Solution to the Personalized Accommodation Review Ranking Task via Tabular Data Approach

Yu Tokutake

The University of Electro-Communications

Outline

Introduction

User-generated Review, Personalized Review Ranking

RecTour 2024 Challenge

Task Description, Dataset, Evaluation Metrics

Solution

Basic Strategy, Candidate Generation, Candidate Ranking

Experiment

Baseline, Results

Conclusion

Outline

Introduction

User-generated Review, Personalized Review Ranking

RecTour 2024 Challenge

Task Description, Dataset, Evaluation Metrics

Solution

Basic Strategy, Candidate Generation, Candidate Ranking

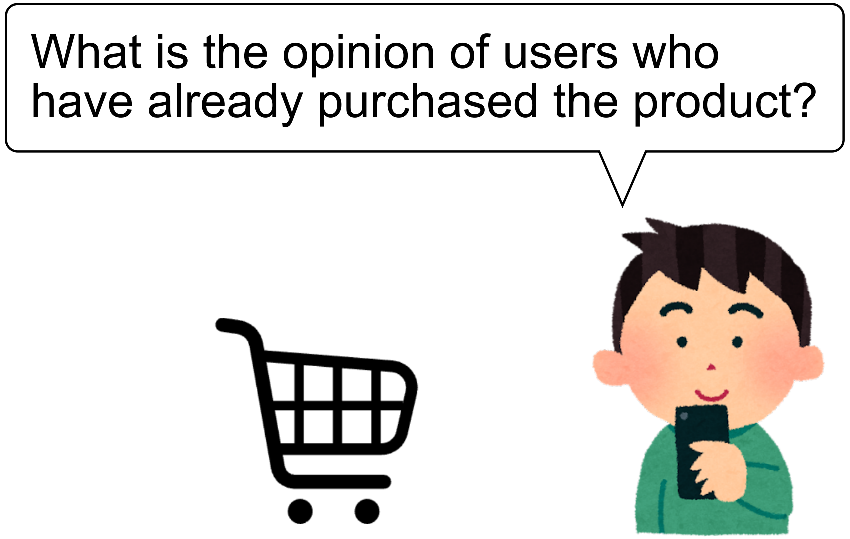
Experiment

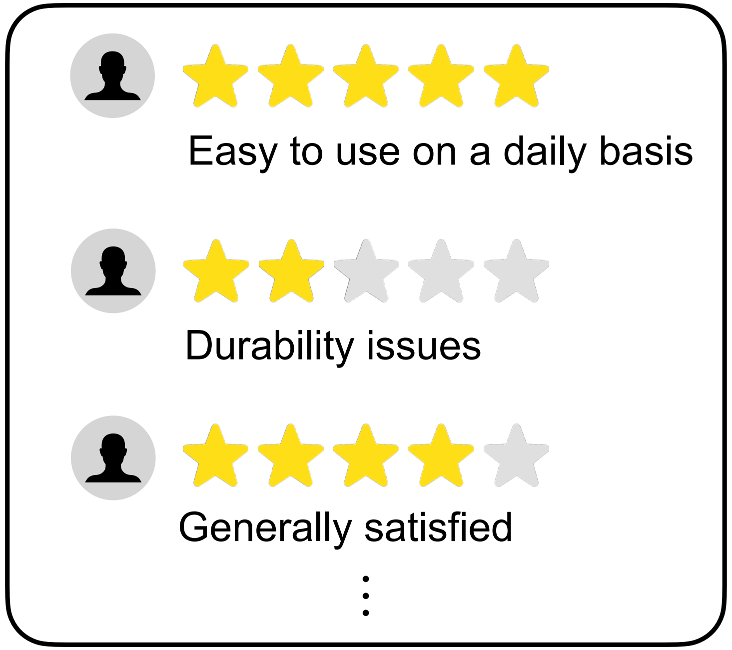
Baseline, Results

Conclusion

User-generated Review

Information to support user's decision-making

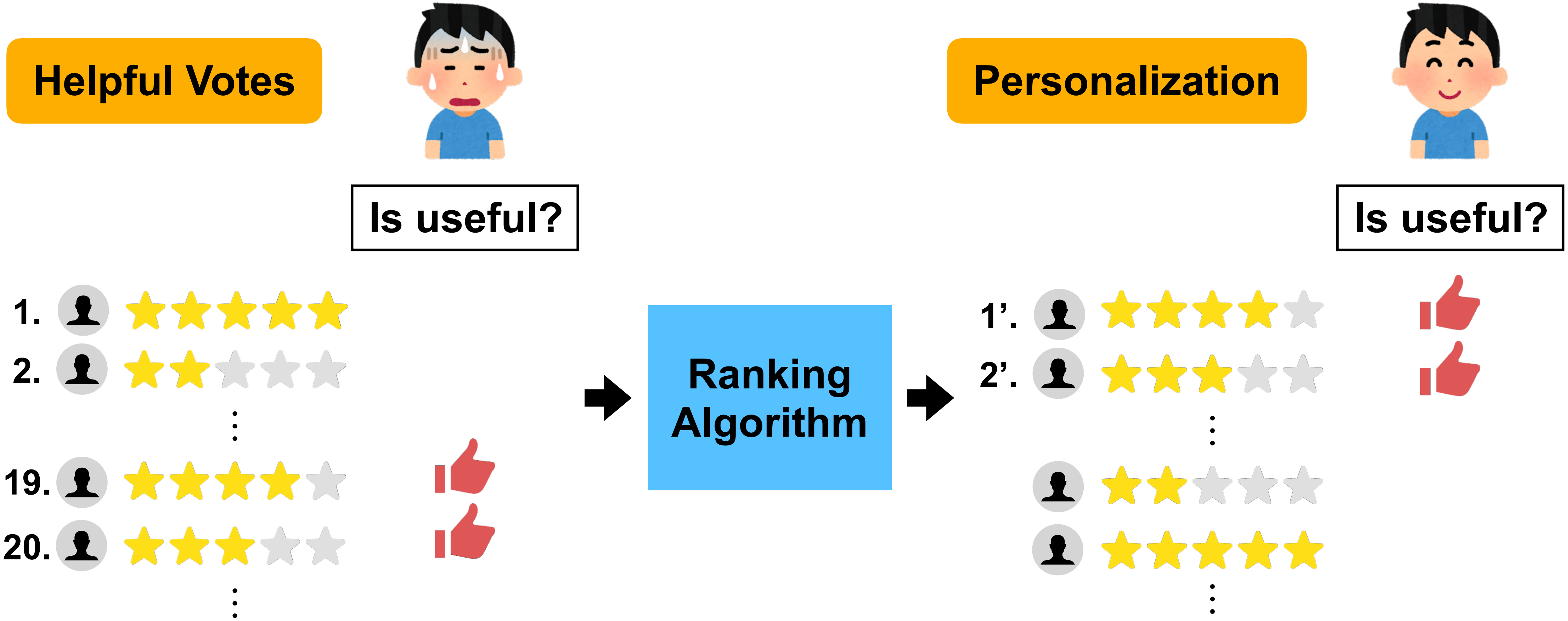




The number of reviews is increasing, and it is difficult for users to examine all reviews.

Personalized Review Ranking

Prioritize displaying reviews that are useful to the user



User-generated Review Dataset

The accommodation review dataset is on a smaller scale.

Dataset	Domain	# Reviews
Amazon Reviews'23	E-commerce	571.5M
Yelp	Restaurant	7.0M
Booking.com Crawling	Accommodation	515K

Booking.com hosted a competition, RecTour 2024 Challenge, using a comprehensive accommodation review dataset.

Outline

Introduction

User-generated Review, Personalized Review Ranking

RecTour 2024 Challenge

