



CIKM 2022

**AnalytiCup - EvalRS: a Rounded Evaluation of Recommender Systems**

# **Track2Vec: Fairness Music Recommendation with a GPU-Free Customizable-Driven Framework**

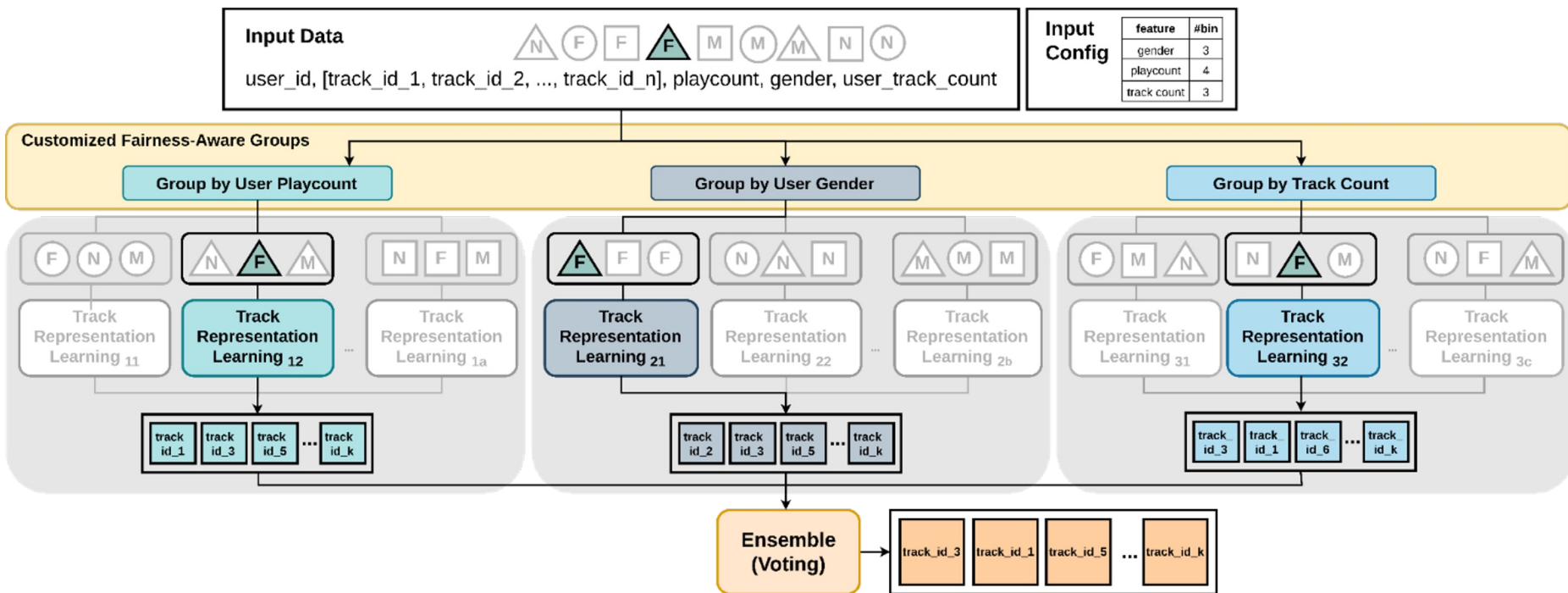
**Wei-Wei Du, Wei-Yao Wang, Wen-Chih Peng**  
National Yang Ming Chiao Tung University, Taiwan

