Cheng YANG, Qiang FAN, Tao WANG, Gang YIN, Xun-hui ZHANG, Yue YU, Hua-min WANG, 2019. RepoLike: a multi-feature-based personalized recommendation approach for open-source repositories. *Frontiers of Information Technology & Electronic Engineering*, 20(2):222-237. https://doi.org/10.1631/FITEE.1700196

RepoLike: a multi-feature-based personalized recommendation approach for open-source repositories

Key words: Social coding; Open-source software; Personal recommendation; GitHub

Corresponding author: Cheng YANG
E-mail: delpiero710@126.com

ORCID: Cheng YANG, http://orcid.org/ 0000-0002-4782-1645

Motivation

1. External contributors are playing a significant role in open-source development.

2. It takes contributors a lot of time to find suitable projects or tasks from thousands of open-source projects in the communities to work on.

Main idea

1. A technical-interest measurement model is proposed, which quantitatively measures the developers' personal interests in joining repositories from technical and social dimensions.

2. A personalized recommendation algorithm using two different approaches including linear combination and a learning-to-rank approach is designed based on the above-mentioned metrics.

Method

RepoLike includes the following four stages:

1. acquisition of the historical development activities of developers;

2. construction of the association network;

3. recommendation model training;

4. recommended test.

Method

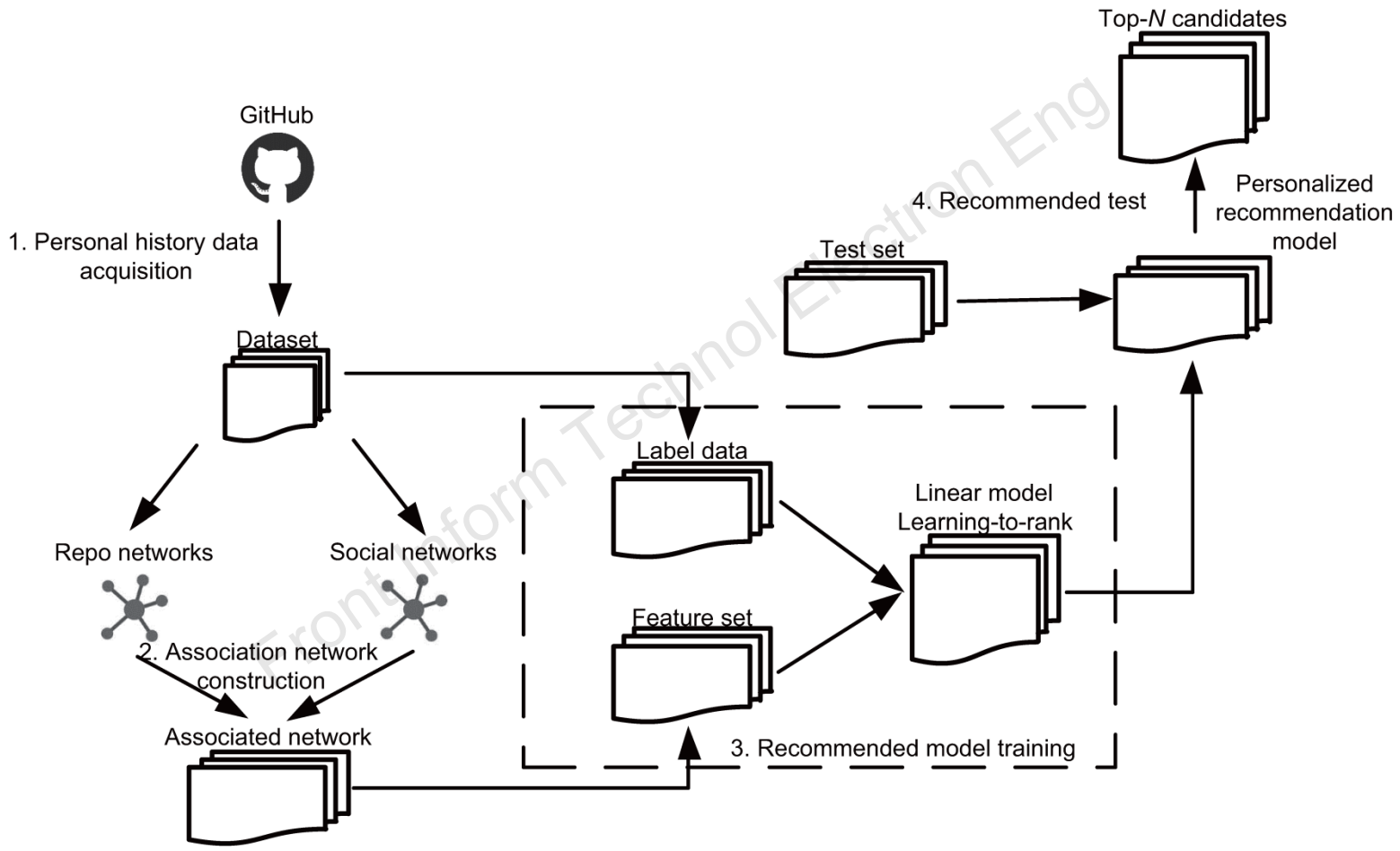

Fig. 3 Overview of the personalized recommendation framework

Major results

1. Comparison of recommendation performance based on different dimensions of features

Table 1 Comparison of recommendation performance based on different dimensions of features

