

Д. М. Понизовкин, С. А. Амелькин

Математическая модель коллаборативных процессов принятия решений

Аннотация. В статье рассматривается математическая модель предпочтений пользователей коллаборативной рекомендательной системы. Вводится критерий оптимальности работы системы. Предложен алгоритм определения эффективности работы системы.

Ключевые слова и фразы: рекомендательные системы, коллаборативная фильтрация.

Введение

Увеличение ассортимента и переход к сверхнасыщенному рынку товаров [1], привели к необходимости организации навигации, позволяющей пользователю найти необходимый товар, наиболее полно удовлетворяющий потребностям пользователя. Особенно важна разработка систем навигации для товаров и услуг символического обмена [2]. Особенностью таких товаров является неопределенность априорной оценки их пользователем, поэтому системы навигации, формируемые продавцами товара, нацелены на формирование стандартных предпочтений (вкусов) [3], обеспечивающих выбор одних и тех же товаров, не обязательно низкого качества. Стандартные предпочтения («мейнстрим») позволяют продавцу снизить издержки за счет снижения ассортимента и увеличения оборачиваемости фондов. Распространение товаров символического обмена через интернет-магазины, особенно в виде электронного контента, дают возможность, наоборот, формировать разнообразие вкусов, не совпадающих с «мейнстримом». Такая политика торгующей организации приводит к расширению целевой аудитории магазина за пределы высококонкурентной области «мейнстрима». Задачи системы навигации в этом случае состоят в идентификации пользователя по принадлежности к той или

Работа проводилась при финансовой поддержке Министерства образования и науки Российской Федерации и при поддержке РФФИ.

ной вкусовой группе и предложению специфического для этой группы товара. Может также быть поставлена задача управления вкусом, если это соответствует целям магазина или общественному интересу, поддерживаемому магазином.

Для решения этих задач необходимо разработать математическую модель предпочтений пользователей, учитывающую возможность кластеризации пользователей в соответствии с вкусовыми группами.

1. Пространство предпочтений

Рассмотрим пользователя, который высказал свое мнение о n товарах путем их парного сравнения. Пусть для любых двух товаров t_i, t_j ($i \neq j, i = 1, \dots, n, j = 1, \dots, n$) определены результаты парного сравнения a_k . Например, $a_k \in A = \{<, >, \cong\}$, в этом случае $k = 1, 2, 3$. Тогда результаты парных сравнений для пользователя формируют матрицу R элементов r_{ij} , принимающих значения в соответствии с результатами парных сравнений a_k из множества $\{1, 2, \dots, k, \dots\}$. Никаких ограничений на транзитивность парных сравнений или на антисимметричность матрицы R не накладывается, так как эти свойства могут выполняться или нет в зависимости от метода опроса пользователя.

Значение матрицы R представляет собой точку в пространстве $A^{n \times n}$. Это пространство метрическое, на нем можно ввести расстояние между точками в соответствии с метрикой Хэмминга. Для этого для любых двух значений $a_k, a_l \in A$ определяются значения $\rho(k, l)$, характеризующие степень несовпадения результатов парного сравнения: $\rho(k, l) \in [0, 1]$, $\rho(k, l) = 0 \iff k = l$.

Величина расстояния между вкусами двух пользователей определяется как

$$\Delta = |R_1, R_2| = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \rho(r_{1ij}, r_{2ij}).$$

В случае, когда не все товары оценены пользователем, матрица R может содержать элементы $r_{ij} = 0$, соответствующие отношению «не сравнивались». При введении расстояния в такой ситуации следует

учесть, что большое количество нулей в матрицах R_1, R_2 не означает близости этих двух пользователей. Поэтому, вводя зависимость $\rho(k, l)$, надо для значений $r_{ij} = 0$ вводить дополнительный штраф ϵ :

$$\bar{\rho}(k, l) = \begin{cases} \rho(k, l), & k \neq 0 \text{ и } l \neq 0, \\ \rho(k, l) + \epsilon, & k = 0 \text{ или } l = 0. \end{cases}$$

2. Формирование кластеров

Рассматривая распределение реализаций парного сравнения на множестве $A^{n \times n}$ по данным реальной активности пользователей, легко отметить, что это распределение существенно отлично от равномерного. Причины неравномерности следующие:

- Значительная часть товаров не оценена; 90–95% матрицы R содержит нулевые значения.
- При формировании матрицы парных сравнений на основе выставленных оценок множество соответствующих страт не велико, поэтому значительное число элементов $r_{ij} = 0$, что соответствует результатам парных сравнений \cong , даже если можно выявить предпочтения внутри страты.

