|  |  |
| --- | --- |
|  |  |
| **HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**  **---------------------------------------** | |
|  | |
|  | |
| **NGUYỄN ĐỨC KHÔI** | |
|  | |
|  | |
|  | |
| **ĐỀ CƯƠNG**  **LUẬN VĂN THẠC SĨ KỸ THUẬT** | |
|  | |
|  | |
| HÀ NỘI - 2016 | |
|  |  |
| **HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**  **---------------------------------------** | |
|  | |
|  | |
| **NGUYỄN ĐỨC KHÔI** | |
|  | |
|  | |
|  | |
| **KỸ THUẬT LỌC CỘNG TÁC TRONG**  **TƯ VẤN NGƯỜI DÙNG TWITTER** | |
| CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH | |
| MÃ SỐ: 60.48.01  **ĐỀ CƯƠNG LUẬN VĂN THẠC SĨ KỸ THUẬT**  **NGƯỜI HƯỚNG DẪN KHOA HỌC**  **PGS.TS. TỪ MINH PHƯƠNG**  HÀ NỘI – 2013 | |
|  | |

1. **MỞ ĐẦU**
2. **Lý do chọn đề tài**

Twitter là một mạng xã hội phổ biến với số lượng người sử dụng lên đến trên 500 triệu người, nó cho phép người dùng chia sẻ các thông tin thông qua việc đăng các tin nhắn trong phạm vi giới hạn 140 ký tự, được gọi là các tweet hoặc các status. Lượng tweet được người dùng đăng lên hàng ngày rất lớn, lên đến 340 triệu tweets mỗi ngày, kèm với đó là một lượng thông tin khổng lồ được chia sẻ và cập nhật mới nhất. Mỗi người dùng có thể lựa chọn theo dõi một cá nhân hoặc tổ chức nào đó, mà người đó quan tâm, và ngược lại cũng có thể được theo dõi bởi các người dùng khác, Twitter sẽ hiển thị những tweet mới nhất được đăng tải bởi các cá nhân hoặc tổ chức mà người dùng đang theo dõi, theo thứ tự thời gian đăng tweet đó.

Một vấn đề đặt ra là khi số lượng tweet tăng lên một cách chóng mặt, do người dùng theo dõi quá nhiều cá nhân hoặc tổ chức khác thì vấn đề lớn mà họ gặp phải chính là sự quá tải thông tin. Rất nhiều thông tin hữu ích có thể sẽ bị mất đi do các tweet khác mới hơn được cập nhật và làm đẩy lùi các tweet trước đó, trong khi những tweet đó không phải là những thông tin thực sự cần thiết mà người dùng quan tâm. Đồng thời, một bài toán khác cũng được quan tâm là rất nhiều người dùng muốn có được những thông tin hữu ích nằm ngoài luồng thông tin mà họ nhận được bởi những người mà họ chủ động theo dõi, những thông tin đó có thể được đăng bởi những người bạn của bạn, hoặc từ những Blog được theo dõi bởi những người bạn của người dùng đó.

Chính vì vậy, việc đưa ra một hệ tư vấn nhằm cung cấp cho người dùng những tweet hữu dụng là một vấn đề vô cùng quan trọng. Một trong những kỹ thuật được sử dụng phổ biến hiện nay và mang lại hiệu quả cao là kỹ thuật lọc cộng tác, đề tài luận văn này của em sẽ tập trung vào tìm hiểu kỹ thuật tư vấn này, dựa trên tính cá nhân hóa của người dùng Twitter nhằm đưa ra một kết quả tư vấn tốt nhất.

1. **Mục đích nghiên cứu**

Bằng trực giác, ta có thể thấy một tweet hữu ích với một người dùng nào đó là khi người đó quan tâm đến hoặc sẵn sàng để đọc những tweet đó. Người dùng quan tâm đến một tweet hay không thường được xác định dựa trên rất nhiều yếu tố, như chất lượng của tweet, hoặc tác giả của tweet đó… Sở thích cá nhân cũng là một yếu tố quan trọng để quyết định một tweet là hữu ích hay không với người dùng. Các phương pháp truyền thống phân tích nội dung của các thông tin được người viết đăng tải dựa trên nội dung của bài viết đó có lẽ sẽ rất khó khăn với trường hợp của Twitter, vì sự giới hạn của 140 ký tự, nên hầu hết các bài viết trên Twitter đều ngắn gọn, phi hình thức, không có cấu trúc ngữ pháp rõ ràng và thông tin bị nhiễu rất nhiều. Những phương pháp này sẽ không mang lại hiệu quả khả quan cho mục đích của hệ tư vấn Twitter.

