УДК 519.87:004.89

МОДЕЛЬ С ЛАТЕНТНЫМИ ПАРАМЕТРАМИ ДЛЯ АНАЛИЗА РЕЙТИНГОВЫХ ОЦЕНОК

Братищенко В.В.

ФГБОУ ВО «Байкальский государственный университет», Иркутск,
e-mail: info@bgu.ru

Рейтинговые оценки являются распространенным способом исследования в самых разных областях. Применяются рейтинговые оценки для построения рейтингов объектов и настройки советующих систем на основе моделей коллаборативной фильтрации. Построение рейтингов выполняется усреднением оценок. Эта процедура не учитывает индивидуальные склонности субъектов завышать или занижать оценки. Для устранения этого недостатка предлагается алгоритм пересчета оценки в шкалу с усредненными параметрами. Пересчет выполняется на основе сравнения распределения вероятностей рейтинговой оценки с усредненным распределением. Для решения этой задачи предложена стохастическая модель рейтинговой оценки с латентными параметрами, характеризующими оценщика и шкалу объекта оценивания. Предложенная модель является аналогом Partial Credit Model, применяемой в Item Response Theory для обработки результатов тестирования. Разработаны процедуры оценивания латентных параметров методом максимального правдоподобия и проверки адекватности модели на основе дисперсионного анализа. Для пересчета оценок предложен алгоритм, использующий функции распределения вероятностей для шкалы оценки и усредненной шкалы. Приведены результаты обработки большого массива рейтинговых оценок. Вычисления подтверждают адекватность модели рейтинговых оценок и процедуры пересчета оценок. Приведенная методика может быть использована для получения более точных рейтингов.

Ключевые слова: рейтинговые оценки, коллаборативная фильтрация, модели с латентными параметрами, сравнение рейтинговых шкал, Partial Credit Model, Item Response Theory

MODEL WITH LATENT PARAMETERS FOR ANALYSIS OF RATINGS

Bratischenko V.V.

Baikal State University, Irkutsk, e-mail: info@bgu.ru

Rating are a common way of research in a wide variety of fields. Ratings are used to build items ratings and set up advisory systems based on collaborative filtering models. Ratings are built by averaging the rating marks. This procedure does not take into account the individual propensities of users to overestimate or underestimate. To eliminate this shortcoming, an algorithm for recalculating the mark scale into a scale with averaged parameters is proposed. The recalculation is performed on the basis of a comparison of the probability distribution of the mark score with the average distribution. To solve this problem, a stochastic rating model with latent parameters characterizing the user and the mark scale is proposed. The proposed model is analogous to the Partial Credit Model used in Item Response Theory to process test results. Procedures for estimating latent parameters by the maximum likelihood method and testing the adequacy of the model based on analysis of variance have been developed. To recalculate the estimates, an algorithm is proposed that uses the probability distribution functions for the mark scale and the average scale. The results of processing a large array of ratings are presented. The calculations confirm the adequacy of the ratings model and the ratings recalculation procedure. The above methodology can be used to obtain more accurate ratings.

Keywords: rating scores, collaborative filtering, models with latent parameters, comparison of rating scales, Partial Credit Model, Item Response Theory

Многие сайты собирают рейтинговые оценки фильмов, книг, музыкальных произведений, товаров и многих других объектов. Каждая рейтинговая оценка принадлежит некоторой порядковой шкале, например $\{0, 1, 2, 3, 4, 5\}$. Обобщение рейтинговых оценок часто представлено в виде усредненных значений. Усреднение является не вполне корректной процедурой, так как порядковые шкалы не являются метрическими. Кроме этого, оценки выполняются разными людьми, которые интерпретируют одну и ту же шкалу по-разному. Обработка таких оценок с учетом подобных различий представляется актуальной задачей.

Для обработки рейтинговых оценок применяют методы коллаборативной фильтрации [1]. Рейтинговая оценка $f_{ur} \in S = \{0,1,\ldots,k\}$ выставляется субъектом $u \in U$ (клиенты, пользователи, …) объекту $r \in R$ (товары, фильмы, ресурсы,…). Матрица $\| f_{ur} \|$ является сильно разреженной. Например, один зритель из нескольких тысяч оценивает десятки фильмов из большого их количества. Традиционное применение коллаборативной фильтрации заключается в прогнозировании не заполненных значений f_{ur}, определении степени сходства субъектов и объектов.

