Модель рекомендательной системы на нечетких множествах как эффективное расширение коллаборативной модели

© Д.М. Понизовкин

IT-Aces,
г. Переславль-Залесский, Россия

denis.ponizovkin@gmail.com

Аннотация. Рассмотрены рекомендательные системы, использующие коллаборативную фильтрацию для решения таких задач, как определение степени близости объекта пользователю (задача прогнозирования) и определение подмножества объектов мощности N, близких пользователю (задача $topN$). Такие системы считаются хорошо изученными и успешно применяются в коммерции, однако существуют открытые проблемы, связанные с использованием таких систем. Эти проблемы описаны в настоящей работе. В качестве метода устранения существующих недостатков предложена модель рекомендательных систем, которая основана на теории нечетких множеств и использует методы коллаборативной фильтрации.

Ключевые слова: рекомендательная система, коллаборативная фильтрация, мера сходства, отношение близости, эффективность, нечеткая логика.

The Model of Recommender Systems based on Fuzzy Logic as the Extension of the Collaborative Filtering Model

© Denis M. Ponizovkin

IT-Aces, Pereslavl-Zalessky, Russia

denis.ponizovkin@gmail.com

Abstract. In this article, we analyze collaborative filtering. We show existing problems connected with using of collaborative filtering. We propose the recommender system model based on the fuzzy logic theory. This model is the extension of the collaborative filtering which removes described problems.

Keywords: recommender system, collaborative filtering, similarity measure, fuzzy logic.

1 Терминология и обозначения

Рекомендательные системы (далее РС) [1] – одна из развивающихся областей Computer Science, начавшая свое существование с конца прошлого столетия [2]. РС являются инструментом, который облегчает пользователю задачу поиска нужной информации путем предоставления рекомендации по использованию соответствующей информации или за счет определения степени близости конкретной информации интересам пользователя.

РС работают со следующими исходными данными:

- $u \in U \subset \mathbb{N}$ – идентификаторы пользователей РС;
- $i \in I \subset \mathbb{N}$ – идентификаторы объектов предметной области РС, например, фильм в РС в области кинематографии; для простоты изложения не будем каждый раз употреблять выражения «пользователь» или «объект», будем обозначать их кратко «объекты»;
- $\rho: U \times I \to [0,1]$ – функция оценки близости и объектов; значение $\rho(u,i)$ показывает, насколько объекты i и u близки; как правило, оценки близости задаются самими пользователями за время работы с РС; будем считать, что чем меньше значение оценки, тем объекты ближе; будем говорить, что между пользователем u и объектом i выполняется отношение близости $\mathcal{R}$, если $\rho(u,i) \le \varepsilon_0 \in \varepsilon(0)$; будем называть такие объекты близкими.

Как правило, РС решают следующие две задачи (пользователь, для которого производится решение, называется *активным* и обозначается символом u_a):

Труды XIX Международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных» (DAMDID/ RCDL'2017), Москва, Россия, 10–13 октября 2017 года

1. *Задача прогнозирования*: спрогнозировать неизвестное значение $\rho(u_a, i_p) = \perp$ (символом $\perp$ будем обозначать неизвестное значение) путем алгоритмического вычисления значения прогнозной функции $\overline{\rho}(u_a, i_p): U \times I \to [0,1]$, где i_p – прогнозируемый объект; при этом требуется, чтобы прогноз был составлен точно, то есть $|\overline{\rho}(u_a, i_p) - \rho(u_a, i_p)| \leq \varepsilon_0$;

2. Задача $topN$ – формирование подмножества объектов

$$I_{topN} = \{i: (u_a \mathcal{R} i) \wedge \rho(u_a, i) = \perp\} \wedge |I_{topN}| = N.$$

Так как неизвестно, выполняется ли отношение $u_a \mathcal{R} i$ в силу того, что $\rho(u_a, i) = \perp$, то выполнение отношения $u_a \mathcal{R} i$ определяется по значению прогнозной функции: $u_a \mathcal{R} i \Leftrightarrow \overline{\rho}(u_a, i) \leq \varepsilon_0$. Решение названных задач производится РС за счет анализа информации о характеристиках пользователей и объектов. X – множество характеристик пользователей, например, социально-демографические показатели банковской РС. Y – множество характеристик объектов, например, наименования кинематографических жанров. Значением характеристик пользователей является значение весовой функции $w_U: U \times X \to [0,1]$, объектов – $w_I: I \times Y \to [0,1]$. Значения весов могут задаваться пользователями, экспертами, алгоритмически и т. д. Структуру данных, представляющую информацию о пользователе u и объекте i назовем контентом пользователя $c_X(u)$ и контентом объекта $c_Y(i)$ соответственно.

