứu khác.

Với tình cảm chân thành, chúng em xin dành lời cảm ơn chân thành đối với các giáo viên trong trường Đại học Ngoại Ngữ Tin Học Thành phố Hồ Chí Minh, đặc biệt là giáo viên hướng dẫn trực tiếp của nhóm nghiên cứu – Thạc Sĩ Trần Khải Thiện. Cảm ơn thầy đã dành nhiều thời gian cũng như công sức và truyền đạt nhiều kiến thức trong việc theo dõi và hướng dẫn nhóm hoàn thành đề tài nghiên cứu.

Do vẫn còn giới hạn về kiến thức, thời gian và khả năng luận lý của nhóm nên không thể tránh khỏi nhiều thiếu sót, kính mong sự chỉ dẫn của các giáo viên cũng như đóng góp ý kiến từ những người quan tâm đề tài của nhóm, để có thể tiếp thu và cải thiện hơn.

Xin chân thành cảm ơn!

MỤC LỤC

Table of Contents

[Chương 1. GIỚI THIỆU CHUNG 1](#_Toc10707717)

[Chương 2. BÀI TOÁN SENTIMENT ANALYSIS 2](#_Toc10707718)

[Chương 3. CÁC HƯỚNG TIẾP CẬN 3](#_Toc10707719)

[Chương 4. THỰC NGHIỆM 4](#_Toc10707720)

[Chương 5. ĐÁNH GIÁ 5](#_Toc10707721)

[KẾT LUẬN 7](#_Toc10707722)

[TÀI LIỆU THAM KHẢO 8](#_Toc10707723)

**DANH SÁCH BẢNG**

(Nếu có)

**DANH SÁCH HÌNH**

(Nếu có)

# GIỚI THIỆU CHUNG

* **Giới thiệu tổng quan về trí tuệ nhân tạo**

Hiện nay, mọi thứ đang trên đà hiện đại hóa, mọi vấn đề đều được xử lý bởi công nghệ nói chung, trong đó nổi trội nhất là mảng trí tuệ nhân tạo nói riêng. Trí tuệ nhân tạo, gọi là AI (Artificial Intelligence) ra đời nhằm thay thế con người trong việc xử lý các vấn đề một cách tự động, hiệu quả cao và chính xác hơn.

Thuật ngữ trí tuệ nhân tạo được sử dụng bởi một nhóm các nhà khóa học máy tính tại hội nghị Dartmouth vào năm 1956, những thập kỷ sau đó, lĩnh vực này đã có những bước phát triển mạnh mẽ vượt bật, và được dự đoán sẽ là chìa khóa mở ra một tương lai mới đối với nền văn minh nhân loại.

Trí tuệ nhân tạo hiện đang được áp dụng trong các mọi lĩnh vực như sinh học, khoa học, giảng dạy, an ninh,… Những ông lớn trong ngành công nghệ cũng áp dụng chúng vào ứng dụng để cải thiện và quản lý. Ví dụ như khả năng nhận dạng khuôn mặt của Facebook hay trình phân loại hình ảnh của Pinterest,…

Vậy trí tuệ nhân tạo bắt đầu từ đâu? Trí tuệ nhân tạo được đào tạo từ một lượng lớn dữ liệu và thực hiện việc học hỏi, đưa ra quyết định hoặc dự đoán về vấn đề liên quan bằng các thuật toán. Việc đào tạo này gọi là học máy (Machine Learning). Học máy cung cấp những kỹ thuật, phương pháp để áp dụng vào thực hiện các khả năng của máy móc, một trong các lĩnh vực ứng dụng tốt nhất cho máy học trong nhiều năm qua là *thị giác máy tính* (computer vision), *khả năng phát hiện hình ảnh* (image detection) hay *xử lý ngôn ngữ tự nhiên* (natural language processing).

