

Выбор оптимальной размерности признакового пространства в реконструкции пользовательских предпочтений

А.А. Бородинов¹, В.В. Мясников^{1,2}

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34а, Самара, Россия, 443086

²Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация

В работе исследована проблема выбора оптимальной размерности признакового пространства в задаче реконструкции пользовательских предпочтений. Сама задача реконструкции предпочтений представлена как задача классификации пар примеров с переходом в пространство парных сравнений. Проведено экспериментальное исследование зависимости ошибки реконструкции предпочтений от размерности признакового пространства, выбора базиса перехода и метода машинного обучения.

Ключевые слова

Пространство большей размерности, реконструкция предпочтений, рекомендательные системы

1. Введение

В работе [1] авторы данной статьи представили новый метод реконструкции пользовательских предпочтений [2-5] с использованием метода парных сравнений [6,7] и исследовали его применение в транспортной рекомендательной системе [8,9]. В предложенном методе выполняется переход в признаковое пространство большей размерности. В данной работе исследуется зависимость ошибки реконструкции пользовательских предпочтений от размерности признакового пространства, базиса перехода и используемого метода машинного обучения.

2. Переход в признаковое пространство большей размерности

В работе используются три базиса для осуществления отображений φ^A и φ^S вида:
 $\varphi: X \rightarrow Y$
 $\mathbf{x} \mapsto y(\mathbf{x})$

- 1) Базис исходного представления.
- 2) Степенной (полиномиальный) базис.

$$y_k = \varphi_k(\mathbf{x}) = \prod_{n=0}^{N-1} x_n^{k_n}, \quad k = \sum_{n=0}^{N-1} K_0^n k_n$$

$$K = \prod_{n=0}^{N-1} K_n = K_0^N = \dim(Y) > \dim(X) = N.$$

- 2) Фурье- базис (гармонический)

$$y_k = \varphi_k(\mathbf{x}) = \prod_{n=0}^{N-1} \cos(\pi k_n x_n), \quad k = \sum_{n=0}^{N-1} K_0^n k_n$$

$$K = \prod_{n=0}^{N-1} K_n = K_0^N = \dim(Y) > \dim(X) = N.$$

3. Экспериментальные исследования

В экспериментальных исследованиях использовались реальные данные о предпочтениях пользователей в выборе маршрута передвижения. Каждый выбор пользователя описан вектором признаков с размерностью 9. Максимальный размер нового признакового пространства – 300, шаг увеличения размерности – 1. В работе рассматривался выбор оптимальной размерности пространства при малых значениях обучающей выборки. Результаты экспериментальных исследования представлены в таблицах 1 и 2.

Таблица 1

Зависимость вероятности ошибки и размерности пространства анализа от размеров обучающей выборки, полиномиальный базис

№ user	γ=1		γ=2		γ=3		γ=4		γ=5		γ=6		γ=7		γ=8		γ=9		γ=10	
	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A
1	0,387	16	0,253	19	0,214	17	0,184	30	0,213	24	0,177	13	0,170	22	0,193	29	0,169	19	0,171	20
2	0,204	24	0,210	15	0,206	20	0,166	16	0,177	21	0,179	24	0,188	20	0,187	15	0,193	21	0,172	29
3	0,214	23	0,184	16	0,178	17	0,244	16	0,196	13	0,174	10	0,202	20	0,188	21	0,184	20	0,179	18
4	0,190	14	0,180	20	0,251	12	0,184	10	0,188	16	0,196	20	0,203	20	0,180	10	0,182	12	0,177	21
среднее	0,249	19,3	0,206	17,5	0,212	16,5	0,194	18	0,193	18,5	0,181	16,8	0,191	20,5	0,187	18,8	0,182	18	0,175	22
медиана	0,209	19,5	0,195	17,1	0,212	16,4	0,197	15	0,189	17,2	0,183	17,7	0,196	20,1	0,186	16,2	0,185	17,8	0,176	22,5

Таблица 2

Зависимость вероятности ошибки и размерности пространства анализа от размеров обучающей выборки, Фурье базис

№ user	γ=1		γ=2		γ=3		γ=4		γ=5		γ=6		γ=7		γ=8		γ=9		γ=10	
	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A	\tilde{d}	K_A
1	0,269	23	0,218	20	0,242	20	0,286	15	0,191	20	0,195	20	0,186	20	0,184	9	0,218	26	0,178	19
2	0,219	20	0,362	11	0,240	20	0,251	9	0,243	20	0,205	20	0,191	12	0,176	28	0,159	27	0,209	9
3	0,178	20	0,267	20	0,198	20	0,173	29	0,182	9	0,183	20	0,187	21	0,188	21	0,171	9	0,169	9
4	0,186	20	0,283	9	0,204	9	0,241	20	0,211	20	0,212	20	0,201	21	0,176	30	0,185	22	0,191	23
среднее	0,213	20,8	0,282	15	0,221	17,3	0,238	18,3	0,207	17,3	0,199	20	0,191	18,5	0,204	20	0,183	21	0,187	15
медиана	0,202	20	0,299	13,6	0,216	16,6	0,226	19,1	0,211	16,6	0,200	20	0,192	18,1	0,186	24,8	0,175	19,8	0,189	14

4. Благодарности

Работа выполнена при частичной финансовой поддержке гранта РФФИ № 18-29-03135-мк.

5. Литература

- [1] Borodinov, A. A Method of Preference and Utility Elicitation by Pairwise Comparisons and its Application to Intelligent Transportation Recommendation Systems / A. Borodinov, A. Agafonov, V. Myasnikov // 10th International Conference on Information Science and Technology, ICIST. – 2020. – P. 77-85.
- [2] Bradley, R.A. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons / R.A. Bradley, M.E. Terry // Biometrika. – 1952. – Vol. 39. – P. 324-345. DOI: 10.2307/2334029.
- [3] Fishburn, P.C. Utility theory for decision making. – Wiley, 1970.

- [4] Fürnkranz, J. Preference Learning / J. Fürnkranz, E. Hüllermeier. – Berlin Heidelberg: Springer-Verlag, 2011. DOI: 10.1007/978-3-642-14125-6.
- [5] Murphy, K.P. Machine Learning: A Probabilistic Perspective. – MIT Press, 2012.
- [6] Tsukida, K. How to analyze paired comparison data / K. Tsukida, M.R. Gupta, 2011.
- [7] Thurstone, L.L. A law of comparative judgment // Psychological Review. – 1927. – Vol. 34. – P. 273-286.
- [8] Campigotto, P. Personalized and Situation-Aware Multimodal Route Recommendations: The FAVOUR Algorithm / P. Campigotto // IEEE Transactions on Intelligent Transportation Systems. – 2017. – Vol. 18(1). – P. 92-102.
- [9] Arentze, T.A. Adaptive personalized travel information systems: A Bayesian method to learn users' personal preferences in multimodal transport networks / T.A. Arentze // IEEE Transactions on Intelligent Transportation Systems. – 2013. – Vol. 14(4). – P. 1957-1966.