Рейтинговые оценки используются для ранжирования объектов. На эту процедуру существенно влияют индивидуальные склонности субъектов к переоценке или недооценке объектов. Возможны ситуации искажения результатов ранжирования в случае существенно неоднородных множеств оценщиков. Если большая часть оценщиков некоторого объекта склонны к завышению оценок, то и средняя оценка будет завышенной. Субъекты, выставляющие в основном высокие или в основном низкие оценки, будут значительно искажать результаты ранжирования. Таким образом, разные субъекты применяют разные шкалы, даже если они содержат одинаковое количество градаций. Для корректной статистической обработки таких данных необходимо установить соответствие разных шкал.

Материалы и методы исследования

Традиционно сопоставление шкал связывают с сопоставлением интервалов [2] интегральных характеристик, по которым определяются рейтинги. Кроме этого, соответствие шкал можно определять на основе статистических распределений оценок. Для данных, изучаемых в данной работе, такой подход оправдан статистической однородностью рассматриваемых наблюдений. Для его реализации нужно строить распределения оценок, зависящие как от параметров субъекта, так и от параметров объекта. Подобные распределения применяет Item Response Theory [3] для описания итогов тестирования с учетом латентных параметров, характеризующих трудности тестовых заданий и подготовленность тестируемых.

Среди методов коллаборативной фильтрации отмечается привлекательность моделей [4, 5], которые используют латентные параметры, характеризующие каждого субъекта и каждый объект. В результате существенно снижаются количество оцениваемых параметров и сложность алгоритмов, а также повышается точность моделирования. Подобный подход применяется [4] для решения традиционных задач коллаборативной фильтрации. Отмечается [5] близость моделей Item Response Theory, применяемых для обработки данных тестирования, и латентных моделей коллаборативной фильтрации.

В данной работе предлагается инструмент построения рейтинга объектов, который учитывает индивидуальные особенности применения порядковой шкалы каждым субъектом на основе моделей с латентными параметрами.

Предлагается следующая модель, описывающая получение оценки в виде последовательности шагов. В случае «неудачи» на l-м шаге ($l = 1,\ldots,k$) с вероятностью $1 - p_i$ оценка принимается равной $l - 1$, в случае «успеха» с вероятностью p_i испытание продолжается на следующем шаге. Предполагается независимость шагов. Успешно выполнив все шаги, субъект выставляет максимальный балл k. Случайная величина X – итоговая оценка – будет иметь следующее распределение вероятностей:

$$P\{X = x\} = (1 - p_{x+1})\prod_{l=0}^{x} p_l, x = 0,\ldots,k,$$

где для удобства полагается $p_0 = 1$, $p_{k+1} = 0$.

Математическое ожидание X можно вычислить в виде следующей суммы:

$$M[X] = \sum_{h=1}^{k}\prod_{l=1}^{h} p_l.$$

По аналогии с моделью «Partial Credit» [3], применяемой для анализа результатов тестирования, склонность субъекта u выставлять повышенные оценки будем измерять параметром θ_u, а l-й шаг оценивания объекта r будем характеризовать параметром δ_{rl}. Параметры определяют для субъекта u вероятности шагов оценивания объекта r:

$$p_l(u,r) = \frac{e^{\theta_u}}{e^{\theta_u} + e^{\delta_{rl}}}, l = 1,\ldots,k$$

и вероятность оценки:

$$P\{F_{ur} = f_{ur}\} = \left(1 - p_{f_{ur}+1}(u,r)\right)\prod_{l=0}^{f_{ur}} p_l(u,r)$$

Очевидно, что, чем больше θ_u, тем сильнее распределение оценки F_{ur} будет смещаться в сторону более высоких оценок. Параметры $\delta_{r1},\ldots, \delta_{rk}$ определяют распределение оценки объекта r при некотором фиксированном θ и могут рассматриваться как характеристики объекта. Предполагается независимость всех шагов оценивания разными субъектами.

