

một phương pháp dự báo nhanh, tương đối đơn giản nhưng độ chính xác khá cao, dễ dàng tích hợp vào mô hình gợi ý.

Từ nhận định trên, chúng tôi đề xuất tích hợp phương pháp dự báo làm trơn (san bằng) hàm mũ (Exponential Smoothing - ES) vào mô hình MF thông qua kỹ thuật phân tích nhân tố tiềm ẩn (Tensor Factorization – TF) với mục tiêu khai thác và tận dụng được các thông tin về thời gian cũng như trình tự người dùng đưa ra phản hồi. Mô hình sẽ được thực nghiệm trên các tập dữ liệu chuẩn trong lĩnh vực gợi ý và đánh giá bằng độ đo RMSE (Root Mean Squared Error) để cho thấy hướng tiếp cận đã đề xuất cho kết quả rất khả quan.

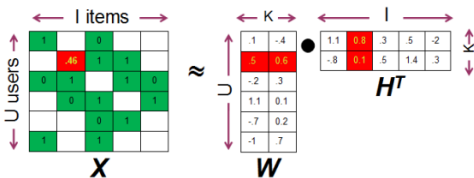
Phần còn lại của bài báo này có cấu trúc như sau: Phần 2 giới thiệu các kỹ thuật được sử dụng để xây dựng mô hình gợi ý. Phần 3 giới thiệu mô hình dự đoán do chúng tôi đề xuất. Phần 4 là kết quả thực nghiệm trên tập dữ liệu mẫu. Cuối cùng phần 5 là phần kết luận.

2 CÁC KỸ THUẬT ĐƯỢC SỬ DỤNG ĐỂ XÂY DỰNG MÔ HÌNH

Trước tiên chúng tôi tóm tắt ngắn gọn kỹ thuật phân rã ma trận (MF) (Koren *et al.*, 2009), kỹ thuật phân tích nhân tử (Tensor Factorization - TF) (Kolda and Bader, 2009) (Dunlavy *et al.*, 2011b) và phương pháp dự đoán làm trơn hàm mũ (ES) (Yorucu, 2003) (Ostertagová and Ostertag, 2012) để làm cơ sở cho việc đề xuất mô hình phân rã ma trận tích hợp yếu tố thời gian.

2.1 Kỹ thuật phân rã ma trận (Matrix Factorization - MF)

Kỹ thuật phân rã ma trận (MF) là việc chia một ma trận lớn \mathbf{X} thành 2 ma trận có kích thước nhỏ hơn rất nhiều so với ma trận ban đầu \mathbf{W} và \mathbf{H} , sao cho \mathbf{X} có thể được xây dựng lại từ hai ma trận nhỏ hơn này càng chính xác càng tốt (Koren *et al.*, 2009), nghĩa là $\mathbf{X} \approx \mathbf{WH}^T$ như minh họa trong Hình 1.



Hình 1: Minh họa kỹ thuật phân rã ma trận

Trong đó, \mathbf{X} là tập hợp tất cả các đánh giá (rating) của người dùng (user) với mục tin (item), bao gồm cả những giá trị chưa biết cần được dự đoán tạo nên một ma trận gọi là Utility Matrix. $\mathbf{W} \in \mathbb{R}^{U \times K}$ là một ma trận mà mỗi dòng u là một

véc tơ bao gồm K nhân tố tiềm ẩn (latent factors) mô tả cho user u , và $\mathbf{H} \in \mathbb{R}^{I \times K}$ là một ma trận mà mỗi dòng i là một véc tơ bao gồm K nhân tố tiềm ẩn mô tả cho item i .

Gọi \mathbf{w}_{uk} và \mathbf{h}_{ik} là các phần tử tương ứng của hai ma trận \mathbf{W} và \mathbf{H} , khi đó rating r của user u trên item i được dự đoán bởi công thức:

$$\hat{r}_{ui} = \sum_{k=1}^K \mathbf{w}_{uk} \mathbf{h}_{ik} = (\mathbf{WH}^T)_{u,i} \quad (1)$$

Như vậy, vấn đề chủ chốt của MF là làm sao tìm được ma trận \mathbf{W} và \mathbf{H} . Hai tham số này có thể được xác định bằng cách tối ưu hóa hàm mục tiêu (objective function) (3) theo RMSE (root mean squared error) như sau:

$$\text{RMSE} = \sqrt{\frac{1}{|D_{\text{test}}|} \sum_{(u,i,r \in D_{\text{test}})} (r_{ui} - \hat{r}_{ui})^2} \quad (2)$$

$$O^{MF} = \sum_{(u,i) \in D^{\text{train}}} (r_{ui} - \hat{r}_{ui})^2 + \lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) \quad (3)$$

Với λ là hệ số chính tắc hóa ($0 \leq \lambda < 1$) để tránh tình trạng quá khớp hay còn gọi là học vẹt (overfitting – xảy ra khi mô hình dự đoán cho kết quả tốt trên tập huấn luyện nhưng cho kết quả thấp trên tập thử nghiệm) (Feng *et al.*, 2009) và $\|\cdot\|_F^2$ là chuẩn Frobenius (Böttcher and Wenzel, 2008).

Một trong những kỹ thuật để tối ưu hóa hàm mục tiêu là dùng SGD (Stochastic gradient descent) (Koren, 2010), tức là các tham số \mathbf{w}_{uk} và \mathbf{h}_{ik} sẽ được cập nhật theo công thức:

$$\mathbf{w}_{uk}^{\text{new}} = \mathbf{w}_{uk}^{\text{old}} - \beta \left(\frac{\partial O^{MF}}{\partial \mathbf{w}_{uk}^{\text{old}}} \right) \quad (4)$$

$$= \mathbf{w}_{uk}^{\text{old}} + 2\beta(r_{ui} - \hat{r}_{ui})\mathbf{h}_{ik}$$

$$\mathbf{h}_{ik}^{\text{new}} = \mathbf{h}_{ik}^{\text{old}} - \beta \left(\frac{\partial O^{MF}}{\partial \mathbf{h}_{ik}^{\text{old}}} \right) \quad (5)$$

$$= \mathbf{h}_{ik}^{\text{old}} + 2\beta(r_{ui} - \hat{r}_{ui})\mathbf{w}_{uk}$$

Với β là tốc độ học (learning rate, $0 < \beta < 1$). Quá trình cập nhật sẽ thực hiện đến khi đạt độ lỗi chấp nhận được hoặc lặp lại đến số lần lặp quy định trước.

2.2 Phân rã ma trận ba chiều (Tensor Factorization – TF)

Tensor Factorization (TF) (Kolda and Bader, 2009; Dunlavy *et al.*, 2011) là một dạng tổng quát của kỹ thuật phân tích ma trận. Cho một tensor \mathbf{Z} với kích thước $U \times I \times T$, với 2 thành phần đầu tiên U và I lần lượt thể hiện cho ma trận nhân tử *user* và *item* trong kỹ thuật phân rã ma trận (MF) được đề cập ở phần trước. Thành phần thứ 3 thể hiện cho