

Báo cáo tuần 1

Tìm hiểu Two-Tower Recommender Architectures & Evaluation metrics in recommender systems



Người phụ trách: Anh Ngô Văn Vĩ

Người thực hiện: Nguyễn Đình Hiếu

1, Two-Tower Recommender Architectures

1.1, Kiến trúc (mô hình) Two-Tower là gì?

Mô hình Two-Tower tách riêng việc tính toán đặc trưng cho User và Item bằng hai mạng neural riêng biệt, gọi là "hai tháp" (two towers):

- User Tower: Nhận đầu vào là các thông tin liên quan đến người dùng như *user ID*, *demographics*, *historical interactions*, *device*, *context*, ... Các dữ liệu này được xử lý qua các layers (*embedding layers*, *MLPs*, *RNNs*, ...) để đưa ra output là một vector duy nhất gọi là the user embedding u (em nghĩ là “vector nhúng người dùng u ”). Vector u này biểu diễn sở thích, hành vi và đặc điểm người dùng trong một *dense, low-dimensional space*.

$$u = \text{Tower_User}(\text{User_Features})$$

- Item Tower: Tương tự như vậy, tháp Item cũng nhận đầu vào là các đặc trưng của Item (*item ID*, *category*, *description*, *image features*, ...) Và xử lý những dữ liệu này qua các layers của riêng nó tạo ra vector nhúng v . Vector này biểu diễn các đặc trưng của Item trong cùng một không gian với vector u .

$$v = \text{Tower_Item}(\text{Item_Features})$$

Để tính độ tương thích giữa user u và item v , thay vì phải đưa toàn bộ đặc trưng vào một mạng lớn duy nhất, Two-Tower chỉ cần tính một similarity function đơn giản. Hàm similarity function thường dùng là:

- Tích vô hướng: $\text{Score}(u, v) = u \cdot v$ (phổ biến nhất)
- Cosine Similarity: $\text{Score}(u, v) = (u \cdot v) / (\|u\| \|v\|)$

Training: Cả hai Tower (User và Item) được huấn luyện cùng nhau (end-to-end). Mục tiêu là học ra các không gian vector nhúng (embedding spaces) sao cho:

- Điểm tương đồng (ví dụ như tích vô hướng $u \cdot v$) giữa người dùng u và các sản phẩm phù hợp v là cao.
- Và giữa người dùng và các sản phẩm không phù hợp là thấp.

1.2, Tại sao lại dùng mô hình Two-Tower?

Đặc điểm lớn nhất của mô hình Two-Tower là việc tách riêng Users và Item. Đối với các mô hình RecSys (hệ gợi ý) khác, việc có hàng tỉ dữ liệu khiến cho chúng kém hiệu quả nhưng Two-Tower có khả năng Scale rất tốt nhờ việc tính trước các giá trị v cho **toàn bộ** Items.

Khi một user request đến chỉ cần tính giá trị vector u của user rồi lấy top-K ứng viên giá trị vector v từ Items.

Vấn đề: Việc tính toán $u \cdot v$ cho hàng tỉ Items lại quá chậm trong real-time nên ở đây ta xử lý bằng kỹ thuật tìm kiếm láng giềng gần xấp xỉ (Approximate Nearest Neighbor – ANN) hiệu quả để tìm ra top-K ứng viên trong mili giây.

Sau khi lọc ra khoảng vài nghìn ứng viên phù hợp ta có thể dùng tiếp các mô hình RecSys khác để tính toán và sắp xếp lại gợi ý tốt nhất.

Ưu điểm: Khả năng mở rộng cao (Highly Scalable); Có thể cải tiến tower người dùng hoặc tower item một cách độc lập, dễ bảo trì và mở rộng.

Hạn chế: Vì tách riêng Users và Items nên mô hình không thể học được các đặc trưng tương tác giữa User-Item. Tuy nhiên điều này có thể khắc phục ở bước sau khi đã thu hẹp các Items phù hợp bằng mô hình Two-Tower thì sử dụng thêm mô hình RecSys khác hiệu quả cho bộ dữ liệu nhỏ này.

Vẫn chưa khắc phục được vấn đề User mới \rightarrow ít dữ liệu \rightarrow kết quả gợi ý kém so với các SysRec khác.