Task Description, Dataset, Evaluation Metrics

Solution

Basic Strategy, Candidate Generation, Candidate Ranking

Experiment

Baseline, Results

Conclusion

RecTour 2024 Challenge

Overview

- Match reviews to users and accommodations
- Given 2M reviews

This study

- Present a solution develop by Team ringo that won first place
- Employ tabular data approach

Challenge: Task Description

Task

- Predict the reviews generated by users for the accommodations they stayed at
- Submit the top 10 reviews

Application

Applying the developed algorithm to reviews of accommodations where users have not yet stayed can provide reviews that are close to the users' opinions.

Challenge: Dataset

Actual review data on the Booking.com platform

Data	Description
Users	Information about users and accommodations
Reviews	Review information generated for accommodations
Matches	Combinations of review by users for accommodations (only for training and validation)

Objective: Predict the matches of test set

Challenge: Dataset Statistics

	# Users	# Accommodations	# Reviews
Training	1,628,989	40,000	1,628,989
Validation	203,787	5,000	203,787
Test	199,138	5,000	199,138

Note:

- Users were unique and correspond one-to-one with reviews
- Each accommodation had at least 10 reviews
- No common accommodations among the sets

Challenge: Evaluation Metrics

The top 10 predicted reviews were evaluated to determine whether they matched the reviews generated by users.