Hiện nay, đã có một loạt các hệ tư vấn trợ giúp người dùng Twitter ra đời, như tư vấn những URL hữu ích cho người dùng, tư vấn những người dùng khác họ nên theo dõi mà có thể phù hợp với sở thích, và cung cấp các thông tin cần thiết họ dựa trên những tweet mà người dùng đó đăng tải, hoặc tư vấn những tweet mà người dùng nên đọc và đăng tải lại.... Tuy nhiên, hầu hết các phương pháp đó đều không mang đầy đủ tính chất các mối quan hệ trong mạng Twitter, hoặc chỉ dựa trên nội dung tweet, hoặc chỉ dựa trên một số cá nhân người dùng cụ thể. Phương pháp được đề xuất trong luận văn này nhằm mục đích lấy được tất cả các ưu điểm trong mối quan hệ về sở thích của người dùng Twitter và xếp hạng cộng tác để xác định được sở thích của người dùng dựa trên dữ liệu cộng đồng của Twitter, và đồng thời tích hợp với nội dung của tweet, mối quan hệ xã hội của người dùng trong mạng và một số đặc trưng khác để cải thiện hiệu quả của việc tư vấn cá nhân hóa người dùng.

1. **Đối tượng và phạm vi nghiên cứu**

* Đối tượng: Kỹ thuật lọc cộng tác trong hệ tư vấn, mạng xã hội Twitter, sở thích và mối quan hệ người dùng trong mạng xã hôiij
* Phạm vi nghiên cứu: Xác định các tác nhân ảnh hưởng đến sở thích người dùng trong mạng xã hội Twitter,

1. **Phương pháp nghiên cứu**

* Tìm hiểu các phương pháp tư vấn hiện có nhằm phục vụ việc tư vấn cho người dùng Twitter, để xác định những hạn chế của các phương pháp này.
* Tìm hiểu kỹ thuật lọc cộng tác, những ưu điểm của phương pháp trong việc xây dựng một hệ tư vấn người dùng.
* Tìm hiểu các đặc trưng của mạng xã hội Twitter, xác định những tác nhân có thể ảnh hưởng đến sở thích người dùng, và những điểm khác biệt của Twitter so với các mạng xã hội khác cũng như các hệ thống cần tư vấn khác, kết hợp với kỹ thuật lọc cộng tác hướng xếp hạng để nâng cao hiệu quả tư vấn một cách tốt nhất.

1. **NỘI DUNG**

**Chương 1: MẠNG XÃ HỘI TWITTER VÀ CÁC ĐẶC TRƯNG**

Giới thiệu chương: Giới thiệu về mạng xã hội Twitter, các khái niệm và đặc trưng trong mạng xã hội này, bao gồm các mối quan hệ trong mạng, những tác nhân ảnh hưởng trực tiếp đến cá nhân người dùng, cách thức đăng tải thông tin thông qua Twitter. Giới thiệu những cách thức tư vấn hiện nay được áp dụng cho Twitter và những hạn chế của các phương pháp này.

Nội dung chương 1 sẽ bố cục theo các mục sau:

* 1. Giới thiệu mạng xã hội Twitter.
  2. Các đặc trưng thông tin của Twitter.
  3. Mối quan hệ người dùng trong mạng Twitter.
  4. Các hệ tư vấn hiện tại hỗ trợ người dùng Twitter.
  5. Kết luận chương

**Chương 2: LỌC CỘNG TÁC VÀ ÁP DỤNG TRONG HỆ TƯ VẤN CHO NGƯỜI DÙNG TWITTER**

Giới thiệu chương: Trình bày tổng quan về hệ tư vấn, một số kỹ thuật trong hệ tư vấn và tập trung và kỹ thuật được sử dụng trong luận văn là kỹ thuật lọc cộng tác

Dựa vào những đặc trưng của mạng xã hội Twitter, đưa ra đề xuất phương pháp áp dụng kỹ thuật lọc cộng tác và xếp hạng cộng tác nhằm tăng hiệu quả tư vấn so với các phương pháp đang tồn tại.

