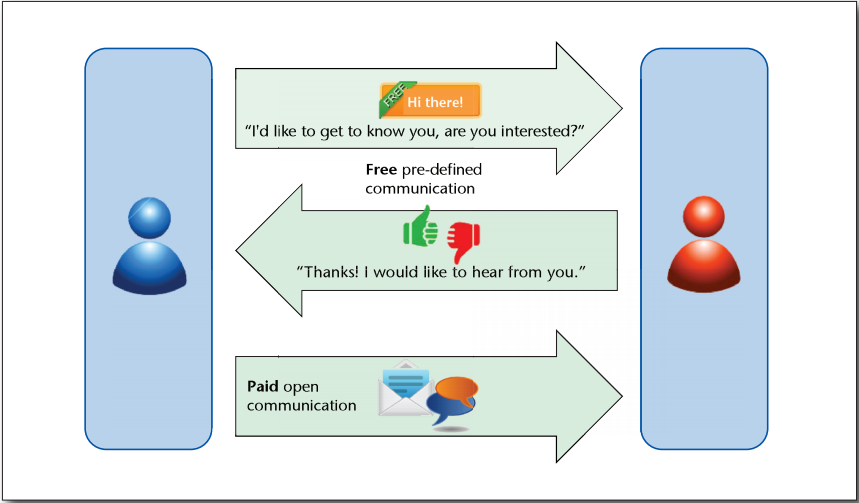
I. Ý tưởng về hệ thống đề xuất chúng tôi muốn xây dựng.

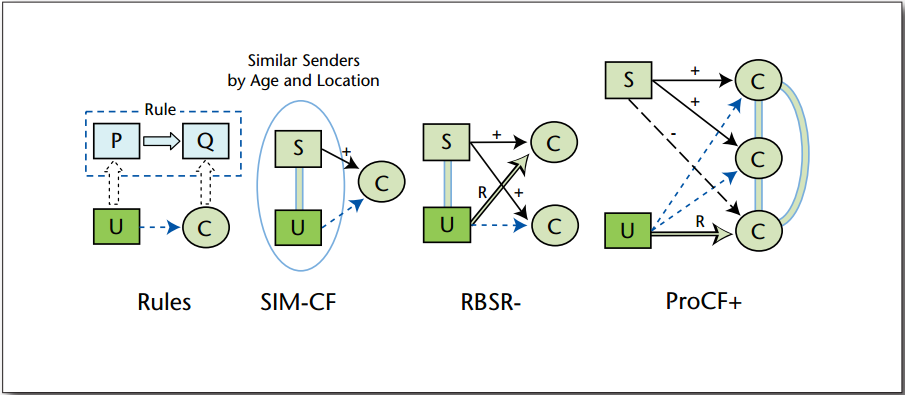
Hệ thống hẹn hò trực tuyến là một trong những ứng dụng phổ biến nhất sử dụng hệ thống đề xuất, hệ tư vấn – Recommender System, hệ tư vấn được sử dụng khi bạn phải đối mặt với một vấn đề có nhiều sự lựa chọn, và bạn chưa có nhiều thông tin về vấn đề. Giả sử bạn được gia nhập vào một hệ tương tác hai chiều: bạn sẽ liên lạc với một người không quen biết, họ có thể phản hồi những điều tích cực hay tiêu cực về bạn, hoặc không hề phản hồi cho bạn bất cứ điều gì. Vậy làm sao để tìm được một “ứng viên” người mà khả năng cao sẽ thích thú khi nói chuyện với bạn ? Hệ tư vấn hay Hệ thống hẹn hò trực tuyến sẽ giúp bạn làm điều này. Từ những dữ liệu ban đầu do chính bạn cung cấp, hệ tư vấn sẽ tìm trong vô vàn những người truy cập hệ thống một người là đối tượng phù hợp với bạn. Thật tuyệt vời phải không nào.

Tất nhiên trên thế giới mạng ảo, thì thông tin là vô cùng khó đoán, những người tham gia hệ thống có thể sử dụng có thể nhập giả thông tin để làm tăng giá trị của bản thân mình lên qua đó có thể nhận được nhiều “đối tượng giá trị” hơn. Và để giải quyết vấn đề này hệ thống được xây dựng ở đây sẽ không đưa cho bạn một đối tượng phù hợp duy nhất, mà sẽ đưa cho bạn một danh sách những đối tượng có xác suất sẽ phù hợp với bạn, qua đó sẽ tăng cơ hội nhận được một tương tác thành công giữa hai người tham gia hệ thống. Nhưng bạn cũng sẽ đối diện với một vấn đề là sẽ có tỉ lệ bạn sẽ mắc phải một tương tác thất bại – chủ yếu do thông tin bạn đưa ra sai lệch so với thực tế, dẫn tới chất lượng tương tác không tốt -> thất bại. Và vấn đề này cực kì dễ giải quyết khi trong thực tế, công ty tổ chực việc hèn hò trực tuyến sẽ đưa cho bạn một danh sách đề nghị thông tin và bạn sẽ phải chọn những thứ định sẵn trong đó, những điều dù tích cực hay tiêu cực, không hề có sự mơ hồ nào xuất hiện ở đây, do đó thông tin sẽ được chọn lọc một cách chuẩn xác hơn đem lại hiệu quả tương tác cao hơn.