Модель РС – это тройка

$$(c_X; c_Y; \Pi), \quad (1)$$

где П – правило алгоритмического вычисления значения прогнозной функции $\overline{\rho}$.

Чтобы определить качество решения задачи, проводится тестирование, для которого исходное множество данных P разбивается на обучающее и тестовое множества P_0 и $P_\perp$ соответственно. Если $\rho(u, i) \in P_0$, будем обозначать такие объекты i_0. Если $\rho(u, i) \in P_\perp$, будем обозначать такие объекты $i_\perp$.

2 Коллаборативные модели

Рассмотрим коллаборативную фильтрацию [3-7], которая является одним из наиболее изученных [3], популярных [4] и успешных [5] правил вычисления П. РС, которые используют коллаборативную фильтрацию в качестве правила П, будем называть коллаборативными РС (далее КРС). Они делятся на два типа по фильтруемому множеству [6]: множеству пользователей или объектов. Будем называть первые субъектно-ориентированными (далее СОК), а последние – объектно-ориентированными (далее ОРС) [7].

Опишем теорию, на которой основаны коллаборативные П. Решение строится по обучающему множеству, а его качество определяется по тестовому. Обучающее множество выступает в роли информации прошедшего времени, тестовое – в роли информации будущего времени.

Правило П СОК основано на утверждении, которое гласит, что если в прошлом пользователи были близки по предпочтениям, то и в будущем они будут близки по предпочтениям. Во введенной терминологии данное утверждение примет следующий вид:

$$u_a \mathcal{R}_u u \text{ выполняется на } P_0 \Rightarrow u_a \mathcal{R}_u u \text{ выполняется на } P_\perp, \quad (2)$$

$\mathcal{R}_u$ – отношение близости пользователей. Выполнение отношения близости $\mathcal{R}_u$ между пользователями устанавливается СРС на основании значений характеристик пользователей. Характеристиками для СОК всегда выступают объекты, а значениями весов – значения $\rho(u, i) \in P_0$, которые были выставлены самими пользователями и характеризуют предпочтения пользователей. Для определения близости по предпочтениям используются так называемые меры близости

$$\delta_u: U \times U \to [0,1]: (1 - \delta_u(u, v)) \leq \varepsilon_0 \Leftrightarrow u \mathcal{R}_u v.$$

Пользователи, между которыми выполняется отношение близости, называются *соседями.*

Правило П СОК задается формулой

$$u \in U, (u_a \mathcal{R} u) \Rightarrow \left|\overline{\rho}(u_a, i_p) - \rho(u_a, i_p)\right| \leq \varepsilon_0, \quad (3)$$

$\overline{\rho}(u_a, i_p) = f(\{\rho(u, i_p)\})$. Правило П СОК говорит о том, что если пользователи u являются соседями для пользователя u_a, то оценки $\rho(u_a, i_p)$, $\rho(u, i_p)$ коррелируют, поэтому неизвестное значение $\rho(u_a, i_p)$ можно функционально определить по значениям $\{\rho(u, i_p)\}$, то есть прогнозная функция является функцией от значений оценок близости соседей.

Правило П ООК основано на утверждении: если пользователю нравится объект i, который близок по характеристикам к объекту j, то пользователю понравится объект j. Во введенной терминологии данное утверждение примет вид

$$(u_a \mathcal{R} i) \wedge (i \mathcal{R}_i j) \Rightarrow u_a \mathcal{R} j, \quad (4)$$

$\mathcal{R}_i$ – отношение близости объектов. Отношение близости $\mathcal{R}_i$ между объектами устанавливается РС на основании значений мер близости: $1 - \delta_i(i, j) \leq \varepsilon_0 \Leftrightarrow i \mathcal{R}_i j$, $\delta_i: I \times I \to [0,1]$ – мера близости объектов. Объекты, между которыми выполняется отношение близости, называются соседями.