Для оценки параметров можно применить метод максимального правдоподобия – найти значения параметров, обеспечивающих максимум вероятности:

$$L = \prod_{i=1}^{N} P\left\{F_{u(i)r(i)} = f_{u(i)r(i)}\right\}$$

наблюдений $\left\{f_{u(i)r(i)}, i = 1,\ldots,N\right\}$, где $u(i)$ – субъект i-й оценки, $r(i)$ – объект i-й оценки, N – число наблюдений. Перейдем от функции максимального правдоподобия:

$$L=\prod_{i=1}^{N}\begin{cases}\prod_{l=1}^{k}\frac{e^{\theta_{u(i)}}}{e^{\theta_{u(i)}}+e^{\delta_{r(i)l}}}, f_{u(i)r(i)}=k\\ \frac{e^{\delta_{r(i)f_{u(i)r(i)}+1}}}{e^{\theta_{u(i)}}+e^{\delta_{r(i)f_{u(i)r(i)}+1}}}\prod_{l=1}^{f_{u(i)r(i)}}\frac{e^{\theta_{u(i)}}}{e^{\theta_{u(i)}}+e^{\delta_{r(i)l}}}, 0<f_{u(i)r(i)}<k\\ \frac{e^{\delta_{r(i)1}}}{e^{\theta_{u(i)}}+e^{\delta_{r(i)1}}}, f_{u(i)r(i)}=0\end{cases}$$

к ее логарифму:

$$ln(L)=\sum_{i=1}^{N}\begin{cases}k\theta_{u(i)}-\sum_{l=1}^{k}\ln\left(e^{\theta_{u(i)}}+e^{\delta_{r(i)l}}\right), f_{u(i)r(i)}=k;\\ f_{u(i)r(i)}\theta_{u(i)}+\delta_{r(i)f_{u(i)r(i)}+1}-\\ -\sum_{l=1}^{f_{u(i)r(i)}+1}\ln\left(e^{\theta_{u(i)}}+e^{\delta_{r(i)l}}\right), 0\le f_{u(i)r(i)}<k;\end{cases}$$

Условием достижения максимума будет равенство нулю частных производных:

$$\frac{\partial ln(L)}{\partial\theta_u}=\sum_{\substack{i=1:\\u(i)=u}}^{N}\begin{cases}f_{ur(i)}-\sum_{l=1}^{k}\frac{e^{\theta_u}}{e^{\theta_u}+e^{\delta_{r(i)l}}}, f_{ur(i)}=k;\\ f_{ur(i)}-\sum_{l=1}^{f_{ur(i)}+1}\frac{e^{\theta_u}}{e^{\theta_u}+e^{\delta_{r(i)l}}}, 0\le f_{ur(i)}<k;\end{cases} \quad (1)$$

$$\frac{\partial ln(L)}{\partial\delta_{rl}}=\sum_{\substack{i=1:\\r(i)=r,f_{u(i)r}\ge l-1}}^{N}\frac{e^{\theta_{u(i)}}}{e^{\theta_{u(i)}}+e^{\delta_{rl}}}-\sum_{\substack{i=1:\\r(i)=r,f_{u(i)r}\ge l}}^{N}1. \quad (2)$$

Для решения таких уравнений в IRT предлагается использовать метод Ньютона. В этом методе в итерационной формуле:

$$x^{(s+1)}=x^{(s)}-\frac{f\left(x^{(k)}\right)}{f'\left(x^{(k)}\right)}$$

в числителе используется первая производная логарифма функции правдоподобия, а в знаменателе – вторая:

$$\frac{\partial^2 ln(L)}{\partial{\theta_u}^2}=\sum_{\substack{i=1:\\u(i)=u}}^{N}\begin{cases}-\sum_{l=1}^{k}\frac{e^{\theta_u}e^{\delta_{r(i)l}}}{\left(e^{\theta_u}+e^{\delta_{r(i)l}}\right)^2}, f_{ur(i)}=k;\\ -\sum_{l=1}^{f_{ur(i)}+1}\frac{e^{\theta_u}e^{\delta_{r(i)l}}}{\left(e^{\theta_u}+e^{\delta_{r(i)l}}\right)^2}, 0\le f_{ur(i)}<k;\end{cases}$$

$$\frac{\partial^2 ln(L)}{\partial{\delta_{rl}}^2}=-\sum_{\substack{i=1:\\r(i)=r,f_{u(i)r}\ge l-1}}^{N}\frac{e^{\theta_{u(i)}}e^{\delta_{rl}}}{\left(e^{\theta_{u(i)}}+e^{\delta_{rl}}\right)^2}$$

Отрицательные значения второй производной соответствуют вогнутости функции, что гарантирует сходимость предложенной процедуры поиска значений параметров.