Nội dung chương 2 sẽ bố cục theo các mục sau:

* 1. Giới thiệu hệ tư vấn và kỹ thuật lọc cộng tác.
  2. Áp dụng kỹ thuật lọc cộng tác tư vấn người dùng Twitter
     1. Các tiêu chuẩn xếp hạng cho việc tư vấn người dùng cá nhân
     2. Các đặc trưng của Twitter phục vụ cho việc tư vấn.
     3. Mối quan hệ người dùng trong hệ tư vấn.
     4. Các đặc trưng khác trong hệ thống.
  3. Kết luận chương

**Chương 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ**

Giới thiệu chương: Sử dụng các thư viện có sẵn do Twitter cung cấp, thực hiện xây dựng bộ dữ liệu từ dữ liệu thực tế hiện có của Twitter cho một số lượng người dùng, sử dụng kỹ thuật xếp hạng cộng tác đã đề xuất ở chương 2 để đưa ra tư vấn về những tweet hữu ích cho một số người dùng nào đó. Đánh giá kết quả so với các phương pháp tư vấn khác, và so với cách làm việc hiện tại của Twitter để cung cấp các tweet cho người dùng.

Nội dụng chương 3 sẽ bố cục theo các mục sau:

* 1. Thu thập dữ liệu thử nghiệm từ Twitter
  2. Các phép đánh giá.
  3. So sánh các phương pháp tư vấn.
  4. Tính thưa thớt dữ liệu.
  5. Độ phức tạp và thời gian thực hiện phương pháp.
  6. Kết luận chương

1. **KẾT LUẬN**

Áp dụng kỹ thuật lọc cộng tác hướng xếp hạng trong việc tư vấn người dùng Twitter mang lại hiệu quả cao, vì phương pháp này đã kết hợp được nhiều yếu tố, và đặc biệt quan tâm đến mối quan hệ xã hội của người dùng Twitter từ đó xác định sở thích của từng người dùng, và tư vấn những tweet phù hợp với người dùng đó. Tuy nhiên phương pháp vẫn còn một số hạn chế cần cải thiện trong tương lai, và đề xuất các hướng đi tiếp theo nhằm nâng cao hơn nữa hiệu quả của việc tư vấn.

1. **DANH MỤC CÁC TÀI LIỆU THAM KHẢO**
2. Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., Yu, Y. *Collaborative Personalized Tweet Recommendation*. SIGIR’12.
3. D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993-1022, 2003.
4. Y. Cao, J. Xu, T. Liu, H. Li, Y. Huang, and H. Hon. Adapting ranking svm to document retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 186-193. ACM, 2006.
5. J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1185-1194. ACM, 2010.
6. P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what? item-level social inuence prediction for users and posts ranking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, pages 185-194. ACM, 2011.
7. A. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271-280. ACM, 2007.
8. Y. Duan, L. Jiang, T. Qin, M. Zhou, and H. Shum. An empirical study on learning to rank of tweets. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 295-303. Association for Computational Linguistics, 2010.
9. J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 199-206. ACM, 2010.
10. S. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: joint friendship and interest propagation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 537-546. ACM, 2011.
11. **DỰ KIẾN KẾ HOẠCH THỰC HIỆN**

|  |  |  |
| --- | --- | --- |
| **TT** | **Nội dung** | **Thời gian** |
| 1 | Tìm hiểu hệ tư vấn và kỹ thuật lọc cộng tác | 04/2013 – 05/2013 |
| 2 | Nghiên cứu các hệ tư vấn Twitter hiện tại, và xác định những mặt hạn chế của các phương pháp này | 05/2013 – 06/2013 |
| 3 | Nghiên cứu kỹ thuật lọc cộng tác hướng xếp hạng và áp dụng trong hệ tư vấn Twitter | 06/2013 – 07/2013 |
| 4 | Cài đặt thuật toán, xây dựng bộ dữ liệu | 07/2013 – 09/2013 |
| 5 | Thử nghiệm và đánh giá thuật toán | 09/2013 |

|  |  |
| --- | --- |
| **Ý KIẾN CỦA GIÁO VIÊN HƯỚNG DẪN**  *(Ký ghi rõ họ tên)* | **NGƯỜI LẬP ĐỀ CƯƠNG**  *(Ký ghi rõ họ tên)* |
| **DUYỆT CỦA TRƯỞNG TIỂU BAN CHẤM ĐỀ CƯƠNG**  *(Ký ghi rõ họ tên)* | |