

Филиппов С.А., Захаров В.Н.

Федеральный исследовательский центр "Информатика и Управление" Российской академии наук,
г. Москва, Россия

**ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДА ПОВЫШЕНИЯ ПЕРТИНЕНТНОСТИ
ИНФОРМАЦИИ В РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМАХ ПОДДЕРЖКИ
ЖИЗНЕОБЕСПЕЧЕНИЯ НА ОСНОВЕ НЕЯВНЫХ ДАННЫХ**

АННОТАЦИЯ

В статье предложен подход к оценке эффективности метода повышения pertinence информации, работающего с неявными большими данными пользователей, на основе результатов экспериментальных исследований программной реализации метода, внедрённой в действующий интернет-магазин Thaisoap. Работа выполнена при поддержке Министерства образования и науки РФ, уникальный идентификатор проекта RFMEFI60414X0139.

КЛЮЧЕВЫЕ СЛОВА

Пертинентность, коллаборативная фильтрация, интернет-магазин, рекомендательная система, неявные данные.

Philippov S.A., Zakharov V.N.

Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, Russia

**EVALUATION OF THE EFFECTIVENESS OF THE PERTINENCE METHOD THAT
INCREASE SALES FOR E-COMMERCE RECOMMENDER SYSTEMS BASED ON REAL
IMPLICIT DATA**

ABSTRACT

This article proposes an approach to the evaluation of the effectiveness of the method that increase pertinence of information, using users implicit big data, based on the results of experimental studies of software implementation of the method are embedded in the real online store Thaisoap. This work was supported by the Ministry of Education and Science of the Russian Federation. A unique identifier RFMEFI60414X0139 project.

KEYWORDS

Pertinence, collaborative filtering, e-commerce recommender system, implicit data targeting.

Ещё на заре электронной коммерции в 1998 году основатель Amazon.com Джефф Безос объявил о необходимости персонализации информационного предложения фразой «Если мы имеем 4,5 млн покупателей... мы должны иметь 4,5 млн магазинов» [1]. На практике, основными целями использования персонализации являются:

1. Предоставление релевантного контента. Персонализация содержимого сайтов позволяет упростить процесс нахождения необходимой информации пользователями, а также повышает общую эффективность сайтов;
2. Адресная (целевая) реклама или таргетинг. Выстраивание рекламы в зависимости от потребностей конкретных групп пользователей, основанное на истории их взаимодействия с сайтом. Существует большое количество видов адресной рекламы, наиболее перспективным из которых является поведенческий таргетинг (формирование рекламных предложений на базе поведенческого профиля пользователя).

В целом все методы персонализации можно разделить на две большие группы: персонализация, основанная на правилах и персонализация, основанная на алгоритмах. Персонализация сайтов, основанная на правилах – это практика использования исторических, географических, демографических и других данных для формирования предложений, основанных на predetermined правилах. Типичный вариант персонализации, основанный на правилах принимает форму «если пользователь совершает действие X, то показывать предложение Y». Например, если пользователь сайта переходит в раздел подержанных автомобилей, то ему

предлагается посетить сайты автосервисов (или продавцов автомобильных комплектующих), расположенных на удалении не более 50 км от местоположения пользователя. Персонализация, основанная на правилах, применима в том случае, когда пользователи могут быть разделены на группы (сегментированы) с явно выраженными признаками (например, сообщества владельцев автомобилей конкретных марок). Для сегментации пользователей и формирования адресных предложений, как правило, используются специализированные сервисы (например, Aigee.co и Monoloop.com).

Персонализация, основанная на алгоритмах, использует предикативные модели, учитывающие самые разные сведения о пользовательской активности с целью формирования контента, который наиболее полно отражает ожидания пользователей. Данный подход считается наиболее эффективным на сегодняшний день, но требует значительных вычислительных ресурсов, как для хранения поведенческих данных, так и для их обработки, анализа и выработки рекомендаций. Подобные алгоритмы составляют основу рекомендательных систем, реализующих с целью выработки рекомендаций анализ поведенческих привычек пользователей или оценку близости товаров по рейтингу их популярности у пользователей.

Небезосновательно считается, что предоставление персонализированного контента пользователям позволяет существенно повысить эффективность сайтов, которая выражается в терминологии маркетинга таким показателем как конверсия (число посетителей, совершивших полезные с точки зрения владельца системы действия к общему числу посетителей, выраженное в процентах). Для качественной персонализации сайтов, ориентированных на работу с большой аудиторией пользователей, как правило, используется комплексный подход, сочетающий маркетинговые исследования и анализ поведения конкретных посетителей сайтов. Информацию о маркетинговых качествах посетителей можно получить, в том числе используя системы веб-аналитики, такие как Adobe Digital Marketing Suite или Google Analytics вместе с Siteapps.com. В сфере электронной коммерции основным инструментом персонализации контента являются рекомендательные системы, обеспечивающие автоматическую обработку данных о пользовательской активности и выработку рекомендаций на товары и услуги, которые могут быть интересны конкретным пользователям [2]. Исходными данными для анализа поведения пользователей являются сведения об их активности, которые могут собираться явным или неявным образом. Явным образом получают персональные данные при регистрации, покупках, выставлении пользователем оценок тем или иным объектам на сайтах, в результате голосований и опросов, другие данные, связанные с уже принятыми пользователем решениями. При этом основное количество информации о пользовательской активности собирается неявным образом посредством протоколирования их действий, т.е. процесса принятия решений. Результаты могут записываться в log файлы (системные журналы) или специализированные базы данных. Предметом отслеживания являются переходы пользователей по ссылкам на сайтах, время их пребывания на отдельных страницах, факты покупки товаров и услуг, географическое местоположение и т.п. Необходимо отметить, что, речь идет об огромных массивах данных, которые являются неоднородными и достаточно сложными для интерпретации.