При решения задачи $topN$ в ООК используется информация только о тех объектах, для которых известно, что $(u_a \mathcal{R} i_0), (u_a \mathcal{R} i_\perp)$, поэтому будем считать, что $P = \{\rho(u, i): u \mathcal{R} i\}$ для задачи $topN$.

Правило П ООК задается формулой

$$(i \mathcal{R}_i i_0) \Rightarrow (\overline{\rho}(u_a, i) = 0) \Rightarrow u_a \mathcal{R} i. \quad (5)$$

Значения $\overline{\rho}(u_a, i)$ задаются равными нулю, потому

что тогда объекты i будут близки активному пользователю при любом пороговом значении ε_0.

Правило вывода ООК говорит о том, что если существует объект i, являющийся соседом объекта i_0, то, следуя эвристическому утверждению, $u_a \mathcal{R} i$, так как $u_a \mathcal{R} i_0$ по принятому для задачи $topN$ виду исходного множества.

3 Проблемы применения коллаборативных моделей

3.1 Выполнение эвристических утверждений

Будем говорить, что РС *эффективна*, если ее правила вывода удовлетворяют некоторому критерию независимо от дополнительных ограничений или условий.

Реальные исходные данные обладают свойствами *динамики* и *неоднородности* [8]. Свойство динамики заключается в том, что множество исходных данных изменяется во времени, так как изменяются предпочтения пользователей, и мощность множеств U, I растет. Пусть выполняется $u_a \mathcal{R}_u u$ для P_0, но в силу динамики возможна ситуация, когда $1 - \delta_u(u_a, i) > \varepsilon_0$ для $P_\perp$. Тогда утверждение СОК (2) и, следовательно, правило П СОК (3) ложны в общем случае для любых исходных данных.

Свойство неоднородности заключается в том, что пользователи предпочитают различные объекты, не обязательно близкие по характеристикам, то есть их вкусы неоднородны: $(u_a \mathcal{R} i) \wedge (u_a \mathcal{R} j) \not\Rightarrow (i \mathcal{R}_i j)$. Тогда $(u_a \mathcal{R} i) \wedge (i \mathcal{R}_i j) \not\Rightarrow u_a \mathcal{R} j$, то есть утверждение ООК (4) и, следовательно, правило вывода ООК (5) ложны в общем случае для любых исходных данных. Таким образом, КРС не являются эффективными по критерию качества решения, так как оно зависит от выполнения эвристических утверждений, что, в свою очередь, зависит от свойств исходных данных.

Таким образом, КРС не являются эффективными по критерию качества решения, так как оно зависит от выполнения эвристических утверждений, что, в свою очередь, зависит от свойств исходных данных.

3.2 Достаточные условия качественного решения

Отношение близости обладает следующими свойствами: рефлексивность, симметричность, транзитивность. Выполнение свойства *транзитивности* отношения близости зависит от выбора функции, используемой в качестве меры близости, и значения порогового значения ε_0.

Пусть правила вывода П СОК (3) и ООК (5) истинны (то есть выполняются эвристические утверждения). Рассмотрим условия, которые влияют на качество решения.

Достаточным условием, при котором СОК гарантирует получение качественного решения задачи прогнозирования, является транзитивность отношения близости на кластере соседей $\mathcal{N}_U = \{u : u_a \mathcal{R} u\}$, который строится для решения задачи: $\forall u_1, u_2 \in \mathcal{N}_U : (u_1 \mathcal{R}_u u_a) \wedge (u_2 \mathcal{R}_u u_a) \Rightarrow u_1 \mathcal{R}_u u_2$. Назовем это условие *условием 1*.

Достаточным условием, при котором ООК гарантирует получение качественного решения задачи $topN$, является транзитивность отношения близости на объединении обучающего, тестового и результирующего множеств:

$$(i \mathcal{R}_i j) \wedge (i \mathcal{R}_i k) \Rightarrow (j \mathcal{R}_i k), i, j, k \in I_0 \cup I_{topN} \cup I_\perp,$$
$$I_\perp = \{i_\perp, I_0 = \{i_0\}.$$

Назовем его *условием 2*.