- Mean Reciprocal Rank (MRR) @10

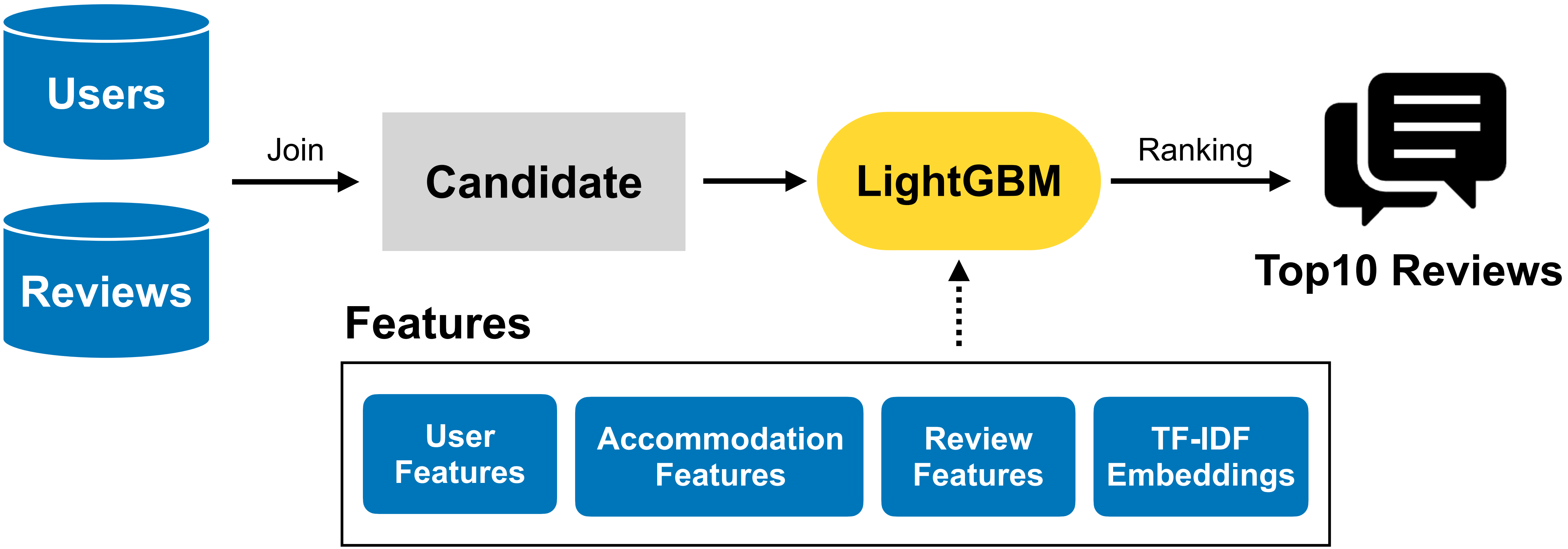
$$\mathrm{MRR@10} = \frac{1}{|U|} \sum_{u \in U} \begin{cases} \frac{1}{\mathrm{rank}_u} & (\text{if } \mathrm{rank}_u \leq 10) \\ 0 & (\text{otherwise}) \end{cases}$$

- Precision@10

$$\mathrm{Precision@10} = \frac{1}{|U|} \sum_{u \in U} \mathrm{I}\left[\mathrm{rank}_u \leq 10\right]$$

