

8: **return**  $\hat{r}_{ui}$  // Trả kết quả dự đoán của user  $u$  cho item  $i$

9: **end function**

Trong đó:

$D^{train}$ : tập dữ liệu huấn luyện – đầu vào

$u$ : user đang xét – đầu vào

$i$ : item đang xét – đầu vào

$threshold$ : là ngưỡng sở thích nhận giá trị thuộc tập  $R = \{1, 2, 3, 4, 5\}$  – đầu vào

$N$ : số người dùng

$r_{ui}$ : giá trị xếp hạng của user  $u'$  cho item  $i$

$I_{uu'}$ : là tập các item được xếp hạng bởi cả user  $u$  và  $u' \geq threshold$

$\hat{r}_{ui}$ : giá trị cần dự đoán cho item  $i$  của user  $u$  – đầu ra.

## 5 KẾT QUẢ THỰC NGHIỆM

### 5.1 Dữ liệu

Để thực nghiệm, chúng tôi sử dụng 4 tập dữ liệu là MovieLens (© 2014 GroupLens, <http://grouplens.org/datasets/movielens/>) 100K (943 users, 1.682 items và 100.000 rating), MovieLens 1M (6.040 users, 3.952 items, 1.000.209 ratings), Flixster (<http://www.recsyswiki.com/wiki/Flixster>) 1K (2.000 users, 4.000 items, 1.089 ratings) và Flixster 10K (10.000 users, 13.000 items, 10.118 ratings).

### 5.2 Độ đo

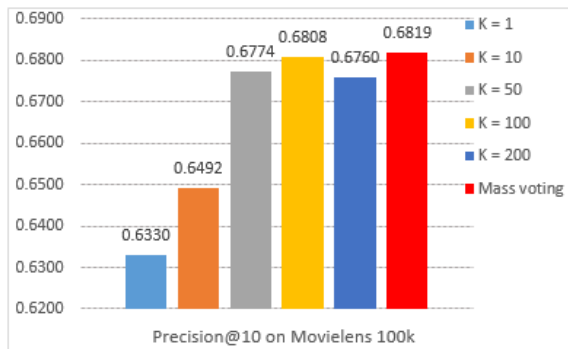
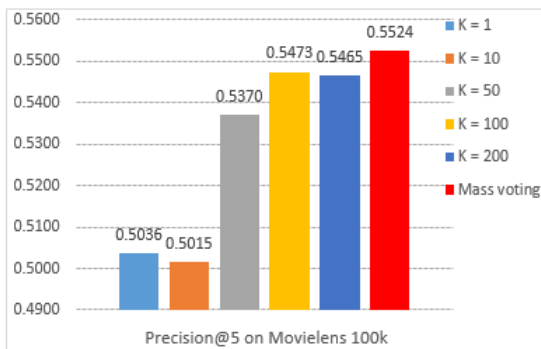
Để đánh giá hiệu quả của giải thuật được đề

xuất, chúng tôi cài đặt giải thuật luật bình chọn số đông kết hợp ngưỡng sở thích đã trình bày ở phần 4 và kỹ thuật User-KNN để so sánh đối chiếu; sử dụng nghi thức kiểm tra k-fold cross validation với  $k = 5$ , dùng độ đo  $precision@5$  và  $precision@10$  (tỉ lệ item gợi ý đúng được trả về trong 5 và 10 items).

### 5.3 Kết quả

Kết quả thực nghiệm trên các tập dữ liệu được trình bày trong các hình trang bên. Kết quả này đã cho thấy phương pháp đề xuất rút ngắn thời gian huấn luyện trong khi độ chính xác không thua kém so với kỹ thuật phổ biến là user-kNN, vì vậy rất phù hợp cho các hệ thống đòi hỏi phải cập nhật/huấn luyện lại mô hình trực tuyến (online training/updating) sau mỗi lần đánh giá của người dùng để có những gợi ý tốt nhất.

Đối với các tập dữ liệu MovieLens, luật bình chọn số đông kết hợp ngưỡng sở thích cho kết quả gợi ý tương đương với kết quả gợi ý khi dùng kỹ thuật User-KNN trong khi thời gian thực thi giải thuật có phần tốt hơn, đặc biệt nếu chọn ngưỡng sở thích càng lớn thời gian thực thi giải thuật càng giảm. Điều này thể hiện rõ khi thực nghiệm trên các tập dữ liệu Flixster, khi mà trận xếp hạng cực thưa, thời gian thực thi giải thuật luật bình chọn số đông kết hợp ngưỡng sở thích càng thể hiện ưu điểm vượt trội so với kỹ thuật User-KNN mặc dù kết quả gợi ý thấp hơn kỹ thuật User-KNN nhưng không đáng kể.



Hình 3: So sánh độ chính xác của các phương pháp trên tập dữ liệu MovieLens 100k