Выполнение достаточных условий зависит от того, какое значение выбрано в качестве порогового значения ε_0, и функции, использующейся в качестве меры близости. К примеру, если $\delta_i = \cos$ и $\varepsilon_0 = 0{,}49$, то транзитивность не гарантируется; коэффициент корреляции Пирсона [6], являющийся традиционной мерой близости СОК, не обладает свойством транзитивности [9].

Если эвристические утверждения выполняются, то правила вывода П гарантируют получение качественного решения, если выполняются достаточные условия, что зависит от разработчиков системы.

Проблемы, описанные в разделах 2.1 и 2.2, подтверждены на практике и продемонстрированы ниже в Разделе 4, в Таблицах 1 и 2.

3.3 Масштабируемость

Стандартные алгоритмы решений КРС обладают следующими асимптотическими сложностями [10]: $O(|I|^2)$ для задачи $topN$, $O(|U|)$ для задачи прогнозирования. Учитывая огромную мощность множеств U, I, такие асимптотические сложности приводят к проблеме масштабируемости КРС [10].

4 Нечеткая контентная модель

4.1 Описание

В нечеткой контентной модели будем представлять контент в виде нечеткое подмножества множества характеристик [11]: $\{(c | w_M(m, c))\}$, где c – характеристика пользователя или объекта, $m \in M$ – множество пользователей или объектов, w_M – характеристическая функция принадлежности. Для СОК и ООК контент пользователя представляется в виде нечеткого множества вида $\{(i | 1 - \rho(u, i))\}$. Между пользователями и объектам введем расстояния ρ_u и ρ_i соответственно как обобщенное расстояние Хэмминга, которое обладает метрическими свойствами.

При представлении контентов в виде нечетких множеств определим нечеткое отображение $\Psi: U \to I$, характеристическая функция которого задана следующей формулой:

$$\nu_\Psi(y) = \max_{x \in X} \min\{\delta_c(x, y); w_U(u, x)\}, \quad (6)$$

и расстояние между пользователем и объектом

$$\overline{\rho}(u, i) = \rho_i(\Psi(u), i). \quad (7)$$

Функция $\delta_c: X \times Y \to [0,1]$ – это функция сходства характеристик пользователей и объектов, задание которой необходимо для построения отображения Ψ. Эта функция может быть определена разработчиками РС, экспертами, алгоритмически и т. д. Будем говорить, что оценка сходства δ_c задана аккуратно, если выполняется неравенство

$$|\rho(u,i) - \overline{\rho}(u,i)| \le \varepsilon_0 \quad (9)$$

Нечеткое правило вычисления Π_f заключается в задании оценки сходства δ_c, нечеткого отображения Ψ (6) и вычислении расстояния между пользователем и объектом, определенного формулой (7):

$$\Pi_f = \{\delta_c, (6), (7)\}. \quad (10)$$

Нечеткая контентная модель – это модель, которая задается следующей тройкой:

$$(c_X; c_Y; \Pi \in \{\Pi_f, \Pi_{\text{СОК}}, \Pi_{\text{ООК}}\}). \quad (11)$$

4.2 Нечеткая модель как эффективное расширение коллаборативной модели

Утверждение 1: нечеткая контентная модель (11) является эффективным расширением СОК по критерию качества.

Утверждение 1 следует из того, что СОК – частный случай модели (11) при использовании $\Pi = \Pi_{\text{СОК}}$. Расширение эффективно по критерию качества, так как выполняется условие 1. Покажем, что это верно: введем следующее дополнительное условие при составлении кластера соседей – $\mathcal{N}_U = \{u: \rho_u(u_a, u) \le \varepsilon_0/2\}$. Покажем, что выполняется достаточное условие. Напомним, что оно заключается в выполнении свойства транзитивности отношения близости $\mathcal{R}_u$ на кластере соседей: $\forall u, v \in \mathcal{N}_U$ верно, что $(u_a \mathcal{R}_u u) \wedge (u_a \mathcal{R}_u v)$.

Так как функция ρ_u обладает метрическими свойствами, то $\rho_u(u, v) \le \rho_u(u_a, u) + \rho_u(u_a, v)$. По дополнительному условию $\rho_u(u_a, u) \le \varepsilon_0/2$, $\rho_u(u_a, v) \le \varepsilon_0/2$, поэтому $\rho_u(u, v) \le \varepsilon_0 \Rightarrow u \mathcal{R}_u v$.

