

CHƯƠNG 1: GIỚI THIỆU

1.1 Giới thiệu

Như chúng ta đã biết từ ngày xa xưa cho đến hiện nay trong nhiều lĩnh vực của xã hội việc thu thập ý kiến, cảm nhận, phản hồi đánh giá của con người là một việc rất phổ biến mà dựa vào đó chúng ta có thể đề đưa ra những đánh giá, nhận xét liên quan. Ở những giai đoạn trước khi công nghệ chưa phát triển hình thức này được diễn ra dưới dạng hòm thư góp ý, lấy ý kiến trực tiếp.v.v. Trong những năm gần đây với sự bùng nổ của ngành công nghệ thông tin công việc này ngày càng được chú trọng và số hóa nhiều hơn.

Một vài nguồn tài nguyên phổ biến cho việc thu thập và sử dụng ý kiến phản hồi được kể ra như dưới đây:

- Kinh nghiệm cá nhân và ý kiến về bất cứ điều gì trong đánh giá, diễn đàn v.v.
- Nhận xét về bài viết, vấn đề, chủ đề, bài đánh giá, v.v.
- Thông tin đăng tại các trang web mạng xã hội, ví dụ: Facebook.
- Đánh giá về các dịch vụ, sản phẩm.

Vậy tại sao những ý kiến này lại quan trọng đến như vậy? Những luận điểm dưới đây sẽ giải đáp những vấn đề này:

- “Ý kiến” là những yếu tố ảnh hưởng quan trọng đến hành vi của một người.
- Những ý kiến đánh giá là một phần quan trọng để đo lường và đánh giá chất lượng sản phẩm hay dịch vụ.
- Thực tế khi chúng ta cần đưa ra quyết định, chúng ta thường tìm kiếm ý kiến của người khác. Với cá nhân sẽ tìm kiếm ý kiến từ bạn bè và gia đình, còn với tổ chức sử dụng khảo sát ý kiến, tư vấn.

Những ứng dụng từ việc phân tích ý kiến cũng được áp dụng rộng rãi trong nhiều lĩnh vực. Đối với các doanh nghiệp và tổ chức, việc phân tích ý kiến hỗ trợ việc cung ứng ra thị trường các sản phẩm phù hợp nhu cầu và xu hướng. Đối với cá nhân, việc phân tích ý

kiến có thể hỗ trợ người dùng trong quá trình ra quyết định sử dụng dịch vụ, thu thập các ý kiến liên quan đến môi trường xã hội xung quanh.

Tuy nhiên hiện nay việc phân tích xử lý các dữ liệu này phần lớn còn được thực hiện một cách thủ công dưới sự đánh giá trực tiếp từ con người. Vì vậy những hệ thống phân tích ý kiến tự động và đưa ra những tổng hợp đánh giá là một nhu cầu cần thiết sẽ mang lại nhiều giá trị trong nhiều lĩnh vực. Trong lĩnh vực giáo dục việc áp dụng một hệ thống phân tích ý kiến dự đoán tự động những ý kiến đánh giá của học sinh, sinh viên về chất lượng giảng viên trong các khóa học, chương trình đào tạo sẽ giúp tiết kiệm một lượng lớn nguồn nhân lực cũng như thời gian đánh giá.

1.2 Tính cấp thiết luận văn

Sau một thời gian tìm hiểu và phân tích chúng tôi nhận thấy việc thu thập ý kiến đánh giá chất lượng giảng dạy của sinh viên trong mỗi học kỳ ở Trường Đại học Công nghệ TP.HCM hiện nay là một bài toán thực tế và có thể áp dụng được mô hình phân tích và đánh giá ý kiến. Với một lượng dữ liệu rất lớn về việc ý kiến đánh giá của sinh viên trong mỗi học kỳ thì việc tổng hợp và đánh giá thủ công thông qua con người sẽ tốn rất nhiều thời gian và chi phí.

Vì vậy luận văn này sẽ thực hiện nghiên cứu và áp dụng phân tích, tổng hợp các ý kiến đánh giá một cách tự động. Mục tiêu của nghiên cứu này là giúp rút ngắn thời gian thực hiện đánh giá, phân tích bên cạnh đó sẽ hỗ trợ đánh giá chất lượng được khách quan hơn.

1.3 Mục tiêu luận văn

Mục tiêu nghiên cứu chính của luận văn là tìm hiểu về các phương pháp phân tích ý kiến và phân lớp dữ liệu. Bên cạnh đó nghiên cứu cũng sẽ so sánh độ hiệu quả giữa các phương pháp phân lớp dữ liệu thông qua bài toán phân tích ý kiến đánh giá giảng viên.