Phương pháp tiếp cận tiếp theo là khả năng khiến AI tự đào tạo bản thân, phương pháp này được gọi là Học sâu (Deep Learning). Deep learning tập trung xử lý các vấn đề liên quan đến mạng thần kinh, được dựa trên mạng thần kinh sinh học của con người, nhằm nâng cao trí tuệ nhân tạo trong các bài toán về nhận diện giọng nói, khuôn mặt hay xử lý ngôn ngữ tự nhiên. Trong vài năm qua, học sâu đã trở thành lĩnh vực nổi nhất trong khoa học máy tính, như khả năng dịch tự dộng, nhận thức sự vật, nhận diện giọng nói, phân tích cảm xúc...

# BÀI TOÁN SENTIMENT ANALYSIS

* **Giới thiệu về text classification**

Bài toán về xử lý văn bản (text classification) xuất hiện khá nhiều trong các ứng dụng về xử lý ngôn ngữ tự nhiên (NLP – Natural Language Processing) được sử dụng nhiều trong kinh doanh và được thực hiện bằng các kỹ thuật học máy, học sâu. Mục đích chính của việc phân loại văn bản được chia thành nhiều mục như:

* Tìm hiểu về cảm xúc của người dùng thông qua các trang mạng xã hội.
* Xác định và phân loại các loại thư rác (spam email).
* Phân loại các bài báo thành các chủ đề xác định.
* Ứng dụng vào tạo chatbot.
* Tóm tắt văn bản.
* Dịch đoạn văn.
* **Bài toán phân tích cảm xúc**

Về đề tài này, chúng ta sẽ tìm hiểu sâu bài toán xử lý văn bản trong vấn đề phân loại cảm xúc người dùng (sentiment analysis). Phân tích cảm xúc (sentiment analysis) còn được gọi là khai thác ý kiến, đây là một trong các lĩnh vực trong xử lý ngôn ngữ tự nhiên, được xây dựng trên các hệ thống nhằm xác định và trích xuất ý kiến người dùng dưới dạng văn bản.

Hiện nay, bài toán phân tích cảm xúc là một chủ đề rất được quan tâm và phát triển vì nó được ứng dụng nhiều trong thực tế, đặc biệt trong lĩnh vực kinh doanh. Một số lượng lớn các phản hồi, ý kiến của người dùng từ các trang web, các diễn đàn, blog và mạng xã hội, với sự giúp đỡ của hệ thống phân tích cảm xúc người dùng, những thông tin phi cấu trúc trên có thể được tự động chuyển thành dữ liệu có cấu trúc về ý kiến, phản hồi của công chúng về sản phẩm, dịch vụ, thương hiệu hoặc bất kỳ chủ đề nào.

Có rất nhiều loại trong phân tích cảm xúc, các công cụ phân tích thường tâp trung vào các loại chính, gồm (tích cực, tiêu cực, trung tính) để hệ thống có thể xác định biểu cảm và cảm xúc như (giận dữ, vui vẻ, buồn bã, thất vọng, …) hoặc xác định ý định như (hứng thú, không hứng thú). Trong đề tài này, chúng ta chỉ tập trung vào cảm xúc tích cự và tiêu cực.

Đây là một bình luận về đề tài sách theo hướng tích cực.

*“I read this book many , many years ago on a very long flight. I could n't put it down.”*

Còn đoạn bình luận này thì mang ý chê bai về đề tài phim.

*“I would n't recommend this movie. Unless you have nothing else to do for the night and want to watch it alone.”*

* **Những thách thức đối với bài toán phân tích cảm xúc:**

Trong những năm gần đây, hầu hết việc phân tích cảm xúc đều tập trung phát triển các loại cảm xúc chính xác hơn bằng cách xử lý một số thách thức và hạn chế trong lĩnh vực này.