Утверждение 2: нечеткая контентная модель (11) является эффективным расширением ООК по критерию качества.

Утверждение 2 следует из того, что ООК – частный случай модели (11) при использовании $\Pi = \Pi_{\text{ООК}}$. Расширение эффективно по критерию качества, так как выполняется условие 2 при $\varepsilon_0 = 0$. Покажем, что выполняется условие 2. Напомним, что оно заключается в выполнении свойства транзитивности отношения близости $\mathcal{R}_i$ на множестве $I_0 \cup I_\perp \cup I_{topN}$. Покажем, что $(i_0 \mathcal{R}_i i) \wedge (i_0 \mathcal{R}_i i_\perp) \Rightarrow i_\perp \mathcal{R}_i i$.

Отношение $i_0 \mathcal{R}_i i_\perp$ выполняется по эвристическому утверждению ООК (4), отношение $i_0 \mathcal{R}_i i$ выполняется по построению решения. Так как функция ρ_i обладает метрическими свойствами, то $\rho_i(i, i_\perp) \le \rho_i(i_0, i) + c_i(i_\perp, i_0)$. По дополнительному условию $\rho_i(i_0, i) = 0$, так как выполняется $i_0 \mathcal{R}_i i_\perp$, то $\rho_i(i_0, i_\perp) \le \varepsilon_0$. Поэтому $\rho_i(i, i_\perp) \le \varepsilon_0$, то есть выполняется отношение $i \mathcal{R}_i i_\perp$.

Таким образом, правила вывода $\Pi_{\text{СОК}}, \Pi_{\text{ООК}}$ в представлении контентов в виде нечетких подмножеств и при использовании метрических расстояний обладают большей эффективностью по критерию качества решения, так как выполняются достаточные условия 1 и 2, и поэтому контентная нечеткая модель является эффективным расширением по критерию качества. Данный вывод подтверждается практическими результатами (см. таблицы 1 и 2).

4.3 Применение нечеткого правила вывода для решения задач

Определим решения в нечеткой контентной модели при использовании Π_f.

Задача $topN$ может быть решена при помощи *линейного поиска* объектов, таких, что $\overline{\rho}(u_a, i) \le \varepsilon_0$. Асимптотическая сложность такого алгоритма равна $O(|I|)$.

Для решения задачи прогнозирования нужно всего лишь рассчитать значение $\overline{\rho}(u_a, i_p)$, поэтому асимптотическая сложность такого решения равна $O(C)$.

Если оценка сходства δ_c *задана аккуратно*, то решения задач контентной нечеткой модели *эффективны* по критерию качества, что будет продемонстрировано на разделе 4. Точность задания δ_c зависит только от разработчиков, но не от свойств исходных данных или дополнительных условий, как в случае с КРС.

Асимптотические сложности алгоритмов решений при использовании правила вывода Π_f на порядок меньше по сравнению со сложностями КРС, поэтому представленная модель более эффективна по критерию масштабируемости, чем КРС. Каждый раз, когда производится вычисление $\overline{c}$, производятся отображение Ψ и расчет δ_c. Сложности вычислений отображения Ψ и δ_c зависят от мощности контента (которое, как правило, значительно меньше мощности множеств пользователей и объектов) и от того, как была задана δ_c разработчиками, поэтому эти сложности не учтены в расчетах, представленных ниже.

Приведенные значения асимптотических сложностей показывают, что контентная нечеткая модель является эффективным расширением КРС по критерию масштабируемости.

5 Практические результаты

Для получения практических результатов было разработано программное обеспечение, которое реализует ООК, СОК и нечеткую контентную РС. С помощью ООК решалась задача $topN$, с помощью СОК – задача прогнозирования. С помощью нечеткой РС решались обе задачи.

Тестирование проводилось на множестве данных, сформированных компанией MovieLens. Множество

данных имеет следующие характеристики:

- $|I|$=10000 – объектами множества являются фильмы, численность которых равна 10000;
- $|Y| = 18$ – множество характеристик объектов состоит из 18 кинематографических жанров;
- $|U| = 670$ – число пользователей данного множества равно 671; пользователи являются реальными людьми, которые проставляли оценки близости различным объектам.