Nhưng không phải lúc nào tương tác được mở ra cũng thành công, hệ thống chỉ có nhiệm vụ mở tương tác giữa hai người chứ không hề biết rằng cuộc tướng tác có thành công, hay nhận lại phản hồi từ người dùng về chất lượng tương tác. Từ vấn đề này hệ thống sẽ luôn cung cấp cho bạn một tỉ lệ thành công đối với mỗi đối tượng liên hệ trong danh sách, và hệ thống sẽ chọn ra N người phù hợp nhất cho bạn để có thể tạo tương tác, N thường nằm ở khoảng 10-100 người (tỉ lệ thành công luôn nhỏ hơn 100%).



II. Phương pháp đề xuất được sử dụng



*\*\*\*Chú thích: Tác nhân tương tác C có thể được đề xuất cho người dùng U nếu:*

*-(Rules) C đáp ứng được những yêu cầu của Q thuộc luật P => Q phát sinh cho U*

*-(SIM-CF) C phản hồi tích cực với một người dùng có thông tin cá nhân (tuổi tác, địa chỉ) tương tự U*

*-(RBSR-CF) C đã phản hồi tích cực với một người dùng S và người dùng S từng tạo một tương tác thành công với U qua Rules của U.*

*-(ProCF+) C là một ứng viên tương tự những người dùng thõa mãn Rules của U hoặc đã có phản hồi tích cực về U.*

Nhưng phương pháp mới được sinh ra nhằm đảm bảo có thể tạo ra danh sách ứng viên tiềm năng cho gần như tất cả người dùng, với những đặc điểm nổi bật:

1) Có tính khả thi để áp dụng trong thương mại

2) Có thể tạo ra các ứng viên gần so với dữ liệu thời gian thực.

3) Cho phép mở rộng số lượng người dùng .

4) Tương tác tốt với dữ liệu người dùng “động”.

**Phân nhóm thích hợp theo quy tắc (Rules)**

Phương pháp này làm việc bằng cách xây dựng lên những Rules (quy tắc) cho mỗi người dùng ứng với dữ liệu đầu vào của họ. Ta sẽ xây dựng được một danh sách u1, u2, … un (điều kiện) và c1, c2, … cn (kết luận). Trong đó ui lần lượt là thông tin của người dùng, và ci là những thông tin của những ứng viên mà chúng ta phải tìm ra cho mỗi ui .Và nếu mỗi ci đáp ứng được những Rules phát sinh ra từ ui thì ci sẽ được xếp vào danh sách những ứng viên của ui . Các ứng viên sẽ được sắp xếp theo mức độ phù hợp với Rules của ui .

Phương pháp này có thể tạo ra một danh sách ứng viên cho một lượng lớn người sử dụng. Nhược điểm của nó là tốn chi phí tính toán quy tắc cho mỗi phần nhóm, và tỉ lệ thành công còn thấp cần được cải tiến.

**Profile-Based User Similarity CF (SIM-CF)**

*Định nghĩa 1: Đối với một người sử dụng, nhóm người tương tự với bạn sẽ là những người có cùng giới tính, cùng tuổi hoặc chênh lệch không quá 5 tuổi, và có cùng địa chỉ.*

Tuổi tác và địa chỉ là hai thuộc tính được xem là cơ sở và sử dụng hai thuộc tính này là phổ biến nhất trong một hệ thống đề xuất người-người. Phân tích cho thấy SIM-CF cung cấp một tỉ lệ thành công cao hơn đồng thời cải thiện sự đa dạng của các ứng viên. Các ứng viên sẽ được xếp hạng theo số lượt tương tác thành công với các ứng viên tương tự với mỗi người dùng U.

**Rule-Based Similar Recipients CF (****RBSR-CF)**

**RBSR-CF** tính toán dựa trên tương tác thành công của người dùng khác với các ứng viên khác dựa trên quy tắc của U, sau đó dùng CF để xếp hạng các ứng viên. Một lợi thế của phương pháp này là nó tạo ra nhiều hơn sự đa dạng từ các ứng viên so với SIM-CF. Nhưng kéo theo theo đó là chi phí tính toán phức tạp vào cao hơn so với SIM-CF.

**Probabilistic CF+ (ProCF+)**

Xuất phát từ mô hình người dùng tương tự nhưng tinh vi hơn khi sử dụng thuật ngữ “tương tác thành công” và “không thành công” biểu diễn bằng mũi tên nhãn + và nhãn trừ (hình 2). Ưu điểm chính của phương pháp này là mức độ thành công khi mở tương tác cao hơn và cải thiện tỉ lệ so với SIM-CF và RBSR-CF (do tính toán chính xác hơn về nhóm người dùng tương tự) nhưng kèm theo đó là một chi phí tính toán cao hơn.