Outline

Solution: Overview

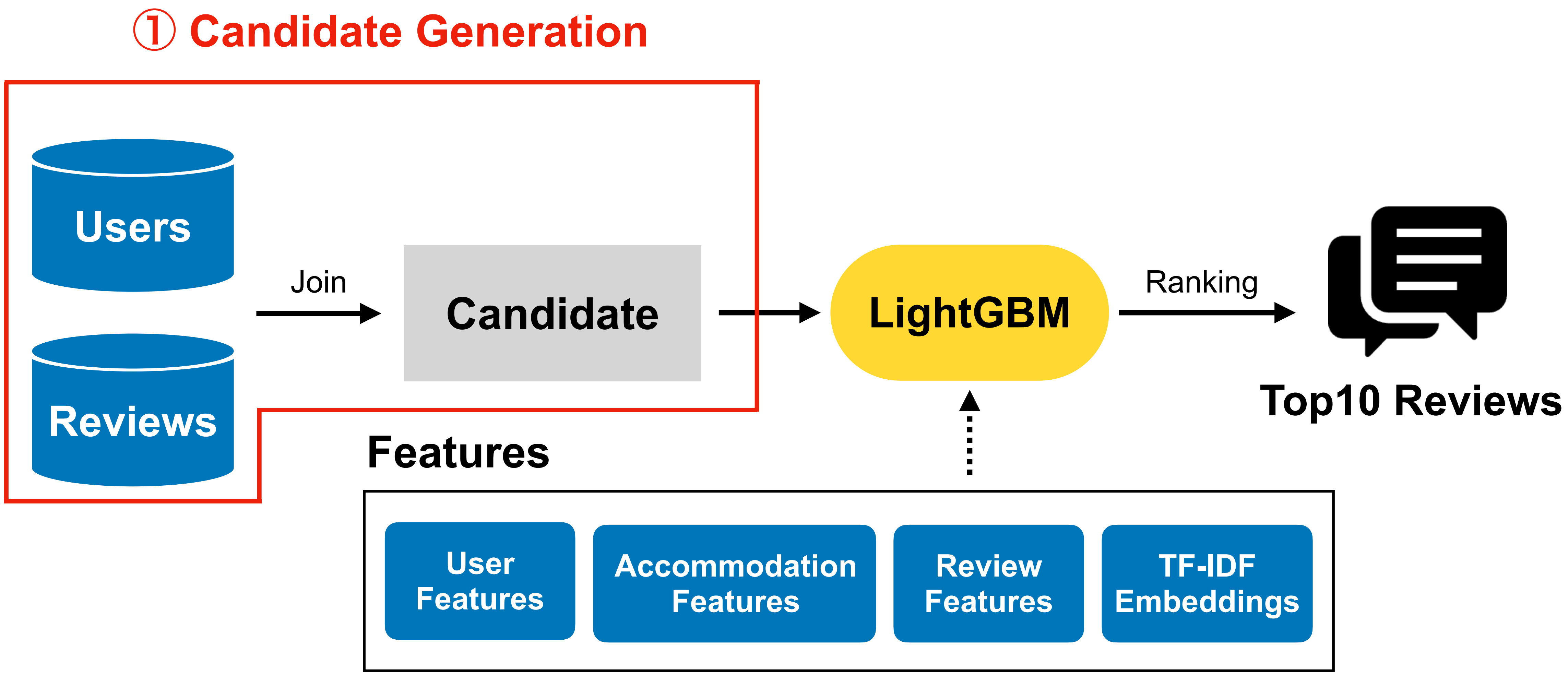


Solution: Strategy

Employ tabular data approach

- Tabular data approach
    1. Feature extraction
    2. Build a supervised model that predicts whether a review was generated by a user for an accommodation
- Inspired by two-stage recommendation approach
    1. Candidate generation
    2. Re-ranking

Solution



① Candidate Generation
Users
Reviews
Join
Candidate
LightGBM
Ranking
Top10 Reviews
Features
User Features
Accommodation Features
Review Features
TF-IDF Embeddings

Candidate Generation: Strategy

- Recent recommendation task
  - Huge number of users and item combination (10B ~ 100B)
  - Not all combination can be used for training and prediction

- This challenge
  - Constraints on the number of combination
  - All combination can be used as candidates

