УКД 004.67

КВЯТКОВСКАЯ И.Ю., ВО ТХИ ХУЕН ЧАНГ, ЧАН КУОК ТОАН

KVYATKOVSKAYA IRINA YURIEVNA, VO THI HUYEN TRANG, TRAN QUOC TOAN

**РАЗРАБОТКА МЕРЫ СХОДСТВА ДЛЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ МЕТОДА АНАЛИЗ СТАТИСТИЧЕСКОЙ ИМПЛИКАЦИИ**

**DEVELOPMENT OF A SIMILARITY MEASURE FOR A RECOMMENDER SYSTEM BASED ON THE METHOD OF ANALYSIS OF STATISTICAL IMPLICATION**

*Мера сходства играет важную роль в рекомендательной системе совместной фильтрации, основанной на пользователях, поскольку она напрямую влияет на результаты рекомендательной системах. Чтобы определить меру сходства между двумя пользователями в рекомендательной системе, предлагается множество решений, таких как: использование статистической корреляции, использование косинусного расстояния между двумя векторами, использование ассоциативных правил и т.д... В этой работе предлагает меру сходства, основанную на методе анализа статистической импликации. Мера сходства между двумя пользователями определяется на основе суммы расстояний статистической импликации ассоциативных правил, которые склонны отдавать предпочтение контр примерам (число объектов, которые удовлетворяют свойству а, но не удовлетворяют свойству b в ассоциативном правиле ), созданным на основе данных ранжирования двух пользователей.*

*Ключевые слова:**мера сходства, рекомендательная система, анализ статистической импликации, мере интенсивности статистической импликации, рейтинг продуктов, рейтинговая матрица, ассоциативное правило.*

*The measure of similarity plays an important role in the recommender system of collaborative filtering based on users, since it directly affects the results of recommender systems. To determine the measure of similarity between two users in a recommender system, many solutions are proposed, such as: using statistical correlation, using cosine distance between two vectors, using association rules, etc… In this paper, he proposes a measure of similarity based on the statistical implication analysis method. The measure of similarity between two users is determined based on the sum of the statistical implication distances of the association rules, which tend to favor counter-examples (the number of objects that satisfy property a but not property b in the association rule) generated from the ranking data of the two users.*

*Key words: similarity measure, recommender system, statistical implication analysis, statistical implication intensity measure, product rating, rating matrix, association rule.*

Мера сходства на основе интенсивности статистической импликации – это мера, используемая для определения значения сходства между двумя пользователями на основе набора ассоциативных правил, созданных на основе данных рейтинга двух пользователей, и меры интенсивности статистической импликации. Мера сходства на основе интенсивности статистической импликации между двумя пользователями  определяется по следующей формуле:

 (1)

где,  – мера сходства между двумя пользователями ;  – величина интенсивности статистической импликации ассоциативного правила ;  – количество правил ассоциации в наборе правил ассоциации, созданном на основе данных ранжирования двух пользователей .

Мера интенсивности статистической импликации  правила  определяется формулой [1]:

. (2)

где,  – количество пользователей,  и  – количество элементов множеств *A* и *B*;  – число контр примеров (это число объектов, которые удовлетворяют свойству , но не удовлетворяют свойству *b*).

*Псевдокод алгоритма для определения меры сходства статистической импликации между двумя пользователями  определяется следующим образом:*

Входные данные: рейтинговые данные для продуктов двух пользователей .

Вывод: значение сходства между двумя пользователями .

Начало

Шаг 1. Производится сгенерировать ассоциативных правил из рейтинговой матрицы пользователя.

Шаг 2. Производится выбрать ассоциативные правила для двух пользователей :

<Выбрать продукты, оцененные пользователями : >;

<Выбрать продукты, не оцененные пользователями : >;

<Выбрать ассоциативные правила вида ,где ;  и >

Шаг 3: Производится вычислить значения параметров  для каждого правила из выбранного набора ассоциативных правил.

Шаг 4: Производится вычислить величину интенсивности статистической импликации для выбранного набора ассоциативных правил.

Шаг 5: Производится вычислить сходство между двумя пользователями :

<Вычислить среднее значение интенсивности статистической импликации для набора правил ассоциации >;

<Определить значение сходства между двумя пользователями , :>.

*Примечание:* , то . Таким образом,.

***Пример 1:*** Учитывая рейтинговую матрицу двух пользователей, оценивающих 4 продукта, представлена в таблице 1.

Таблица 1. Рейтинговая матрица двух пользователей

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | i1 | i2 | i3 | i4 |
| u1 | 0 | 4 | 4 | 1 |
| u2 | 0 | 0 | 4 | 0 |

Первым шагом является создание ассоциативных правил из матрицы ранжирования. Затем производится выбор правила для двух пользователей . Ассоциативные правила двух пользователей показаны в таблице 2.