Đối với bài toán phân tích ý kiến đánh giá của sinh viên về chất lượng giảng dạy tại Trường Đại học Công nghệ TP.HCM tôi dự kiến tạo được một hệ thống phân tích các ý

kiến thu thập được một cách tự động, xác định được cụ thể ý kiến là đánh giá tích cực hay tiêu cực.

Đưa ra các kết luận, đánh giá về kết quả đạt được đồng thời cũng nêu ra phương hướng để giải quyết các vấn đề còn tồn tại.

Ngoài ra luận văn này có thể phát triển thêm ở việc xác định khía cạnh đánh giá của ý kiến, hướng phát triển này phụ thuộc vào độ hiệu quả của việc đánh giá ý trước trước đó.

1.4 Nội dung nghiên cứu

Dựa vào các mục tiêu đã xác định luận văn sẽ tiến hành nghiên cứu các nội dung sau:

- Nghiên cứu về phân lớp chủ quan về phân lớp cảm nghĩ.
- Nghiên cứu về tóm tắt ý kiến.
- Nghiên cứu về phương pháp phân lớp dữ liệu.
- Nghiên cứu về phân loại ý kiến dựa trên học không giám sát.
- Xây dựng bộ phân lớp dữ liệu đánh giá giảng viên.
- So sánh độ hiệu quả của bộ phân lớp qua các phương pháp khác nhau.
- Kết luận đưa ra các đánh giá.

Thực nghiệm và đánh giá trên CSDL khảo sát sinh viên năm học 2016-2017.

1.5 Phương pháp nghiên cứu

Tìm hiểu các tài liệu về phân tích ý kiến, cảm xúc thông qua các từ khóa phổ biến như: opinion mining, data mining opinion, data mining and education, v.v.

Tìm hiểu các phương pháp liên quan đến khai thác văn bản, ý kiến, phân lớp dữ liệu, học giám sát, học không giám sát so sánh độ hiệu quả giữa các phương pháp thông qua các ứng dụng thực tế đã.

Tìm hiểu các kỹ thuật xử lý văn bản, phân lớp văn bản lựa chọn ra các phương pháp phù hợp để áp dụng vào bài toán của luận văn.

Cài đặt các thuật toán của các phương pháp đã nghiên cứu

Chạy thực nghiệm các dữ liệu đánh giá giảng viên trên các thuật toán đã cài đặt, ghi nhận kết quả và đánh giá nhận xét.

1.6 Bố cục luận văn

Luận văn sẽ dự kiến thực hiện 5 chương:

- Chương 1: Giới thiệu tổng quan:
 - 1.1 Giới thiệu.
 - 1.2 Tính cấp thiết luận văn.
 - 1.3 Mục tiêu nghiên cứu.
 - 1.4 Nội dung nghiên cứu.
 - 1.5 Phương pháp nghiên cứu.
- Chương 2: Cơ sở lý thuyết:
 - 2.1 Trình bày cơ sở lý thuyết về phân tích ý kiến, phân loại cảm xúc, phân lớp câu chủ quan, các mô hình biểu diễn văn bản, tóm tắt văn bản, từ vựng văn bản.
 - 2.2 Tìm hiểu các nghiên cứu đã có về phân tích ý kiến, phân loại cảm xúc, phân lớp câu chủ quan.
- Chương 3: Phương pháp thực hiện:
 - 3.1 Thu thập dữ liệu và tiền xử lý dữ liệu bằng các phương pháp hiệu quả.
 - 3.2 Gán nhãn dữ liệu lựa chọn theo quy tắc.
 - 3.3 Lựa chọn các phương pháp phân tích ý kiến, phân lớp cảm xúc để áp dụng.
- Chương 4: Thực nghiệm và đánh giá:
 - 4.1 Từ nguồn dữ liệu khảo sát sinh viên học kỳ 2 năm học 2016-2017 tại trường Đại học Công Nghệ Tp.HCM tiến hành trích xuất và tiền xử lý.
 - 4.2 Chuẩn bị môi trường thiết lập các thuật toán thực nghiệm.
 - 4.3 Trình bày về các công cụ cần cho thực nghiệm.
 - 4.4 Cài đặt các thuật toán đã tìm hiểu trên môi trường đã chuẩn bị.
 - 4.5: Chạy thực nghiệm dữ trên các phương pháp khác nhau.
 - 4.6: Trình bày kết quả thực nghiệm trên tập dữ liệu đánh giá sinh viên trên các phương pháp khác nhau.
 - 4.7 Phân tích so sánh kết quả thu được thông qua các phương pháp.
- Chương 5: Kết luận và hướng phát triển.