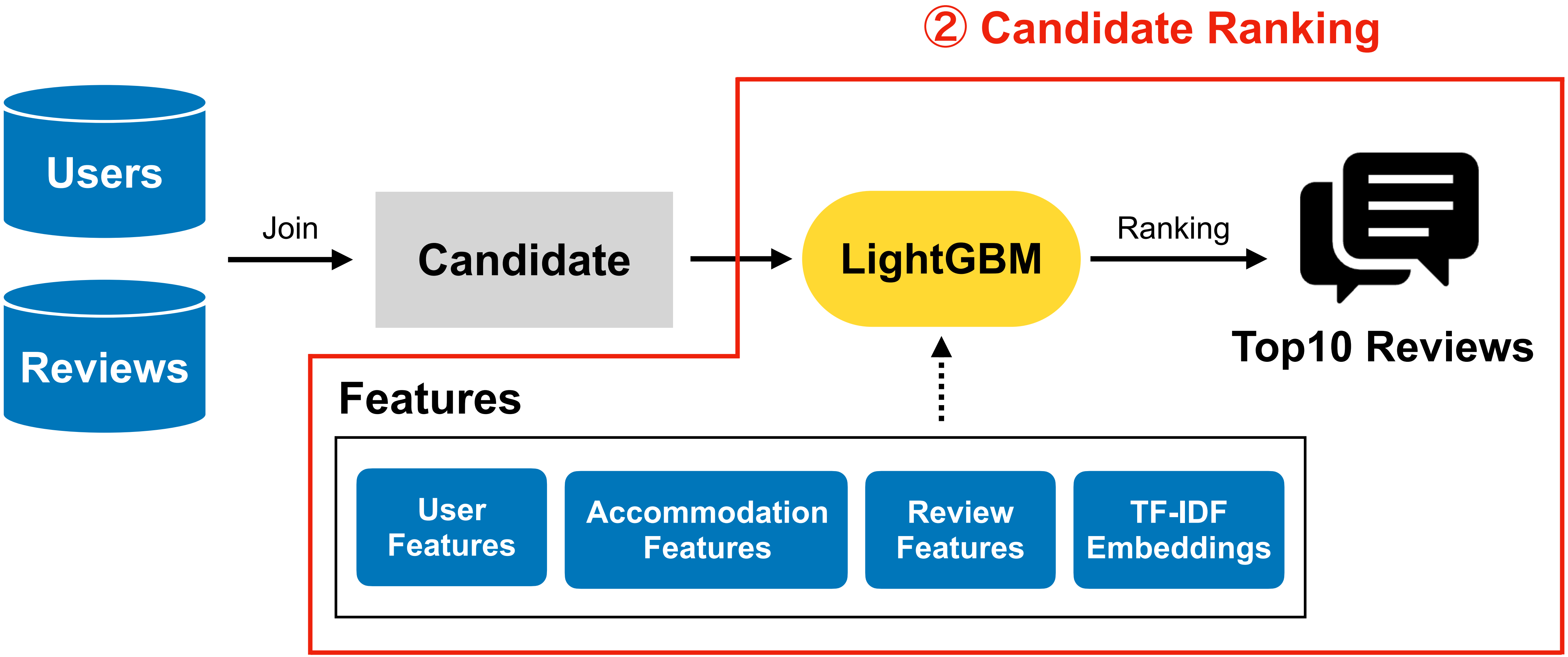
Candidate Generation: Procedure

1. Join users (user_id, accommodation_id) and reviews (review_id, accommodation_id)
2. Merge matches (**Add binary ground truth**)

→ Formulate as a binary classification task

	# Candidates	# Positive	# Negative	Positive:Negative
Training	214,311,737	1,628,989	212,682,748	1 : 131
Validation	29,676,751	203,787	29,472,964	1 : 145
Test	24,066,438	199,138	23,867,300	1 : 120

Solution


② Candidate Ranking
Users
Reviews
Join
Candidate
LightGBM
Ranking
Top10 Reviews
Features
User Features
Accommodation Features
Review Features
TF-IDF Embeddings