Таблица 2. Ассоциативные правила двух пользователей 

|  |  |
| --- | --- |
| № | Ассоциативные правила  |
| 1 | {i2=4}=>{i1=0} |
| 2 | {i4=1}=>{i1=0} |
| 3 | {i3=4}=>{i1=0} |
| 4 | {i3=4}=>{i2=0} |
| 5 | {i3=4}=>{i4=0} |
| 6 | {i2=4,i4=1}=>{i1=0} |
| 7 | {i2=4,i3=4}=>{i1=0} |
| 8 | {i3=4,i4=1}=>{i1=0} |
| 9 | {i2=4,i3=4,i4=1}=>{i1=0} |

Следующим шагом является определение параметров  для каждого правила ассоциации и вычисление величины интенсивности статистической импликации на основе этих параметров. Результаты этого шага представлены в таблице 3.

##### Таблица 3. Значение параметров и величина интенсивности статистической импликации каждого правила ассоциации

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № | Правило ассоциации | 𝑛 | 𝑛𝐴 | 𝑛𝐵 | 𝑛𝐴𝐵̅ | Интенсивность |
| 1 | {i2=4}=>{i1=0} | 2 | 1 | 1 | 0 | 0.49 |
| 2 | {i4=1}=>{i1=0} | 2 | 1 | 1 | 0 | 0.49 |
| 3 | {i3=4}=>{i1=0} | 2 | 2 | 1 | 1 | 0.38 |
| 4 | {i3=4}=>{i2=0} | 2 | 2 | 1 | 1 | 0.38 |
| 5 | {i3=4}=>{i4=0} | 2 | 2 | 1 | 1 | 0.38 |
| 6 | {i2=4,i4=1}=>{i1=0} | 2 | 1 | 1 | 0 | 0.49 |
| 7 | {i2=4,i3=4}=> {i1=0} | 2 | 1 | 1 | 0 | 0.49 |
| 8 | {i3=4,i4=1}=> {i1=0} | 2 | 1 | 1 | 0 | 0.49 |
| 9 | {i2=4,i3=4,i4=1}=>{i1=0} | 2 | 1 | 1 | 0 | 0.49 |

Наконец, сходство между пользователямиопределяется следующим образом: 

**Алгоритм поддержки принятия решения для прогноза рейтинга продуктов при отсутствующих оценках на основе меры сходства статистической импликации**

Рекомендательная модель совместной фильтрации, основанная на мере сходства статистической импликации определяется следующим образом:

 (3)

где,  – множество  пользователей системы; – множество продуктов системы; – рейтинговая матрица пользователей для продуктов, где каждая строка представляет пользователя  (), каждый столбец представляет продукт  (), – это значение рейтинга пользователя  для продукта ;  – это вычислительная функция, позволяющая определить продукты, которые необходимо рекомендовать пользователю : 

*Псевдокод алгоритма поддержки принятия решения для прогноза рейтинга продуктов при отсутствующих оценках на основе меры сходства статистической импликации:*

*Входные данные:*Набор пользователей , набор продуктов , рейтинговая матрица пользователей для продуктов, пользователю нужна рекомендация.

*Выход:*Продукты, которые необходимо рекомендовать пользователю :.

Начало

Шаг 1: Определить список  пользователей, похожих на пользователя .

Для каждого пользователя  выполнить:

<Определение значения меры сходства между  и  с использованием меры сходства статистической импликации: >;

<Сортировать список пользователей по убыванию значения сходства>;

<Выберите первых k пользователей с наибольшим значением меры сходства: >.

Шаг 2. Производится рассчитать прогнозируемые рейтинги продуктов:

<Определить продукты, которые пользователь  не оценил ()>;

<Рассчитать прогнозируемые рейтинги для этих продуктов по формуле: , где  – значения сходства между  и ,  – значение рейтинга пользователя  для продукта >.

Шаг 3: Выбрать продукты для рекомендации пользователю :

<Сортировать продукты по убыванию значения прогнозируемых рейтингов продуктов>;

<Выбрать продукты с наивысшим значением прогнозируемых рейтингов, чтобы представить пользователю >.

***Пример 2*.** Чтобы более наглядно увидеть шаги алгоритма, предположим, что система предлагает пользователям выбрать 8 продуктов (от  до ), и в настоящее время в системе есть 10 пользователей (от  до ), которым присвоены рейтинги продуктов. Товары оцениваются по шкале от 1 до 5 («1» — самая низкая оценка; «5» — самая высокая оценка; а «?» — продукт не имеет рейтинга от пользователя). Система должна рекомендовать продукты новому пользователю  с коэффициентом , равным 4 (рассчитывается для 4 похожих пользователей), как показано на таблице 4.