2, Evaluation metrics in recommender systems

Evaluation Metrics là các chỉ số được dùng để đánh giá chất lượng gợi ý mà hệ thống tạo ra. Tùy theo mục tiêu, các metric được chia thành nhiều nhóm:

2.1, Candidate Generation Metrics

Candidate Generation Metrics đóng vai trò trong hiệu quả của RecSys lọc ra các lựa chọn phù hợp nhất với người dùng dựa trên sở thích, lịch sử tìm kiếm và hành vi của người dùng.

Một số metric quan trọng trong Candidate Generation:

- Novelty: Gợi ý item mới, lạ với người dùng.
- Diversity: Các item trong danh sách nên đa dạng, không trùng lặp.
- Serendipity: Gợi ý bất ngờ nhưng vẫn đúng sở thích.
- Catalog Coverage: Bao phủ được bao nhiêu % item trong toàn bộ hệ thống.
- Distributional Coverage: Đảm bảo phân phối item được gợi ý không quá lệch về các item phổ biến.

2.2, Predictive Metrics

Predictive Metrics là các chỉ số dùng để đánh giá độ chính xác dự đoán của hệ thống gợi ý – tức là hệ thống dự đoán sở thích hoặc điểm đánh giá của người dùng tốt đến đâu.

- RMSE (Root Mean Squared Error): Sai số căn bậc hai trung bình
- MAE (Mean Absolute Error): Sai số tuyệt đối trung bình
- MSE (Mean Squared Error): Sai số bình phương trung bình
- MAPE (Mean Absolute Percentage Error): Sai số phần trăm tuyệt đối trung bình
- R^2 (R-squared): Hệ số xác định (đo mức độ phù hợp của mô hình)
- Explained Variance: Tỷ lệ phương sai được giải thích bởi mô hình

2.3, Predictive Metrics

Ranking-Based Metrics là các chỉ số đánh giá chất lượng thứ tự sắp xếp các item mà hệ thống gợi ý tạo ra, nhằm đo lường mức độ hài lòng của người dùng với danh sách gợi ý.

- Mean Reciprocal Rank (MRR): Đo vị trí của item đúng đầu tiên trong danh sách gợi ý.
- ARHR: Phiên bản có trọng số của MRR, ưu tiên item đúng ở top đầu.
- nDCG (Normalized Discounted Cumulative Gain): Đánh giá chất lượng thứ tự toàn bộ danh sách, với trọng số giảm dần theo vị trí.
- Precision: Tỷ lệ item đúng trong top-K gợi ý.
- Recall: Tỷ lệ item đúng được gợi ý trong top-K so với tổng số item đúng.
- F1: Trung bình điều hòa của Precision và Recall.
- Average Recall: Trung bình Recall trên nhiều người dùng.
- Average Precision: Trung bình Precision trên nhiều người dùng.
- MAP (Mean Average Precision): Trung bình các giá trị Average Precision trên toàn bộ người dùng.

2.4, Business Metrics

Business metrics đóng vai trò quan trọng trong việc đánh giá hiệu quả và tác động của RecSys.

- Click-through Rate (CTR): Đo số lượt click vào các mục được gợi ý. Chỉ số phổ biến thể hiện mức độ hấp dẫn của gợi ý.
- Adoption & Conversion: Đo mức độ người dùng *thực sự tương tác sâu* với item, ví dụ: “Long CTR” (xem video đến một tỷ lệ nhất định), “Take rate” (thực sự xem sau gợi ý).
- Sales & Revenue: Phản ánh hiệu quả kinh doanh thực tế. Tuy nhiên khó tách biệt ảnh hưởng trực tiếp từ hệ thống gợi ý.
- Sales Distribution: So sánh doanh số trước và sau khi áp dụng hệ gợi ý. Giúp đánh giá ảnh hưởng đến sự đa dạng sản phẩm.
- User Engagement & Retention: Đo mức độ người dùng quay lại, ở lại lâu hơn nhờ hệ gợi ý (ví dụ như trên Spotify).

Nguồn tham khảo:

[1] *The Two-Tower Model for Recommendation Systems: A Deep Dive*

shaped.ai/blog/the-two-tower-model-for-recommendation-systems-a-deep-dive

[2] *A Comprehensive Survey of Evaluation Techniques for Recommendation Systems*

<https://arxiv.org/html/2312.16015v2>