* Tính chủ quan và giọng điệu (subjectivity and tone):

Việc xác định tính chủ quan và tính khách quan trong đoạn văn rất quan trọng cũng như việc phân tích giọng điệu của nó. Thực tế, văn bản được gọi là khách quan thường không chứa cảm xúc rõ ràng. Như ví dụ sau:

*“the package is nice.”*

*“the package is red.”*

Hầu hết mọi người sẽ nhận định đoạn đầu tiên là tích cực và đoạn thứ hai là trung tính. Tất cả các vị ngữ (tính từ, động từ, danh từ) không nên được xem như nhau trong việc chúng tạo ra cảm xúc. Với ví dụ trên, từ *‘nice’* mang tính chủ quan hơn là từ *‘red’*.

* Bối cảnh và phân cực (context and polarity):

Tất cả các ý kiến được nêu ra tại một số thời điểm, một số nơi, và bởi một số người, thì chúng ta sẽ nắm bắt được nội dung câu chuyện đó. Mọi câu nói đều đi chung với bối cảnh, nên việc phân tích cảm xúc mà không dựa vào bối cảnh sẽ gây khá nhiều khó khăn. Tuy nhiên, máy móc không thể hiểu được bối cảnh nếu nó không được đề cập một cách rõ ràng. Một trong những vấn đề nảy sinh từ bối cảnh sẽ dẫn đến thay đổi trong tính phân cực (loại cảm xúc của câu văn).

*“Everything of it.”*

*“Absolutely nothing!”*

Với ví dụ trên đóng vai trò là câu trả lời, giả sử câu hỏi là *“What did you like about the event?”*, đoạn đầu tiên sẽ trở nên tích cực và đoạn sau tiêu cực. Còn nếu ta thay đổi câu hỏi thành *“What did you DISLIKE about the event?”,* tính phân cực của hai đoạn trên sẽ đảo ngược lại.

* Mỉa mai và châm biếm (irony and sarcasm):

Sự khác biệt giữa nghĩa đen và nghĩa bóng, và một định nghĩa khác của sự xúc phạm hoặc kiểu chế giễu hơn của sự mỉa mai thường thay đổi cảm xúc từ tích cực sang tiêu cực, trong khi cảm xúc tiêu cực hoặc trung tính có thể chuyển sang hướng tích cực. Tuy nhiên, việc xác định sự mỉa mai hoặc hàm ý trong đoạn văn cần có nhiều phân tích về bối cảnh, do vậy, việc xác định một cách tự động sẽ rất khó khăn.

Ví dụ như câu hỏi *“Have you had a nice customer experience with us?”*

*“Yeah. Sure”*

*“Not one, but many!”*

Giả sử chúng ta nghe câu phản hồi đầu tiên nhiều lần liên tục, ta sẽ nghĩ rằng câu trả lời đó là tiêu cực. Vấn đề ở đây là không có văn bản gợi ý để máy có thể biết rằng đó là cảm xúc tiêu cực khi lặp lại nhiều lần, từ *‘yeah’* và *‘sure’* là những từ tích cực hoặc trung tính.

Với câu trả lời thứ hai, theo ngữ cảnh thì đây là cảm xúc tích cực. Nhưng với những ngữ cảnh khác thì câu trả lời trên thể hiện sự tiêu cực.

* So sánh (comparations):

Việc xử lý sự so sánh trong văn bản cũng là một thử thách đối với bài toán phân tích cảm xúc.

*“This product is second to none.”*

*“This is better than old tools.”*

*“This is better than nothing.”*

Trong ví dụ đầu tiên, dịch ra có nghĩa là ‘không sản phẩm nào khác có thể sánh được’, đây là một dạng câu so sánh mà ta không cần ngữ cảnh vẫn có thể phân loại một cách chính xác.

Với ví dụ thứ hai và ba thì hơi khó để phân loại. Chúng ta thường sẽ phân loại ví dụ hai vào tính tích cực và cái cuối vào trung tính. Nhưng ngữ cảnh sẽ quyết định việc phân loại.