Oct. 21, 2022

# Introduction

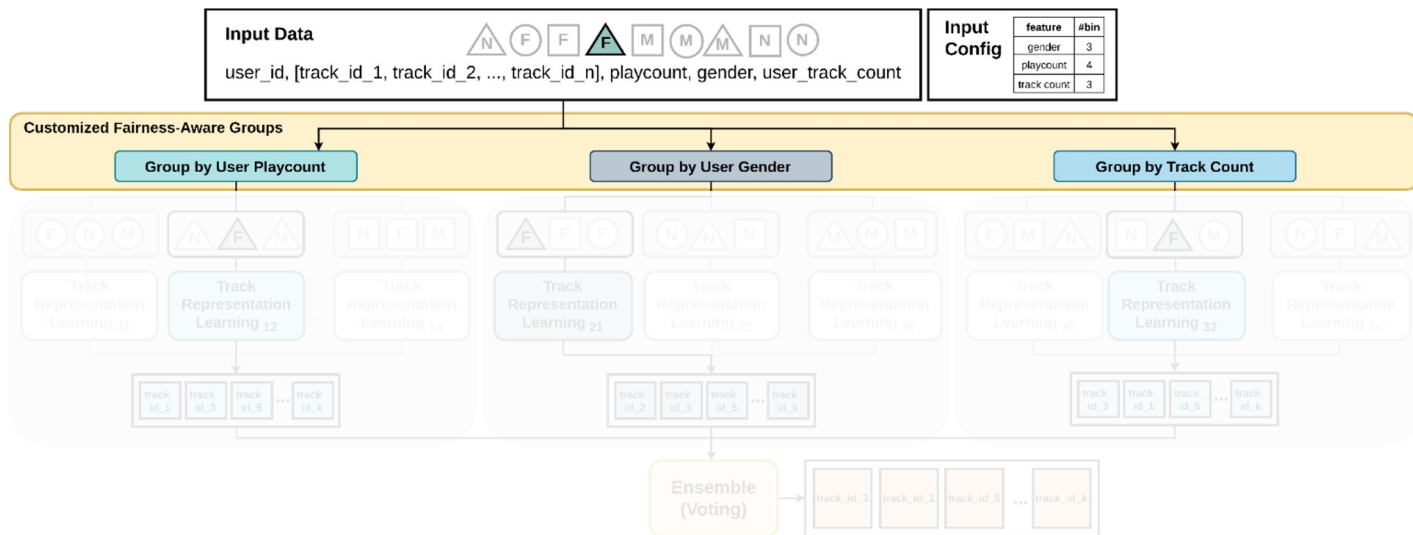
- We propose **Track2Vec** as a fairness recommendation system with a **customizable-driven framework** in a **GPU-free** environment
  - **Customized fairness-aware groups** for modeling different features
  - **Track representation learning module** for learning better user embedding
  - **Ensemble module** for ranking the recommendation results from different track representation learning modules
- Also, we introduce a novel metric, **MR-ITF**, to measure the predictive distribution of the model by **weighting importance based on the number of predictions of each class**
- Experiments show that our framework achieves **4th prize** ranking

# Model Framework: Track2Vec



# 1. Customized Fairness-Aware Groups

- **Discretize each feature** based on the feature distribution
- **Bunch users into different groups** by the customizable input configuration to avoid the unbalance issue (e.g., majority dominating the model behavior)