Особенностью предложенной модели является то, что значение вероятностей не изменится, если все латентные параметры изменить на одну и ту же величину.

Для получения начальных значений положим в (1) $e^{\delta_{r(i)l}} = 1$. В этом случае параметр вычисляется как логарифм средней оценки субъекта *u*:

$$\theta_u = ln\left(\frac{1}{m(u)} \sum_{\substack{i=1: \\ u(i)=u}}^{N} f_{ur(i)}\right),$$

где $m(u) = \sum_{\substack{i=1: \\ u(i)=u}}^{N} 1$ – количество оценок субъекта *u*.

Аналогично полагая в (2) $e^{\theta_{u(i)}} = 1$, получаем:

$$\delta_{rl} = ln\left(\frac{n(r,l-1) - n(r,l)}{n(r,l)}\right),$$

где $n(r,l) = \sum_{\substack{i=1: \\ r(i)=r, f_{u(i)r} \geq l}}^{N} 1$ – количество оценок субъекта *r*, не меньших *l*.

Если аргумент логарифма равен нулю, то начальное значение параметра приравнивается некоторому отрицательному значению, например $\delta_{rl} = -5$, что соответствует почти единичной вероятности успеха на *l*-м шаге оценивания.

Для апробирования модели был выбран набор данных ml-latest-small с рейтинговыми оценками фильмов, который содержит 100 836 оценок для 9 308 фильмов от 610 зрителей. Данный набор данных размещен на сайте Social Computing Research at the University of Minnesota. Оценки включали половинки баллов от 0,5 до 5 и были переведены в целочисленную шкалу от 0 до 9. Для данного набора данных были определены латентные параметры субъектов (зрителей) и объектов (фильмов).

Для оценки адекватности модели предлагается применить дисперсионный анализ. Дисперсионный анализ позволяет оценить зависимость оценок от субъектов и объектов. Для оценки влияния субъектов сравниваются усредненные выборочные дисперсии по субъектам:

$$M_2 = \frac{1}{N-n} \sum \left(f_{ur} - \widehat{f_{u*}}\right)^2, \quad u$$

$$\widehat{f_{u*}} = \frac{1}{m(u)} \sum_{\substack{k=1: \\ u(k)=u}}^{N} f_{u(k)r(k)},$$

где *n* – количество субъектов,

с межгрупповой дисперсией:

$$M_1 = \frac{1}{n-1} \sum \left(\widehat{f_{u*}} - \widehat{f_{**}}\right)^2, \widehat{f_{**}} = \frac{1}{N} \sum f_{ur}.$$

Статистики M_1 и M_2 являются разными оценками дисперсии одной и той же случайной величины в случае отсутствия влияния субъекта на оценку. Статистика $F = M_1 / M_2$, при условии одинакового нормального распределения и независимости вариаций среди оценок, будет иметь распределение Фишера со степенями свободы $n - 1$ и $N - n$. Как и следовало ожидать, проверка подтвердила предположение о наличии влияния субъектов на оценки. Аналогично была подтверждена зависимость оценок от объектов.

По такой же схеме были проверены гипотезы о независимости остатков $f'_{ur} = f_{ur} - M[F_{ur}]$ от субъектов и объектов. Вычисления подтвердили эти гипотезы с доверительной вероятностью, близкой к единице.

Коэффициент детерминации:

$$R^2 = 1 - \sum \left(f_{ur} - M[F_{ur}]\right)^2 / \sum \left(f_{ur} - \widehat{f_{**}}\right)^2$$

определяет «долю изменчивости», которую описывает модель. Для предложенной модели и массива оценок коэффициент детерминации составил 0,43. Не слишком большое значение коэффициента детерминации показывает, что оценки являются в большой степени случайными, а также в модели не учитывается множество других факторов, например характеристик фильмов и зрителей, оказывающих значимое влияние на формирование оценок. В целом предложенная модель позволяет получить более точные оценки по сравнению с простым усреднением.