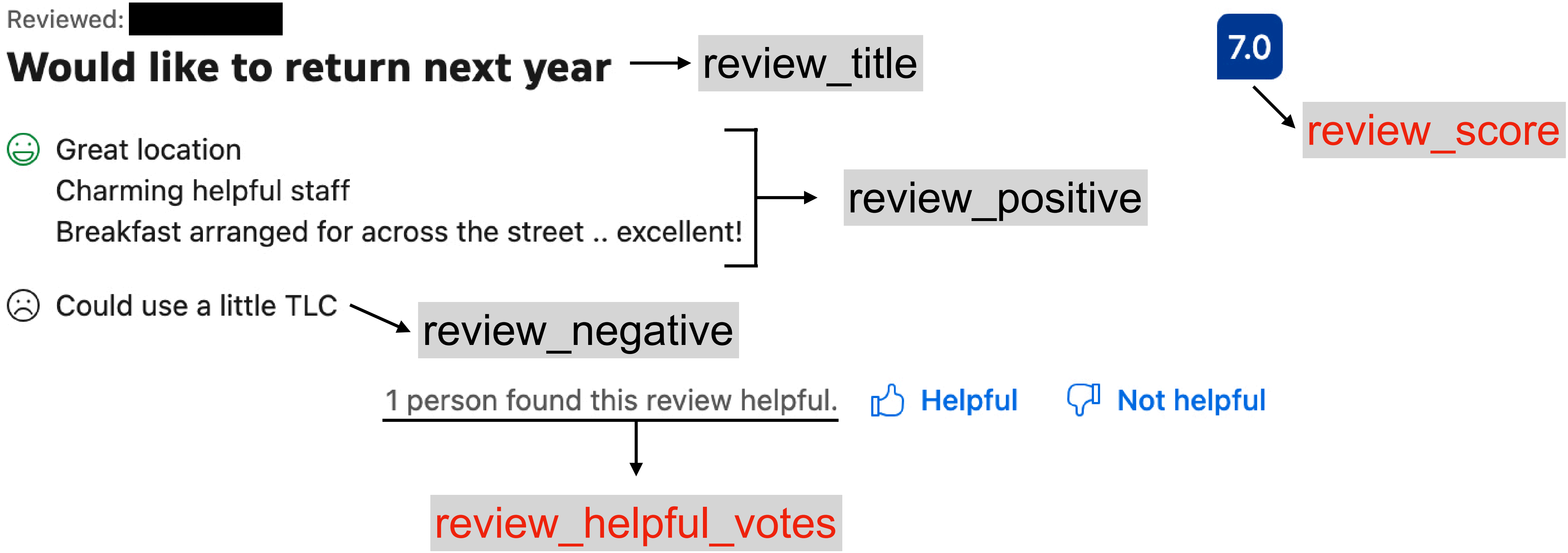
Features for Candidate Ranking

Most of the features used are derived from the original data

Type	Features
User	`guest_type`, `guest_country`, `room_nights`
Accommodation	`accommodation_type`, `accommodation_country`, `accommodation_score`, `accommodation_star_rating`, `location_is_beach`, `location_is_ski`, `location_is_city_center`

Features for Candidate Ranking

Review Features



Features for Candidate Ranking

<u>Added Features</u>

- Aggregate features
    - Frequency of each accommodation
    - Average score of each accommodation
    - Review text length
- Sentiment analysis score using a RoBERTa-based model (only `review_title`)

Features for Candidate Ranking

Added Features

One of the model variations used TF-IDF embeddings of user accommodation and review data.

- Concatenate features from original data: `<field_name>:<field_value>\n`
- Reduced to 100 dimensions each using ICA

Outline

Introduction

User-generated Review, Personalized Review Ranking

RecTour 2024 Challenge

Task Description, Dataset, Evaluation Metrics

Solution

Basic Strategy, Candidate Generation, Candidate Ranking

Experiment

Baseline, Results

Conclusion

Experiment

EQ1：Does performance improve by changing the number of negative candidates in the training data (candidates) or by adding TF-IDF embeddings as features?

EQ2：How does the proposed method perform compared to the baseline methods?

The number of positives in the training set is very small compared to the negatives (1:131).

