

Gower's coefficients

Абрамов, Миахутдинов, Бабаян, Кашникова

1 Gower's similarity

Для двух объектов i и j с эмбеддингами длины m коэффициент сходства Говера вычисляется по формуле

$$S_{ij} = \frac{\sum_{k=1}^m w_{ijk} s_{ijk}}{\sum_{k=1}^m w_{ijk}},$$

где w_{ijk} неотрицательные веса, а s_{ijk} - близость двух объектов по k -ой координате

На основе этого коэффициента мы строим матрицы сходства для пользователей и/или товаров и сравниваем качество рекомендаций с базовым подходом на косинусном сходстве

2 Сравниваемые Алгоритмы

В качестве базовых моделей мы реализовали: KNN, EASER, fsSLIM, EIGENREC, HybridSVD.

3 Feature importance

Мы проверили гипотезу, что вместо равных весов w_{ijk} можно использовать веса, зависящие от "полезности" признаков. Веса оценивали с помощью моделей классического ML (например, регрессии для предсказания рейтинга). На практике такая схема не дала стабильного улучшения: качество чаще снижалось. Вероятная причина — важность признаков для предсказания рейтинга не совпадает с важностью для ранжирования в рекомендациях.

4 Основные результаты

В большинстве экспериментов наблюдаются небольшие улучшения по рассматриваемым метрикам. В отдельных случаях встречаются заметные (кратные) приросты по одной из метрик, однако эффект нестабилен.

Таблица 1: Результаты работы алгоритмов на Movielens 1m

Model	Is_Gower	w_type	HR	MRR	Cov
knn	False	None	0.006169	0.002085	0.270496
	True	None	0.003926	0.001064	0.105968
	True	builtin	0.005609	0.001493	0.178472
	True	shap	0.007291	0.001584	0.146960
	True	permutation	0.004487	0.001287	0.161461
EASER	False	None	0.064498	0.023526	0.288622
	True	None	0.064498	0.023526	0.288622
	True	builtin	0.064498	0.023621	0.288622
	True	shap	0.064498	0.023527	0.288901
	True	permutation	0.064498	0.023527	0.288622
fsSLIM	False	None	0.060572	0.021082	0.257111
	True	None	0.065059	0.021077	0.258784
	True	builtin	0.058890	0.019749	0.251534
	True	shap	0.060011	0.020449	0.258784
	True	permutation	0.062255	0.021245	0.253765
HySVD	False	None	0.008413	0.002012	0.013385
	True	None	0.023556	0.009088	0.013385
	True	builtin	0.015143	0.005775	0.014222
	True	shap	0.019069	0.006933	0.012828
	True	permutation	0.015143	0.005954	0.013943
EigenRec	False	None	0.045990	0.018376	0.085332
	True	None	0.047672	0.017062	0.098438
	True	builtin	0.038699	0.014502	0.114334
	True	shap	0.043186	0.016214	0.115170
	True	permutation	0.037577	0.015324	0.112660
Random			0.002243	0.001121	0.994701

Таблица 2: Результаты работы алгоритмов на датасете BookCrossing

Model	Is_Gower	HR	MRR	Cov
knn	False	0.017351	0.006230	0.848276
	True	0.017882	0.006551	0.823908
EASER	False	0.021335	0.007361	0.259770
	True	0.022486	0.008475	0.896552
fsSLIM	False	0.004338	0.001312	0.068046
	True	0.004338	0.001312	0.068046
EigenRec	False	0.009738	0.002605	0.191264
	True	0.007436	0.002605	0.435862
Random		0.004869	0.001183	1.0

5 Используемые статьи

1. HybridSVD: When Collaborative Information is Not Enough
2. EigenRec: Generalizing PureSVD for Effective and Efficient Top-N Recommendations
3. Item-based collaborative filtering recommendation algorithms
4. Random Forests
5. A Unified Approach to Interpreting Model Predictions
6. CatBoost: unbiased boosting with categorical features