Для каждого субъекта рейтинговая оценка будет иметь свое распределение вероятностей. Усреднение таких оценок может приводить к значительным искажениям. Для исключения этого нужен инструмент определения соответствия значений случайных величин с разными распределениями. Для непрерывных величин *X* и *Y* можно использовать функцию:

$$y = F_Y^{-1}\left(F_X(x)\right)$$

где *F* – функция распределения соответствующей случайной величины. Для дискретных случайных величин такой подход не является точным. Идея следующего метода связана с сопоставлением значений двух вариационных рядов с одинаковыми номерами. Для дискретных случайных величин в вариационном ряду одинаковые значения образуют интервал. Сопоставление компонентов вариационных рядов приводит к ситуации, когда интервалу одинаковых значений x будет соответствовать часть вариационного ряда для Y, возможно, с неодинаковыми значениями y. Если вместо вариационных рядов использовать вероятности, то получается следующая схема. Значению x соответствует (рис. 1) интервал $(F_X(x), F_X(x+1)]$. Этому интервалу могут соответствовать несколько значений Y:

$$y_1, y_1+1, \ldots, y_2$$

$$y_1 : F_Y(y_1-1) < F_X(x) \le F_Y(y_1),$$

$$y_2 : F_Y(y_2) < F_X(x+1) \le F_Y(y_2+1).$$

Значению x можно поставить в соответствие среднее значение y из интервала, соответствующего интервалу $(F_X(x), F_X(x+1))$. Определим случайную величину Z на основе распределения Y:

$$P\{Z = y_1\} = \frac{F_Y(y_1+1) - F_X(x)}{P\{X = x\}}$$

$$P\{Z = y_2\} = \frac{F_X(x+1) - F_Y(y_2)}{P\{X = x\}}$$

$$P\{Z = z\} = \frac{P\{Y = z\}}{P\{X = x\}}, y_1 < z < y_2$$

$$P\{Z = z\} = 0, \; z < y_1 \text{ или } z > y_2$$

Значению x ставится в соответствие $M[Z]$.

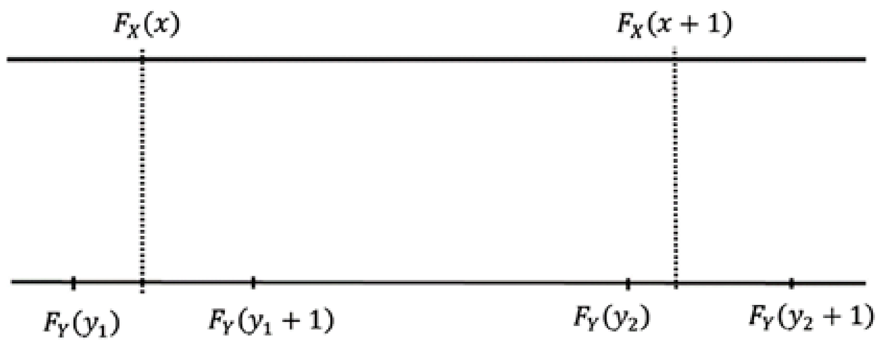

Рис. 1. Соответствие двух шкал на основе сопоставления функций распределения $F_X(x)$ и $F_Y(y)$

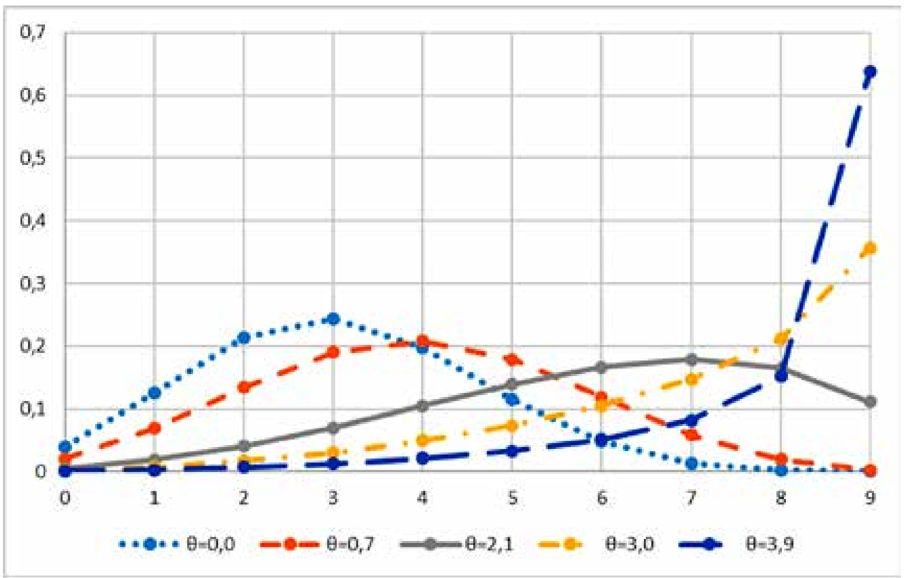

Рис. 2. Распределения вероятностей оценок с одинаковыми латентными параметрами объектов и разными параметрами θ_u субъектов