Таблице 4 – Рейтинговая матрица между пользователями и продуктами

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  | ? | 4,0 | 4,0 | 1,0 | 2,0 | 2,0 | ? | ? |
|  | 3,0 | ? | ? | ? | 5,0 | 1,0 | ? | ? |
|  | 4,0 | ? | ? | 3,0 | 2,0 | 2,0 | ? | 3,0 |
|  | 3,0 | ? | 3,0 | 2,0 | 1,0 | ? | 5,0 | 4,0 |
|  | 1,0 | 1,0 | ? | ? | ? | ? | ? | 1,0 |
|  | ? | 1,0 | ? | ? | ? | 1,0 | ? | 1,0 |
|  | 1,0 | 4,0 | ? | 2,0 | ? | 4,0 | 4,0 | ? |
|  | 5,0 | ? | 4,0 | 3,0 | ? | 2,0 | 3,0 | 1,0 |
|  | ? | 1,0 | 3,0 | ? | ? | ? | 1,0 | ? |
|  | 2,0 | ? | ? | ? | ? | 3,0 | 2,0 | 1,0 |
|  | ? | ? | 4,0 | 2,0 | ? | 1,0 | ? | 3,0 |

Из этого требования система определяет список пользователей, похожих на , на основе меры сходства статистической импликации, включая:,,,, как показано на рисунке 1:



Рисунок 1 – Определение список пользователей, похожих на пользователя  при *k* = 4

Таблице 5 – Расчет списка прогнозируемых продуктов для пользователя 

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  | ? | ? | 4,0 | 2,0 | ? | 1,0 | ? | 3,0 |
|  | 3,0 | 4,0 |  |  | 1,5 | 4,0 |  |  |

На основе списка похожих пользователей система производится вычислить значение рейтинга для продуктов пользователя , которых без рейтинга и дает пользователю  рекомендуемые продукты, включая: i1, i2, i7, как показано на таблице 5.

**Заключение**

В этой работе разработана мера сходства для рекомендательной системы на основе метода анализ статистической импликации. Разработан алгоритм поддержки принятия решения для прогноза рейтинга продуктов при отсутствующих оценках на основе меры сходства статистической импликации.

**СПИСОК ЛИТЕРАТУРЫ**

1. Квятковская И. Ю., Во Тхи Хуен Чанг, Чан Куок Тоан. Модель и алгоритм поддержки принятия решения по выбору продуктов для рекомендации пользователю на основе метода анализа статистической импликации // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика 2023. № 2. С. 116-124. https://doi.org/10.24143/2072-9502-2023-2-116-124. EDN UHNZRL/
2. David H. Glass. Confirmation measures of association rule interestingness // Knowledge-Based Systems, – 2013, 44, pp. 65-77.
3. Шуршев В.Ф., Кочкин Г.А., Кочкина В.Р. [Модель системы поддержки принятия решений на основе рассуждений по прецедентам](https://www.elibrary.ru/item.asp?id=19424474) // [Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика](https://www.elibrary.ru/contents.asp?id=33836435). 2013. [№ 2](https://www.elibrary.ru/contents.asp?id=33836435&selid=19424474). С. 175-183.
4. Kvyatkovskaya I.Y., Shurshev V.F., Frenkel M.B. [Methodology of a support of making management decisions for poorly structured problems](https://www.elibrary.ru/item.asp?id=26811374) // [Communications in Computer and Information Science](https://www.elibrary.ru/contents.asp?id=34274239). 2015. Т. 535. С. 278-291.
5. R. Gras, P. Kuntz. An overview of the Statistical Implicative Analysis (SIA) development // Statistical Implicative Analysis - Studies in Computational Intelligence, Springer-Verlag, 2008,127, pp. 11-40.
6. R. Gras, Pascale Kuntz, Nicolas Greffard. Notion of Implicative Fields in Statistical Implicative Analysis // VIII Colloque International - VIII International Conference A.S.I. Analyse Statistique Implicative - Statistical Implicative Analysis Radès (Tunisie) - Novembre 2015, pp. 29-46.

**Квятковская Ирина Юрьевна**

Астраханский государственный технический университет

Д.т.н, профессор, заведующая кафедрой «Высшая и прикладная математика»

Email: i.kvyatkovskaya@astu.org

**Во Тхи Хуен Чанг**

Астраханский государственный технический университет

Аспирант кафедры «Автоматизированные системы обработки информации и управления»

Email: vthtrang@mail.ru

**Чан Куок Тоан**

Астраханский государственный технический университет

К.т.н. кафедры «Высшая и прикладная математика»

Email: hoaivan219@mail.ru