* Biểu tượng cảm xúc (emojis):

Biểu tượng cảm xúc đúng với cái tên gọi của nó, mục đích được dùng để biểu diễn cảm xúc của người dùng được đính kèm trong đoạn văn, nó đơn giản là sự kết hợp của các ký tự unicode để tạo ra những hình ảnh mang biểu cảm. Theo Guibon, biểu tượng cảm xúc được chia ra hai loại chính, đó là *Western Emojis* (ví dụ như :D :v ) và *Eastern Emojis* (ví dụ như

( ¯ \ \_ (ツ) \_ / ¯) ).

Để phân tích cảm xúc của văn bản có đính kèm những biểu tượng này, chúng ta cần chú ý đến cấp độ ký tự (character level) cũng như cấp độ của cụm từ (word-level).

# CÁC HƯỚNG TIẾP CẬN PHÂN TÍCH CẢM XÚC

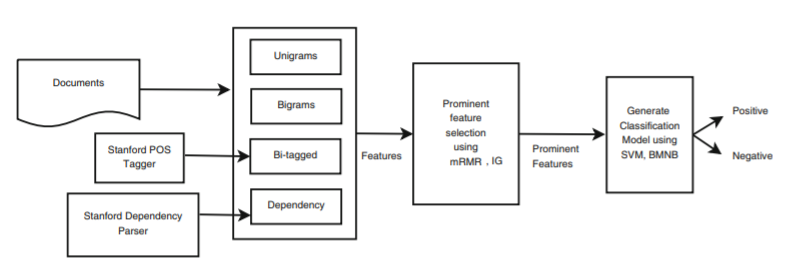
Có ba hướng tiếp cận phân tích cảm xúc trong văn bản: theo hướng học máy (machine learning approach), theo định hướng ngữ nghĩa (semantic orientation approach) và theo học sâu (deep learning approach).

* **Tiếp cận theo hướng học máy**

Các giải thuật học máy đã và đang được sử dụng rộng rãi cho bài toán phân tích cảm xúc, Bag-of-Word (BoW) là thuật toán điển hình được sử dụng phổ biến đối với bài toán này.

Phương pháp BoW chú trọng phần từ và bỏ qua tầm quan trọng của ngữ nghĩa và tính chủ quan của đoạn văn. Tất cả các từ trong đoạn văn đều được xem là quan trọng như nhau.

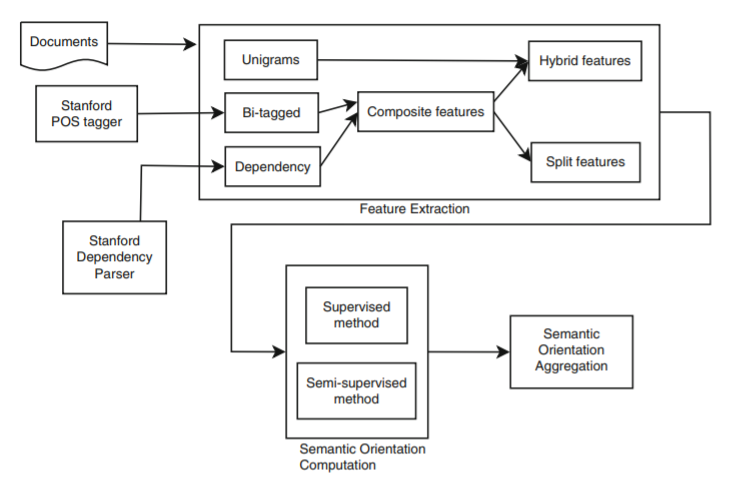
Phương pháp này trả ra độ cao của chiều trong các không gian tính năng (high dimensionality of the feature space), thuật toán của học máy sẽ giảm thiểu độ cao này với sự giúp đỡ của kỹ thuật lựa chọn tính năng, chỉ chọn những tính năng quan trọng bằng cách loại bỏ những tính năng dư thừa không liên quan. Gần đây, các mô hình hoạt động trên nền học máy đạt được nhiều thành công trong lĩnh vực phân tích cảm xúc.



* **Tiếp cận theo hướng ngữ nghĩa**

Tiếp cận theo hướng ngữ nghĩa được phân loại thành hướng tiếp cận dựa trên ngữ liệu (Corpus-based approach) và hướng tiếp cận dựa trên từ vựng (Lexicon-based approach).