Для решения задач $topN$ и прогнозирования в ООК и СОК соответственно были использованы стандартные алгоритмы и подходы [6, 12]. При решении задачи $topN$ в ООК использовалась мера сходства косинус, при решении задачи прогнозирования в СОК – коэффициент корреляции Пирсона. Те же алгоритмы были применены при решении задач в нечеткой контентной модели, но при этом использовались расстояния ρ_i и ρ_u. Пороговое значение ε_0 было принято равным 0,1.

Чтобы применить Π_f, была задана функция δ_c на основании эвристического предположения о том, что между оценкой пользователя и жанрами объектов существует корреляция:

$$\delta_c(i,y) = (|like_y| - |dislike_y|)/|P_u|.$$

Если $\delta_c(i,y) < 0$, то $\delta_c(i,y) = 0{,}0001$;

$$like_y = \{i: (\rho(u,i) \le \varepsilon_0) \wedge w_U(i,y) \neq 0\},$$

$$dislike_y = \{i: (\rho(u,i) > \varepsilon_0) \wedge w_U(i,y) \neq 0\},$$

$$P_u = \{i: (\rho(u,i) \neq \perp)\}.$$

Такое эвристическое предположение верно не для всех пользователей, так как их вкусы могут быть неоднородными. Поэтому для некоторых пользователей функция δ_c задана аккуратно, а для некоторых – нет.

Стандартно при проведении тестирования данные о пользователе случайно разбивались в следующем отношении: 80% – обучающее множество, 20% – тестовое. Обозначим такое разбиение цифрой 1. Помимо стандартного разбиения использовались и другие специально сформированные разбиения 2 и 3. Разбиение 2 составлено так, что обучающее множество состоит из таких объектов i, для которых выолняется отношение $\mathcal{R}_i$, тестовое множество состоит из таких объектов j, для которых отношение $i\,\mathcal{R}_i\,j$ не выполняется. Такое разбиение создано для того, чтобы подтвердить или опровергнуть влияние свойства неоднородности данных на эффективность по критерию качества. Разбиение 3 составлено так же, как и стандартное разбиение, но в нем участвуют только те пользователи, для которых функция δ_c задана аккуратно.

Эффективность решений задач по критерию качества определяется усредненными по числу тестов (равному 1000 для каждой задачи, разбиению и модели) значениями функций. Эффективность решения задачи $topN$ по критерию качества оценивалась значениями функций точность (P), точность по списку длины L, средняя точность, NDCG. В результате тестирования среднее значение этих функций мало отличалось, поэтому в Таблице 1 приведены только значения точности. Большее значение точности свидетельствует о то, что решение более эффективно. Эффективность решения задачи прогнозирования по критерию качества оценивалась значениями функций MAE, NMAE, RMSE, меньшее значение которых говорит о более эффективном решении.

Таблица 1

№	Модель/Правило вычисления	Разбиение	P
1	ООК/ $\Pi_{ООК}$	1	0.32
2	ООК/ $\Pi_{ООК}$	2	0,24
3	Нечеткая контентная / $\Pi_{ООК}$	1	0,55
4	Нечеткая контентная / $\Pi_{ООК}$	2	0,53
5	Нечеткая контентная/ Π_f	1	0.39
6	Нечеткая контентная/ Π_f	2	0,36
7	Нечеткая контентная/ Π_f	3	0,81

Прокомментируем данные таблиц 1 и 2. Результаты 1 эффективнее результатов 2 и результаты 3 эффективнее результатов 4, что подтверждает теоретические выводы о влиянии свойства неоднородности на эффективность по критерию качества при применении ООК. Разбиение 2 задано так, что свойства неоднородности влияют на эффективность решения, так как между объектами обучающего и тестового множеств не выполняется отношение сходства, в результате чего нарушается утверждение ООК (4). Разбиение 2 увеличивает вероятность того, что утверждение СОК (5) может быть неверным, поэтому результаты 1 и 3 эффективней результатов 2 и 4 Таблицы 2.

