Tóm tắt:

Phần 3: So sánh hiệu năng cách gợi ý người dùng kiểu mới (Novel Recommender) và truyền thống (baseline) thông qua thực tiễn.

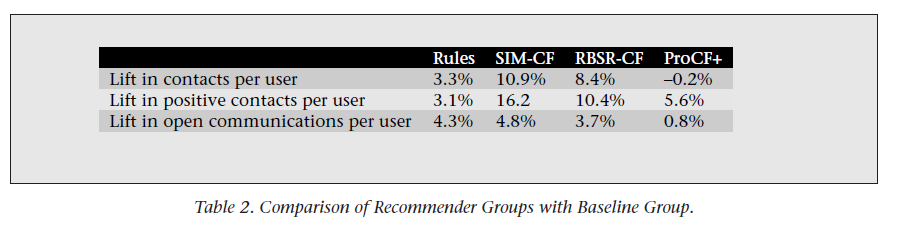
Các nhà nghiên cứu dự án hợp tác với công ty để làm một cuộc khảo sát số lượng lớn. Các thông tin thu được đều đã được thống nhất về cách tính, sàng lọc, và theo một tiêu chuẩn đã được định trước và quy định rất kĩ để dữ liệu được so sánh một cách khách quan nhất.

Mỗi cách gợi ý người dùng (ở phần 2) sẽ được phân hco 10% số người dùng của trang web và các người dùng mới sẽ được tự phân loại và nhóm (mỗi nhóm có một nhóm người dùng phân biệt, không có ai ở trong cùng 1 lúc 2 nhóm)

Cuộc khảo sát diễn ra trong 6 tuần từ tháng 5 đến giữa tháng 6 2012. Sau cuộc khảo sát người ta lấy ra được 3 bảng sau:

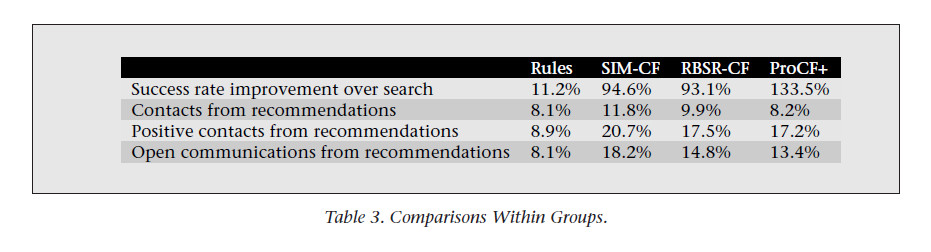
Bảng 2: So sánh nhóm người dùng của Recommender với baseline

* Phản ánh tỉ lệ người dùng phản ứng với các phương pháp gơi ý khác nhau
* Độ tăng trưởng danh sách liên lạc
* Độ tăng trưởng về liên lạc tích cực (kiểu như favorite ấy)
* Độ tăng trưởng giao tiếp (cái này quan trọng do nó quyết định doanh thu)



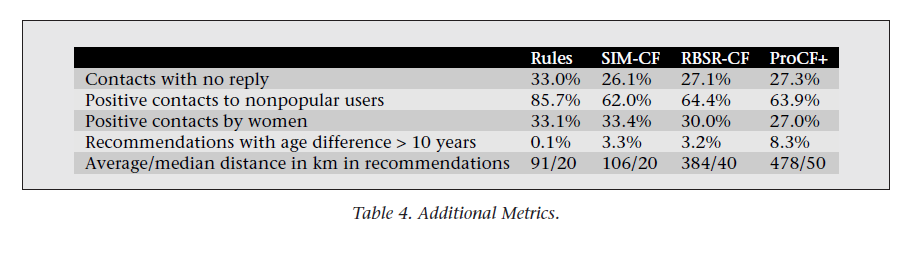
Bảng 3: So sánh hành vi của người dùng trong một nhóm

* Tỉ lệ độ tỉ lệ thành công tăng khi người dùng tìm kiếm theo thời gian
* Tỉ lệ tiến hành liên lạc từ gợi ý
* Tỉ lệ tiến hành liên lạc tích cực từ gợi ý
* Tỉ lệ giao tiếp tăng từ gợi ý



Bảng 4: bảng phụ, đưa ra sau buổi khảo sát dài 6 tuần

* Liên lặc nhưng bị lơ
* Người dùng không nổi nhưng được nhiều người theo
* Được nhiều gái theo
* Gợi ý đưa ra các gợi ý có độ tuổi chênh lệch trên 10 tuổi
* Khoảng cách trung bình (về địa lý) của 2 ứng viên khi được gợi ý



* Kết luận ngắn gọn: chọn SIM-CF (do nỗi trội và cái ăn tiền nhất thì nó thắng – bảng 2, phần open communication)

Phần 4:Quá trình triển khai

Có 2 giai đoạn xử lý

Giai đoạn 1: Tạo ứng viên bằng cách sử dụng các điểm giống nhau giữa người dùng

Giao đoạn 2: Dùng cây quyết định tính toán từ 1 bộ dữ liệu lớn

Công ty dùng Java ->lưu bảng trong bộ nhớ để dữ liệu chạy với thời gian gần thực nhất có thể -> lọc lại ứng viên bằng khoảng cách tuổi tác và địa lý.

Quá trình diễn ra trong vòng 3,5 tháng (thay đổi gì, thiết kế ra sao là công ty làm còn nhóm nghiên cứu chỉ đưa ra gợi ý):

Giữa tháng 6 – 2012: kết thúc khảo sát

Tháng 8 – 2012: chọn SIM-CF. Bắt đầu thiết kế và phát triển cách dùng SIM-CF thẳng trong bộ nhớ.

Tháng 9 – 2012: Bản thử nghiệm đầu tiên (triển khai cho 60% người dùng của trang)

Tháng 10-2012: Đưa ra cho mọi người dùng, vẫn dùng dữ liệu offline

Tháng 11-2012: Dùng cho dữ liệu trực tuyến

* Do thiết kế tốt nên chương trình chạy ổn và các cây quyết định cũng không làm tràn bộ nhớ và do thiết kế tốt nên nếu muốn đổi luật để tạo cây quyết định trong tương lại thi có thể làm dễ dàng