

Е. Н. Степанов

Алгоритмическое и программное обеспечение для кластеризации данных в системах коллаборативной фильтрации

Научный руководитель: к.т.н. С. А. Амелькин

Аннотация. Разработка алгоритма для определения объективности оценок работ по результатам рецензирования.

Ключевые слова и фразы: Совместная фильтрация, Рецензирование, Конференция, Кластеризация.

Введение

Качество любой работы — это лишь субъективная оценка эксперта, который принимал данную работу. На экспертную оценку могут влиять различные факторы, даже незначительные. Кроме того, оценки — это величины нечисловой природы, т.к. инвариантны к любым монотонным преобразованиям. Поэтому вычисление функциональных зависимостей, аргументами которых являются эти оценки, неинформативно.

1. Постановка задачи

Требуется определить объективность оценки рецензентов.

2. Теоретическая часть

Имеется N рецензентов и N работ. Каждый рецензент оценивает каждую работу кроме своей. Из полученных оценок составляем матрицу размера $N \times N$.

Рец1	X	V_{12}	...	V_{1n}
Рец2	V_{21}	X	...	V_{2n}
...
РецN	V_{n1}	V_{n2}	...	X

Для каждого рецензента получается свой вектор оценок $V_{ij} = (V_{ij}^1, V_{ij}^2, \dots, V_{ij}^N)$. Кроме оценок рецензентов имеется вектор оценок, который считается эталонным — W_j . Эталонный вектор, в данном случае, это вектор правильных оценок. Данные оценки выставляются преподавателем. Поскольку алгоритм ещё не протестирован, нельзя использовать какие-то другие оценки.

Эталон	W_1	W_2	...	W_j
--------	-------	-------	-----	-------

Для каждого набора оценок вычисляем матрицу предпочтений A_{ij}^k . Поскольку рецензент не оценивает свою работу, на пересечении мы получаем пустые ячейки. Количество пустых ячеек равно N .

	V_{i1}	V_{i2}	V_{i3}	...	V_{in}
V_{i1}	X	знач	знач	...	знач
V_{j2}	знач	X	знач	...	знач
V_{j3}	знач	знач	X	...	знач
...
V_{jn}	знач	знач	знач	...	X

Правила расчета матрицы предпочтений:

Если $V_{i1} > V_{i2}$, то $V_{i1} = 1$.

Если $V_{i1} < V_{i2}$, то $V_{i1} = -1$.

Если $V_{i1} = V_{i2}$, то $V_{i1} = 0$.

Для вычисления степени близости оценок рецензента к эталонным оценкам, необходимо взять матрицу предпочтений рецензента (A_i) и эталонную матрицу предпочтений (A_e). В обеих матрицах вычеркиваем i -ый столбец и строку (рецензент не оценивает свою работу). Вводим такое понятие как близость — P . Теперь сравниваем матрицы поэлементно. Если $A_{ij}^k = A_{ej}^k$, то увеличиваем P на один. Конечное P — это количество совпадений в оценках рецензентов и эталонных оценках. Для того чтобы рассчитать степень близости оценок рецензента и эталонных оценок, необходимо рассчитать максимальное количество совпадений P_{\max} оценок рецензента и эталонных оценок. Рассчитываем его по формуле:

$$(1) \quad P_{\max} = \frac{(N-2)^2 - (N-2)}{2} = \frac{N^2 - 5N + 6}{2}$$

Минимальное количество совпадений P_{\min} всегда будет равно нулю. Введем новую переменную — R . R — это коэффициент близости оценок рецензента и эталонных оценок. R рассчитывается по формуле:

$$(2) \quad R = \frac{P}{P_{\max}}$$

Следовательно, значение R всегда находится в пределах от нуля до единицы. Чем ближе к единице R , тем сильнее близость оценок рецензента к эталонным оценкам.

3. Практическая часть

Данный алгоритм был разработан для информационной системы рейтинговой оценки успеваемости. Апробация алгоритма происходит на данных, полученных после проведения экзамена по Теории вероятностей и математической статистике во время зимней сессии 2009/10 учебного года.

Список литературы

- [1] Herlocker J.L., Konstan J.A., Riedl J. *Explaining Collaborative Filtering Recommendations*, 2000, p.10, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.26.5594>.
- [2] O'Connor M., Herlocker J.L. *Clustering Items for Collaborative Filtering*, 2001, p. 4, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.6325>.
- [3] Marlin B. *Modeling User Rating Profiles For Collaborative Filtering*, 2003, p. 8, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.1932>.

УГП, 5И52

Е. N. Stepanov. *Algorithmic maintenance and software for data clusterization in collaborate filtering systems* // Proceedings of Junior research and development conference of Ailamazyan Pereslavl university. — Pereslavl, 2010. — p. 215–217. (*in Russian*).

ABSTRACT. Creation of algorithm for definition of objectivity assessments of works by results of estimations of reviewers.

Key Words and Phrases: Collaborative filtering, Cluster, Reviewing, Conference.