Таблица 2

№	Модель/ Правило вычисления	Разбиение	MAE	NMAE	RMSE
1	ООК/ $\Pi_{СОК}$	1	0.14	0.23	0.19
2	ООК/ $\Pi_{СОК}$	2	0.16	0.26	0.21
3	Нечеткая контентная / $\Pi_{СОК}$	1	0.08	0.19	0.13
4	Нечеткая контентная / $\Pi_{СОК}$	2	0.10	0.17	0.18
5	Нечеткая контентная/ Π_f	1	0.14	0.23	0.21
6	Нечеткая контентная/ Π_f	2	0.16	0.26	0.22
7	Нечеткая контентная/ Π_f	3	0.05	0.04	0.1

Результаты 3 и 4 эффективнее результатов 1 и 2, что подтверждает вывод о том, что нечеткая контентная модель является эффективным расширением, так как в ней выполняются достаточные условия 1 и 2. Эти же результаты подтверждают выводы о влиянии меры сходства на эффективность ООК и СОК по критерию качества.

Результаты 7 эффективнее результатов 3–6, так как для разбиения 7 функция δ_c задана аккуратно. Результаты 7 эффективнее результатов 5 и 6, так как для 5 и 6 в общем случае δ_c не задана аккуратно, и поэтому же 5 и 6 не эффективнее 3 и 4. Результаты 5 эффективнее 6, так как функция δ_c задавалась на основании данных обучающего множества, поэтому свойство неоднородности влияет на аккуратность функции так же, как и на эффективность КРС по критерию качества. Использование Π_f может быть неэффективным, если о пользователях известна только та информация, которая принадлежит исходному множество P. В такой ситуации эффективнее использовать нечеткую модель Π_{OOK} или Π_{COK}. Для задания функции δ_c можно использовать информацию, которая никак не зависит от мощности и свойств исходных данных, и тогда решения задач в нечеткой контентной модели не будут зависеть от свойств исходных данных. Такой информацией может выступать, к примеру, контекстная информация [13].

6 Заключение

Нечеткая контентная модель РС, представленная в настоящей работе, является эффективным расширением КРС по критериям качества решений и масштабируемости.

Литература

[1] Resnick, P., Varian, H.R.: Recommender systems. Communications of the ACM, 40 (2), pp. 56-58 (1997)

[2] Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35 (12), pp. 61-70 (1992)

[3] Asanov, D.: Algorithms and Methods in Recommender Systems. Berlin Institute of Technology. https://www.snet.tu-berlin.de/fileadmin/ fg220/courses/SS11/snet-project/ recommender-systems_asanov.pdf

[4] Yao, W., Xudong, L., Min, X., Ester, M., Qing, Y.: CCCF: Improving Collaborative Filtering via Scalable User-Item Co-Clustering. WSDM '16 Proc. of the Ninth ACM Int. Conf. on Web Search and Data Mining, pp. 73-82 (2016)

[5] Hu, R., Pu, P.: Using personality information in collaborative filtering for new users. Recommender Systems and the Social Web, pp. 17-24 (2010)

[6] Su, X., Khoshgoftaar, T.: A survey of collaborative filtering based social recommender systems. Computer Communications, 41, pp. 1-10 (2014)

[7] Wang Jun: Unifying user-based and item-based collaborative filtering approaches by similarity fusion. SIGIR'06 Proc. of the 29th Annual International ACM, pp. 501-508 (2006)

[8] Посыпанова, О.: Экономическая психология: психологические аспекты поведения потребителей. Калуга: Изд-во Калужского государственного университета им. К.Э. Циолковского, 296 с. (2012)

[9] Castro Sotos, A., Vanhoof, S., Van den Noortgate, W., Onghena, P.: The non-transitivity of Pearson's correlation coefficient: an educational perspective. Proc. of the 56th Session of the ISI, 62, pp. 4609-4613 (2007)

[10] Linden, G., Smith, B., York, J.: Amazon.com Recommendations Item-to-Item Collaborative Filtering. Internet Computing, IEEE, 7, pp. 76-80 (2003)

[11] Амелькин, С.А., Понизовкин, Д.П.: Математическая модель задачи topN для контентных рекомендательных систем. Изв. МГТУ МАМИ, 2, сс. 26-31 (2013)

[12] Deshpande, M., Karypis, G.: Item-Based Top-N Recommendation Algorithms. ACM Transactions on Information Systems, 22 (1), pp. 143-177 (2004)

[13] Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. Conference: Proc. of the 2008 ACM Conference on Recommender Systems, RecSys 2008, Lausanne, Switzerland, October 23–25, 2008. doi: 10.1007/978-1-4899-7637-6_6