→ **Randomly undersampling of negative candidates**

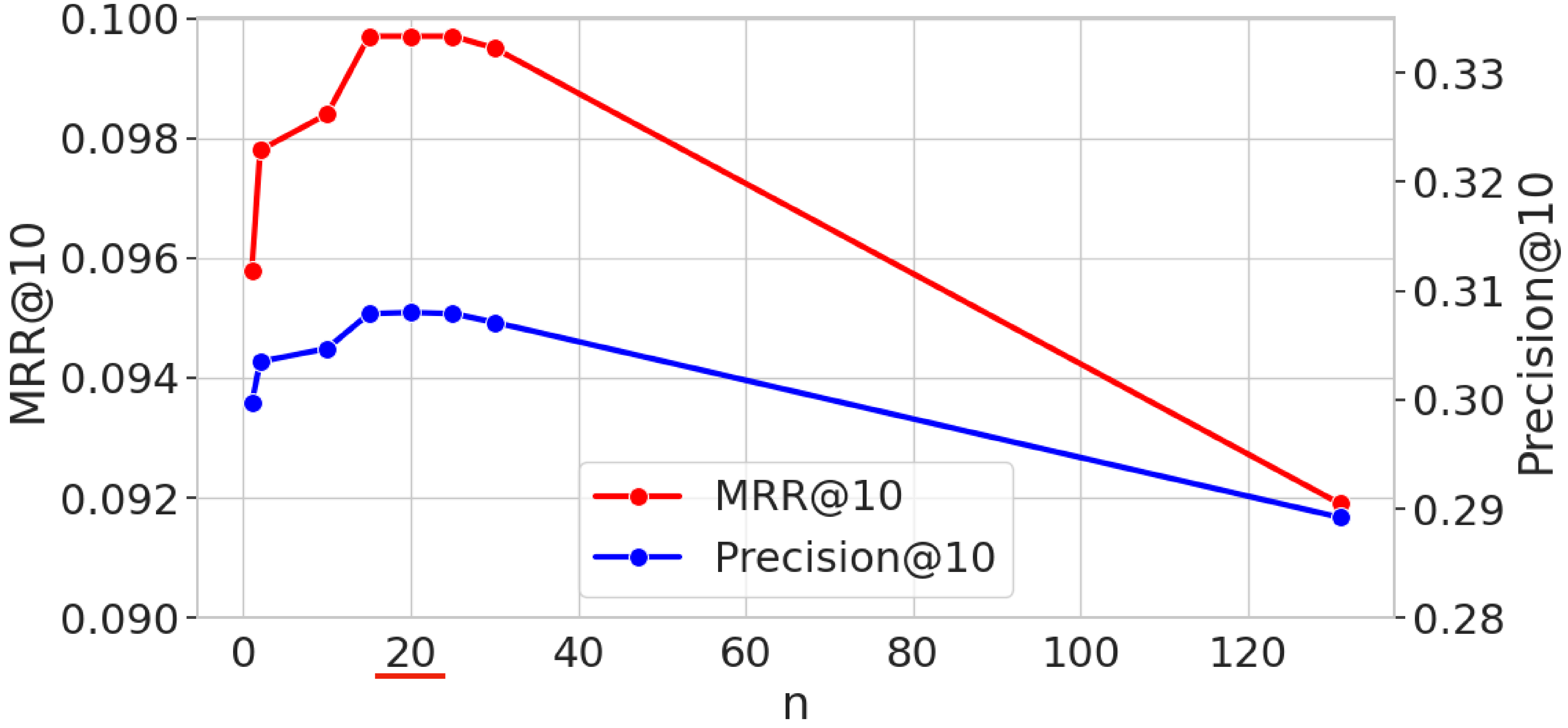
Experiment: Baseline

Baseline	Description
RAND	Randomly select 10 reviews from possible candidates
Helpful Votes	Select top 10 reviews from the candidates based on `review_helpful_votes`
LGBM	Proposed method

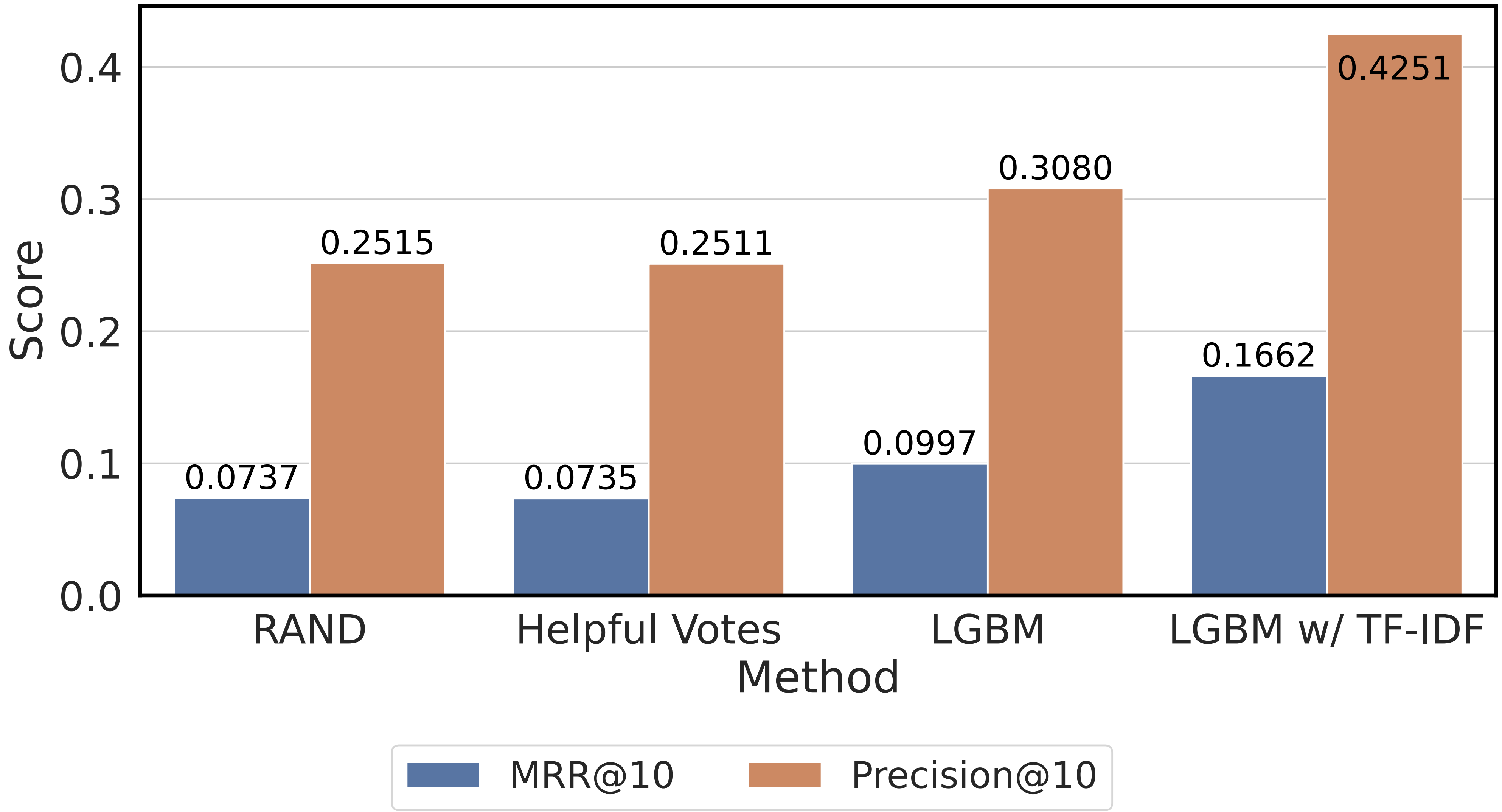
LGBM changes the ratio n of negative to positive in training set.