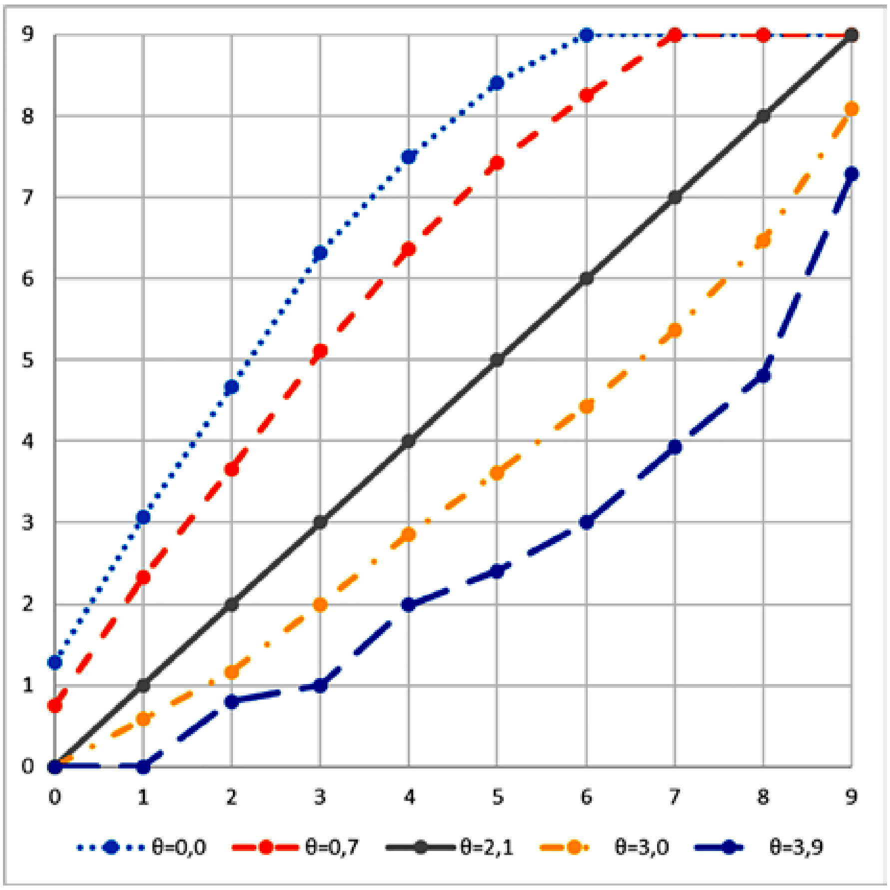

Рис. 3. Преобразование оценок с распределениями для разных параметров $\theta_и$ к распределению $\theta_и = 2,1$