Feature dimension	Hit rate (%)			
	Top 5	Top 10	Top 15	Top 20
TL	6.04	8.52	10.45	11.77
SL	6.24	8.47	10.40	11.60

TL: technical dependence; SL: social relevance degree

Major results

2. Comparison of recommendation performance for combinations of two-dimensional information

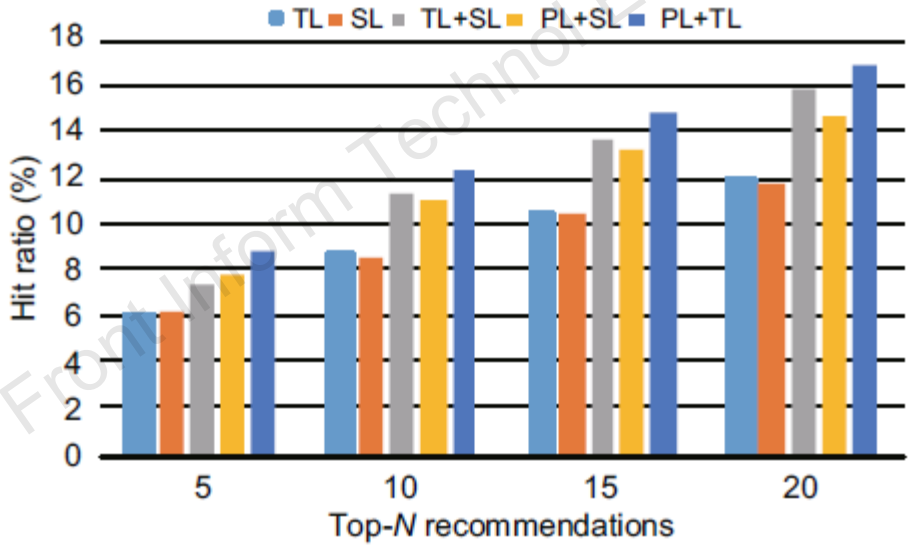

Fig. 6 Comparison of recommendation performance for combinations of two-dimensional information

TL: technical dependence; SL: social relevance degree; PL: project popularity

Major results

3. Comparison of recommendation performance for combinations of multi-dimensional information

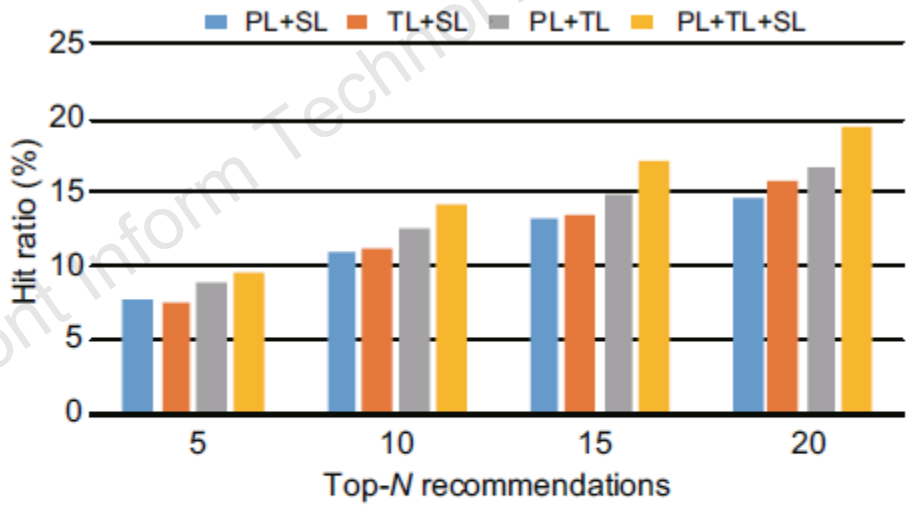

Fig. 7 Comparison of recommendation performance for combinations of multi-dimensional information

TL: technical dependence; SL: social relevance degree; PL: project popularity

Major results

4. Comparison of recommendation performance for the linear-combination model and LTR model

Table 2 Comparison of recommendation performance for linear-combination model and LTR model

Recommendation	Hit ratio (%)			MRR (%)		
	Linear mode	LTR	Rate of increase	Linear mode	LTR	Rate of increase
Top-5	9.46	11.44	20.93	0.0538	0.0657	22.12
Top-10	14.00	15.28	9.14	0.0599	0.0708	18.20
Top-15	17.02	17.44	2.47	0.0622	0.0725	16.56
Top-20	19.38	19.60	1.14	0.0635	0.0734	15.59

LTR: Learning-to-rank; MRR: mean reciprocal rank

Major results

5. Comparison of recommendation performance for the linear-combination model and LTR model after introducing additional social connection information

Table 3 Comparison of recommendation performance for the linear-combination model and LTR model after introducing additional social connection information

Recommendation	Hit ratio (%)			MRR (%)		
	Linear mode	LTR	Rate of increase	Linear mode	LTR	Rate of increase
Top-5	10.57	16.04	51.75	0.0609	0.1172	92.45
Top-10	15.07	20.20	34.04	0.0669	0.1227	83.41
Top-15	18.44	23.16	25.60	0.0695	0.1251	80.00
Top-20	20.99	25.04	19.29	0.0709	0.1262	78.00

LTR: Learning-to-rank; MRR: mean reciprocal rank

Conclusions

1. We have proposed a method called “RepoLike,” which explores multi-dimensional features to measure the potential correlation between a developer and a project.

2. The experimental results showed that our method can achieve high accuracy for recommending open-source repositories to developers.