- $n \in \{1, 2, 10, 15, 20, 25, 30, 131\}$ ($n = 131$ is the original ratio)

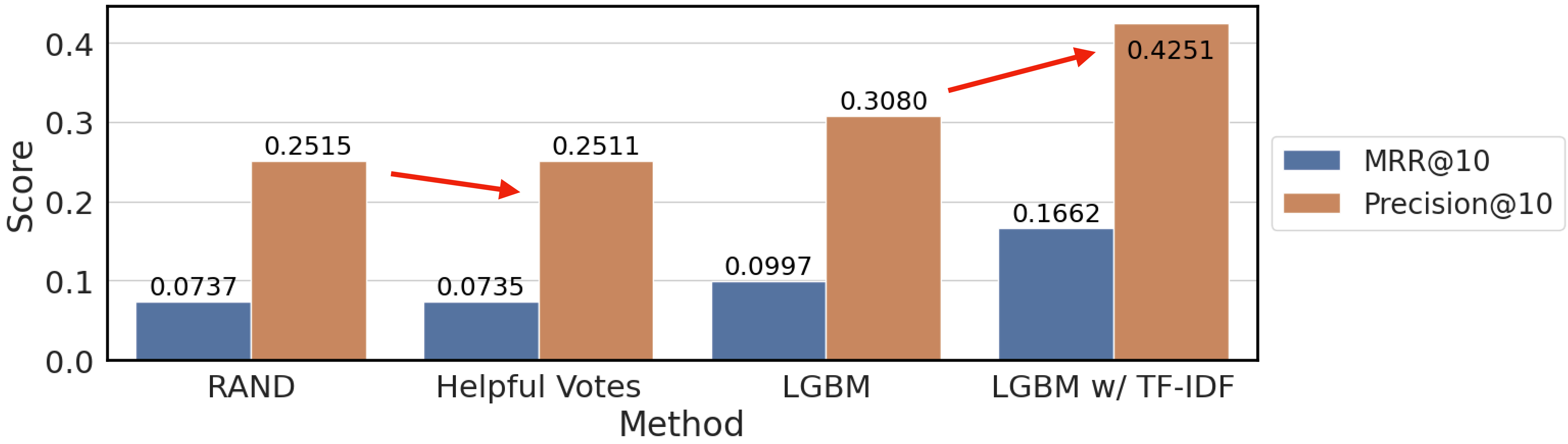
EQ1: LGBM w/o TF-IDF (Change n)



EQ2: Comparison with Baselines



EQ2: Comparison with Baselines



Helpful Votes

- Many reviews are sparse.
- **Helpful Votes < RAND**

LGBM

Improved performance by using TF-IDF embedding as features

Insights and Reflections

*1: R. Igebaria, et al., "Enhancing Travel Decision-Making: A Contrastive Learning Approach for Personalized Review Rankings in Accommodations," arXiv:2407.00787, 2024.

- Room for improvement since LGBM did not outperform the Booking.com-provided baseline

	LGBM w/ TF-IDF	**Booking.com**[*1]
Precision@10	0.425	0.549

- Feature extraction was difficult
    - Could not express user preferences due to all users being unique (e.g., review perspective, average rating)

Outline

Introduction

User-generated Review, Personalized Review Ranking

RecTour 2024 Challenge

Task Description, Dataset, Evaluation Metrics

Solution

Basic Strategy, Candidate Generation, Candidate Ranking

Experiment

Baseline, Results

Conclusion

Conclusion

Summary

- Propose review ranking algorithm using tabular data approach
- Improved prediction accuracy by devising training data and features (EQ1)
- Outperformed helpful votes performance (EQ2)

Future studies

- Improve undersampling methods
- Integration with other NLP approaches (e.g., LLM's fine-tuning)

Thank you!

If you have any question, please contact

tokutakeyuu@uec.ac.jp