Hướng tiếp cận dựa trên ngữ liệu chủ yếu phụ thuộc vào phương pháp xác định tính phân cực của cụm từ, cách tiếp cận này không hoạt động tốt do tính phân cực của từ phải dựa trên bối cảnh và không có kho ngữ liệu nào có thể cung cấp tính phân cực của từ dựa vào bối cảnh.

Đối với hướng tiếp cận về từ vựng, còn được gọi là hướng tiếp cận dựa trên kiến thức (knowledge-based approach), hướng này phụ thuộc trên những cơ sở kiến thức đã được phát triển như SentiWordNet, WordNet, … Hướng tiếp cận này là sự bao quát về các cơ sở kiến thức vì hầu hết những kiến thức có sẵn đều chứa kiến thức tổng hợp, điều này không đủ để có thể xác định tính phân cực của văn bản.

Hướng tiếp cận theo hướng ngữ nghĩa chủ yếu dựa vào các từ mang cảm xúc, sau khi xác định bằng nhiều phương thức khác nhau, cuối cùng tính phân cực của đoạn văn sẽ được xác định bởi một tập hợp các hướng ngữ nghĩa của tất cả chức năng.

* **Tiếp cận theo deep learning:**

Với sự phát triển nhanh chóng tốc độ xử lý của GPU, CPU. Điều này làm tiền đề cho phương pháp học sâu bằng mạng thần kinh phát triển mạnh mẽ.

Đối với bài toán phân tích cảm xúc, phương pháp học sâu sẽ sử dụng các mô hình mạng phổ biến để giải quyết như mạng CNN (mạng nơ-rôn tích chập), RNN (mạng nơ-rôn quy hồi), cùng biến thể là LSTM (mạng nơ-rôn phụ thuộc gần xa), vượt trội hơn nữa là sử dụng cơ chế Attention trong mô hình seq2seq.

Chúng ta sẽ đi sâu vào từng mô hình ở chương tiếp theo.

# CÁC MÔ HÌNH MẠNG NEURAL

* **Mạng nơ-rôn (neural network):**

Mạng neural là một mô hình mạng toán học được lập trình mô phỏng theo mạng thần kinh của con người.

Kết cấu của một mạng neural là các neural riêng lẻ, được gọi là các *preceptron*. Một *preceptron* nhận vào một hoặc nhiều thông tin đầu vào nhưng trả ra một kết quả duy nhất. Khi neural nhận tín hiệu đầu vào từ các *dendrite*, khi tín hiệu vượt qua một ngưỡng (threshold) thì tín hiệu sẽ được truyền đi sang một neural khách theo sợi *axon*. Công thức tính output *y* như sau:

*y = a ()*

Giải thích đơn giản thì output sẽ trả ra một kết quả dựa trên tổng của các input và trọng số của chúng. Đầu vào là các *x* cùng trọng số *w* tương ứng.

*0 if ≤ threshold*

*output =*

*1 if > threshold*

Ví dụ như vấn đề đặt ra là “đi chơi”, cùng các trường hợp sau:

*“Thời tiết có nắng không?”*

*“Chuẩn bị mọi thứ cho chuyến đi chưa?”*

*“Có bạn bè đi cùng không?”*

*“Đã giải quyết xong công việc chưa?”*

Output sẽ trả ra kết quả có hoặc không, cùng với nhiều trường hợp quyết định việc đó. Những trường hợp đó sẽ là các và trọng số ở đây mang nghĩa là độ ưu tiên,ví dụ như = 0.3, = 0.45, = 0.08, = 0.9.