Эти причины неравномерности, однако, не описывают отклонения от постоянной плотности распределения точек R_ν . Введем значения расстояний $\Delta_{12} = |R_1, R_2|$ такие, что:

- если $\Delta_{12} \leq \delta^0$, то пользователи признаются идентичными;
- если $\Delta_{12} \leq \delta^*$, то пользователи признаются принадлежащими к одному и тому же кластеру (вкусовой группе).

Центром кластера назовем такую точку R^* , для которой

$$\frac{1}{N} \sum_{\nu=1}^N |R^*, R_\nu| \rightarrow \min_{R^*},$$

где N — число пользователей в кластере.

Наличие вкусовых групп, среди которых лидирующей является группа «мейнстрим», является также причиной неравномерности распределения матриц сравнения в пространстве $A^{n \times n}$.

Таким образом, кластер представляет собой шар радиуса δ^* в пространстве $A^{n \times n}$, который с точностью δ^0 покрывает область повышенной плотности распределения точек R_ν .

3. Расчет рекомендаций

Работа рекомендательной системы заключается в выделении таких объектов, которые не были оценены пользователем и ожидаемая оценка (результаты парных сравнений) для которых достаточно высока. Для этого выбирается строка матрицы R , состоящая из нулей, и рассчитываются значения ожидаемых результатов парных сравнений для объектов, оцененных пользователем. Рекомендация производится с учетом значений матриц парных оценок других пользователей, формирующих данную вкусовую группу. Стандартные алгоритмы коллаборативной фильтрации [4, 5] предлагают наличие множества пользователей, чьи предпочтения усредняются для восстановления значений матрицы R . При таком подходе не гарантируется сохранение положения пользователя в заданном кластере, а, наоборот, в ряде случаев рекомендации способствуют переходу пользователя в кластер «мейнстрим».

Поэтому предлагается разработка рекомендаций с учетом интересов навигатора (интернет-магазина) в виде решения задачи

$$|R^*, R| \rightarrow \min,$$

где управлениями являются значения r_{ij} в заданной пользователем i -ой строке. В случае произвольного выбора товара, следует выбрать такой i -ый товар, который обеспечит наибольшее приближение к центру кластера, соответствующего заданной вкусовой группе. Таким решением, очевидно, будет соответствующая строка значений матрицы R^* .

Эффективность работы рекомендательных систем оценивается по контрольным выборкам. Ожидается, что рекомендации, генерируемые системой, должны совпадать с реальными (но неизвестными системе) оценками пользователей. Среднее квадратическое и среднее абсолютное отклонения между рекомендациями и оценками из контрольных групп являются показателями эффективности. Такие показатели, однако, не учитывают знак ошибки рекомендации по направлениям к центру вкусовой группы пользователя и к центру «мейнстрима». Можно выделить ряд процессов, влияние которых не учитывается такими контрольными показателями эффективности. Прежде всего, это — влияние рекомендательной системы на формирование вкусов потребителя. Определение и учет влияния таких процессов — задача пока не решенная.

Алгоритмы, направленные на выделение кластеров пользователей, позволяют не только осуществлять рекомендательный прогноз, но и выделять целевые группы пользователей для товаров. Локализация кластеров с учетом влияния рекомендаций на вкусы пользователя в конечном счете повышает эффективность работы системы за счет уменьшения радиуса δ^* .

Список литературы

- [1] Сысоева С. В., Бузукова Е. А. Категорийный менеджмент. Курс управления ассортиментом в рознице. СПб : Питер, 2009. — 336 с. ↑[]
- [2] Долгин А. Б. Экономика символического обмена. М. : Инфра-М, 2006. — 632 с. ↑[]
- [3] Макаров А. Ю. *Системные программы обучения* // Деловое обозрение, 2004, № 4, с. 15–28 ↑[]
- [4] Manning C. D., Raghavan P., Schuetze H. Introduction to Information Retrieval. Cambridge : CUP, 2008. — 134 p. ↑3
- [5] Bishop C. M. Pattern Recognition and Mashine Learning. New York : Springer, 2007. — 256 p. ↑3

D. M. Ponizovkin, S. A. Amelkin. *A Mathematical Model of Collaborative Decision-Making Processes.*

ABSTRACT. Mathematical model of consumer's preferences for a collaborative recommendation service. A problem on optimal regime of the recommendation service is formulated. An algorithm of efficiency estimation is suggested.

Key Words and Phrases: recommendation service, collaborative filtering.

Образец ссылки на статью:

Д. М. Понизовкин, С. А. Амелькин. *Математическая модель коллаборативных процессов принятия решений* // Программные системы: теория и приложения : электрон. научн. журн. 2011. № 4(8), с. 95–99. URL: http://psta.psiras.ru/read/psta2011_4_95-99.pdf