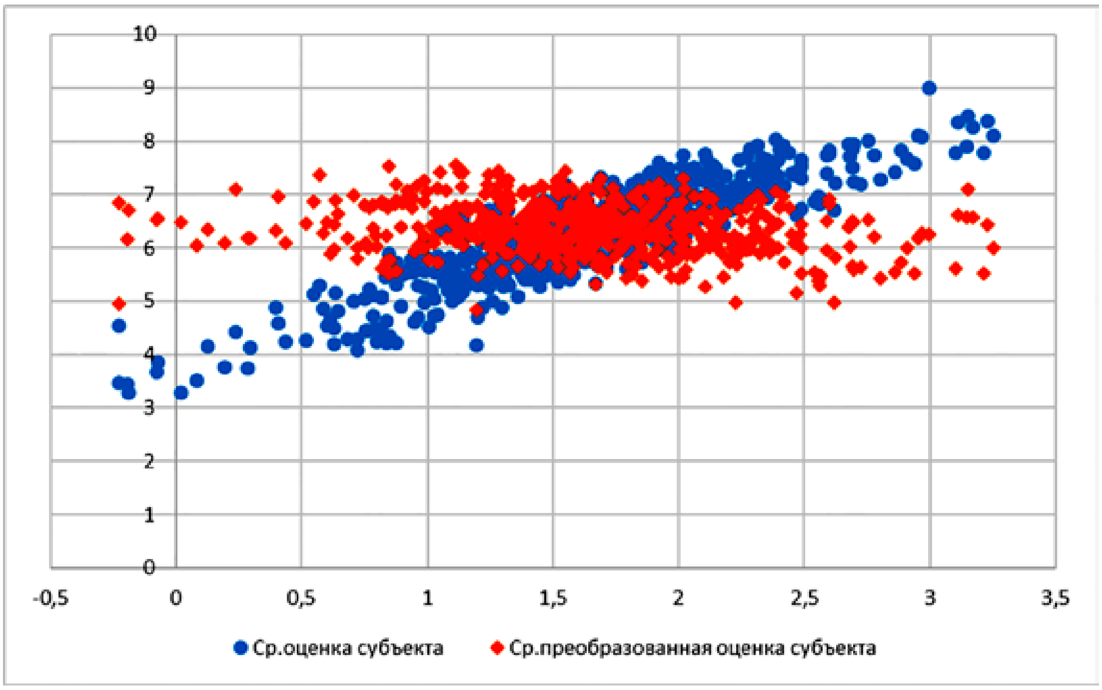

Рис. 4. Разброс усредненных исходных и преобразованных оценок субъектов в зависимости от значений латентного параметра субъекта

Преобразование оценки на основе разности распределений следует проводить для исключения тенденции субъекта ставить завышенные или заниженные оценки. Такая тенденция субъекта определяется его латентным параметром: чем больше параметр θ_u субъекта, тем больше вероятность высокой оценки (рис. 2).

На рисунке 3 приведены преобразования оценок с разными распределениями, ось ординат соответствует исходным оценкам, ось абсцисс – преобразованным. Как было естественно ожидать, преобразование увеличивает оценки субъектов с низким значением латентных параметров – склонных занижать оценки – и уменьшает оценки субъектов с большими значениями латентных параметров – склонных завышать оценки. Пилообразность преобразования для $\theta_u = 3{,}9$ объясняется неравномерным перекрытием интервалов, соответствующих одинаковым значениям случайных величин с разными распределениями вероятностей.

На рисунке 4 представлена диаграмма разброса исходных и преобразованных оценок в зависимости от значений латентного параметра субъекта. Диаграмма демонстрирует, что преобразование компенсирует смещение оценок под влиянием тенденции субъекта завышать или занижать оценки.

Выводы

Предложенная модель рейтинговых оценок позволила построить вероятностные распределения рейтинговых оценок в зависимости от свойств субъектов и объектов, задаваемых латентными параметрами. Результаты обработки большого множества наблюдений подтверждают адекватность модели. Полученные распределения позволили определить преобразование оценок на основании распределения вероятностей рейтинговой оценки и усредненного распределения. Такое преобразование позволяет исключить субъективность оценок. В предложенную модель можно включить дополнительные параметры субъектов и объектов оценивания для учета факторов, которые могут повысить точность моделирования. Применение предложенной методики пересчета оценок позволяет получить более объективные оценки и более точные рейтинги объектов.

Список литературы

1. Dong X., Yu L., Wu Z., Sun Y., Yuan L., Zhang F. A Hybrid Collaborative Filtering Model with Deep Structure for Recommender Systems. Proceedings of the AAAI Conference on Artificial Intelligence. 2017. V. 31(1). DOI: 10.1609/aaai.v31i1.10747.

2. Айвазян С.А., Головань С.В., Карминский А.М., Пересецкий А.А. О подходах к сопоставлению рейтинговых шкал // Прикладная эконометрика. 2011. № 3 (23). С. 13-40.

3. Hamel J.-F., Sébille V., Challet-Bouju G., Hardouin J.-B. Partial Credit Model: Estimations and Tests of Fit with Pcmodel. The Stata Journal. 2016. V. 16(2). P. 464–481. DOI: 10.1177/1536867X1601600212.

4. Kumar B. A novel latent factor model for recommender system. JISTEM. Journal of Information Systems and Technology Management. 2016. V. 13. Is. 3. P. 497-514. DOI: 10.4301/S1807-17752016000300008.

5. Bergner Y., Halpin P., Vie J.J. Multidimensional Item Response Theory in the Style of Collaborative Filtering. Psychometrika. 2022. V. 87(1). P. 266-288. DOI: 10.1007/s11336-021-09788-9.