Với đầu ra và đầu vào là dạng nhị phân, chúng ta rất khó để điều chỉnh một lượng nhỏ đầu vào để đầu ra thay đổi chút ít, nên chúng ta sẽ mở rộng chúng ra trong khoảng [0, 1], được tính bởi công thức sigmoid  *f* :

*f =* ( với *z* là )

Hàm *sigmoid* là một hàm kích hoạt phi tuyến tính, giống như các hàm *tanh, ReLU, …* Hàm kích hoạt phải là một hàm phi tuyến tính, vì nếu là tuyến tính thì khi kết hợp với các phép toán tuyến tính thì đầu ra cũng là một thao tác tuyến tính, lúc này phép tính sẽ trở nên vô nghĩa.

Đây là một số hàm kích hoạt phi tuyến tính như đã nêu trên:



Kiến trúc của một mạng neural, còn được gọi là các preceptron đa tầng, gồm có ba kiểu tầng: tầng dữ liệu đầu vào (input layer), tâng dữ liệu đầu ra (output layer) và tầng thể hiện tính logic của mạng (hidden layer). Một mạng neural chỉ có duy nhất một tầng đầu vào và một tầng đầu ra, nhưng lại có một hoặc nhiều các tầng ẩn.

Đầu ra của mạng neural trả ra bao nhiêu node, phụ thuộc vào bài toán đang xử lý. Đối với bài toán phân tích cảm xúc, kết quả quan tâm là tính phân cực của đoạn văn gồm tích cực (positive) và tiêu cực (negative), nên lúc này, tầng output sẽ trả ra hai node.



1. **Mạng neural tích chập (CNN):**

Các hệ thống xử lý ảnh như các trang mạng xã hội Facebook, hay trang tìm kiếm hàng đầu Google đã đưa và ứng dụng của họ những chức năng thông minh như nhận diện khuôn mặt, phát hiện thông tin người dùng nhờ hình ảnh, … những thành tựu trên được gọi là Thị giác máy tính (Computer Vision). Mạng CNN là một trong những mô hình học sâu được xây dựng để xử lý hình ảnh, bên cạnh đó, CNN còn được dùng để xử lý giọng nói, văn bản, âm thanh, …

Mạng CNN (Convolution Neural Network) được gọi là mạng tích chập do nó sử dụng một biểu thức toán tích chập, CNN là mạng neural xử lý dữ liệu theo dạng lưới, trượt trên các ma trận.



Mạng CNN là tập hợp của nhiều lớp chồng lên nhau, gồm có các lớp Convolution, lớp Pooling, lớp Fully-Connected, … sau đó kết hợp với các hàm kích hoạt phi tuyến tính (activate function) để tạo ra các thông tin trừu tượng hơn cho tầng tiếp theo. Các tầng liên kết với nhau thông qua cơ chế tích chập, như tên gọi của mạng này, tầng tiếp theo sẽ là kết quả tích chập từ tầng trước đó, mỗi tầng được áp đặt các bộ lọc (filters) khác nhau.



Như đã nêu trên, mạng CNN không chỉ có thể xử lý trong lĩnh vực hình ảnh, chúng ta cũng có thể áp dụng vào các bài toán về phân loại văn bản (text classification).

Đối với xử lý hình ảnh, đầu vào của mạng là các điểm ảnh (pixels), các bộ lọc sẽ trượt qua các điểm ảnh này. Còn trong xử lý văn bản, đầu vào sẽ là các mệnh đề, các câu văn được biểu diễn dưới dạng một ma trận, mỗi dòng của ma trận sẽ ứng với một từ, cũng có thể là một ký tự, các bộ lọc sẽ trượt qua tất cả các dòng của ma trận. Mỗi dòng của ma trận chính là một vector đại diện cho một từ, thông thường các vector này được biểu diễn ở mức thấp (low-dimensional) như word2vec hay glove.

Dưới đây là sơ đồ hoạt động của mạng CNN áp dụng vào bài toán xử lý ngôn ngữ tự nhiên:



# THỰC NGHIỆM VÀ ĐÁNH GIÁ

# KẾT LUẬN

1. Những vấn đề đã làm được

2. Những vấn đề chưa làm được

3. Hướng phát triển

# TÀI LIỆU THAM KHẢO