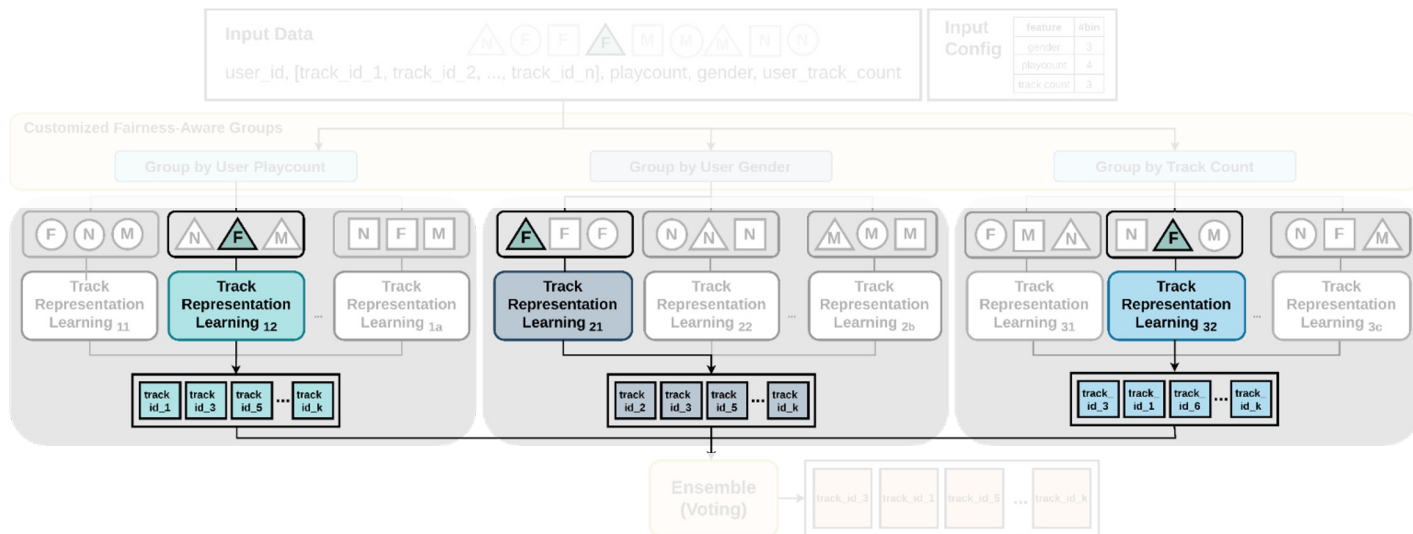


# 1. Customized Fairness-Aware Groups

- Discretize each feature based on the feature distribution
- Bunch users into different groups by the customizable input configuration to avoid the unbalance issue (e.g., majority dominating the model behavior)
- In this paper, we use **user playcount, user gender and track count** as the three most important factors
- Details
  - The playcount group and track count group use **logarithmic bucketing** in base 10
  - The gender group divides each sequence into male, female and neutral

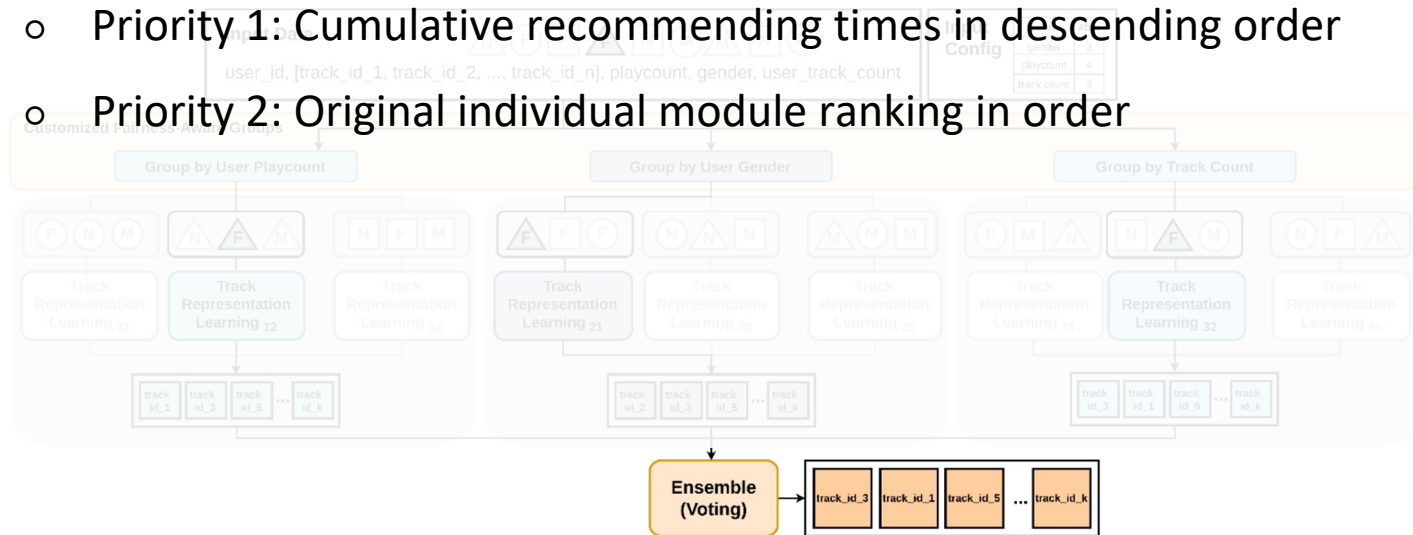
## 2. Track Representation Learning

- Due to **time constraint** in this challenge, we focus on an **unsupervised method**
- We use **Word2Vec** to train track embeddings by calculating the interactions between tracks



# 3. Ensemble Techniques

- To produce more robust and diverse recommendations, voting is used for ensembling each group with different priorities
- Ensemble re-ranking strategy
  - Priority 1: Cumulative recommending times in descending order
  - Priority 2: Original individual module ranking in order



# Evaluation Metric: MR-ITF

- Previous Problems
  - **HR**, **nDCG** and **MRR** fail to reflect the model behavior
  - **MRED**, **being less wrong** and **latent diversity** [1] require human settings for the number of bins and hard to generalize
- Motivation: TF-IDF (term frequency - inverse document frequency)

[1] P. J. Chia, J. Tagliabue, F. Bianchi, C. He, B. Ko, Beyond ndcg: behavioral testing of recommender systems with reclist, in: Companion Proceedings of the Web Conference 2022, 2022, pp. 99–104.



# Evaluation Metric: MR-ITF

## (Miss Rate - Inverse ground Truth Frequency)

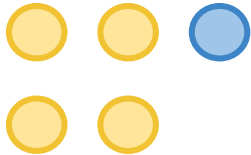
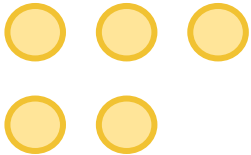
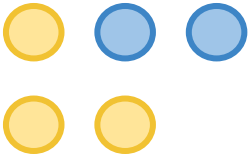


$$MR - ITF = \frac{\sum_{i=1}^{|C|} MR_i \times ITF_i}{\sum_{j=1}^N MR_j}, \quad (1)$$

$$ITF_i = \log\left(\frac{\#total\ predictions}{\#track_i}\right), \quad (2)$$

$C$  is the number of classes,  $N$  is the number of total instances,

$MR_i$  is the miss-rate of the  $i$ -th track

# Evaluation Metric: MR-ITF

	Ground Truth	Model 1 prediction	Model 2 prediction
			
Hit rate		80%	80%
MR-ITF		-25 	-0.39 

# Experiment: Ablation Study

- Track2Vec **performs the best** compared to each module

	G	P	U	G+P	G+U	P+U	Track2Vec (ours)
Standard RSs metrics (1)	0.0103	0.0118	0.0128	0.0127	0.0136	<u>0.0143</u>	<b>0.0146</b>
Standard metrics on a per-group (2)	-0.0073	<b>-0.0039</b>	-0.0061	-0.0052	-0.0055	<u>-0.0044</u>	-0.0055
Behavioral tests (3)	-0.0138	0.0014	0.0008	0.0188	0.0156	<u>0.0223</u>	<b>0.0271</b>
MR-ITF (ours)	-4.3862	-4.3863	<u>-4.3861</u>	-4.3862	<u>-4.3861</u>	<b>-4.3860</b>	<u>-4.3861</u>
Total Score	-0.0048	0.0008	-0.0003	0.0041	0.0035	<u>0.0057</u>	<b>0.0062</b>

G: Gender. P: Playcount. U: User track count.

Total score is computed as  $((1) + (2) + (3)) / 3$

# Experiment: Testing Performance

- In phase 2, our method won **4th** prize among all the participants

Rank	Model	Score	Standard RSs metrics	Standard metrics on a per-group	Behavioral tests
4	Track2Vec	1.1847	0.0088	2.9481	0.2050
-	CBOWRecSysBaseline	-1.2122	0.0512	-3.7194	0.4527
Improvements (%)	-	<b>198</b>	-83	<b>179</b>	-55

# Conclusion

- We propose a framework called **Track2Vec** as a fairness recommendation system including three modules, customizable-driven groups, track representation learning and an ensemble technique
- We introduce **MR-ITF** by weighting different degrees of importance for each class based on the corresponding frequency
- In phase 2, our framework achieves **4th prize** ranking and outperforms the official baseline about 200% **without any GPU**

CIKM 2022

**AnalytiCup - EvalRS: a Rounded Evaluation of Recommender Systems**

**Thanks for Your Attention!**

Code