В связи с большими объемами и неоднородностью данных о пользовательской активности, при исследовании этих данных часто используются методы интеллектуального анализа данных (Data Mining). Данные методы предназначены для выявления скрытых (или ранее неизвестных) знаний в исследуемых наборах данных (как правило, слабо структурированных). Знания, полученные посредством применения методов интеллектуального анализа данных, как правило, представляются в виде закономерностей (паттернов). В качестве них могут выступать ассоциативные правила, деревья решений, кластеры, математические функции. Наиболее распространенные задачи Data Mining - классификация, кластеризация, ассоциация, прогнозирование и визуализация. Одним из наиболее быстро развивающихся направлений в Data Mining является анализ связей между данными (link analysis), которое имеет приложения в таких областях как биоинформатика, цифровые библиотеки и защита против терроризма [3]. Все методы анализа данных, используемые в Data Mining, можно разделить на две большие группы: статистические методы и кибернетические методы. К статистическим методам относятся дескриптивный анализ и описание исходных данных, анализ связей (корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ), многомерный статистический анализ (кластерный анализ, компонентный анализ, дискриминантный анализ, многомерный регрессионный анализ, канонические корреляции и др.), анализ временных рядов (динамические модели и прогнозирование). К группе кибернетических методов относятся искусственные нейронные сети (распознавание, кластеризация, прогноз), эволюционное программирование (в т. ч. алгоритмы метода группового учета аргументов), генетические алгоритмы (оптимизация),

ассоциативная память (поиск аналогов, прототипов), нечеткая логика, деревья решений, системы обработки экспертных знаний [3, 4]. В современных рекомендательных системах для формирования предикативной модели часто используются подходы, использующие алгоритмы интеллектуального анализа данных. В частности, хорошее распространение в этой области получило использование кластерного анализа.

Для выявления пользовательских предпочтений рекомендательные системы собирают и анализируют большое количество данных о пользователях, включая их географическое местоположение, время пребывания на различных страницах целевого ресурса, переходы по ссылкам, ассортимент покупок и многое другое. Собранные данные о пользовательской активности характеризуются большим объемом и разнородностью, а также быстрым изменением (обновлением) во времени. Традиционные базы данных малоприменимы для работы с этими данными по причине больших объемов данных и повышенных требований к производительности [5]. Как правило, используются так называемые NoSQL системы управления данными (HBase, Cassandra). Их характерными особенностями являются отказ от транзакций, практически линейная масштабируемость, высокая скорость обработки запросов, отсутствие жесткой схемы данных.

Необходимо отметить, что адаптация под конкретного пользователя – весьма сложная задача, поскольку для ее решения необходимо принимать во внимание как присущие человеку неопределенность и спонтанность в рамках конкретного интернет-ресурса, так и множество неопределенностей, связанных с особенностями функционирования сети Интернет. В контексте проблемы персонализации контента (а также прогнозирования, выявления предпочтений и групп схожих ресурсов) встает задача обработки собранных данных и выявления определенных закономерностей, позволяющих сделать выводы о конкретных предпочтениях пользователей. Таким образом, основной целью обработки данных о пользовательской активности является извлечение полезной информации, которая может, в свою очередь, использоваться для решения следующих задач:

- кластеризация информационных единиц (в случае электронной коммерции – товаров);
- формирование поведенческих профилей пользователей (обобщение всей совокупности явных и неявных данных);
- кластеризация профилей пользователей;
- формирование для пользователя персонального информационного предложения из набора информационных единиц.

Таким образом основная задача рекомендательной системы – формирование контента, максимально соответствующего ожиданиям, в том числе неявным, конкретного пользователя. Для решения этой задачи в большинстве современных рекомендательных систем используются следующие базовые подходы: коллаборативная фильтрация (collaborative filtering, CF) и контентная фильтрация (content-based filtering, CbF) [6]. Метод контентной фильтрации фокусируется на выявлении объектов со схожими характеристиками по отношению к тем объектам, которые уже заинтересовали пользователя. При этом учитывается модель поведения пользователя и характеристики (контент) заинтересовавших его объектов. При выработке рекомендаций выявляются объекты со схожими характеристиками (контентом). Для эффективной работы метода контентной фильтрации, как правило, необходимо подробное описание характеристик объектов (так в проекте Music Genome Project музыкальный аналитик оценивает каждую композицию по сотням различных музыкальных характеристик), а также сведения о конкретном пользователе (например, ответы на конкретные вопросы в анкете).

В основе метода коллаборативной фильтрации лежит предположение о консервативности пользовательских предпочтений (т.е. пользователи, одинаково оценивающие определенные объекты, скорее всего аналогичным образом будут оценивать и новые объекты со сходными характеристиками) [7]. По существу, рекомендации базируются на автоматическом сотрудничестве множества пользователей и на выделении (методом фильтрации) тех пользователей, которые демонстрируют схожие предпочтения или шаблоны поведения. Таким образом, метод коллаборативной фильтрации вырабатывает рекомендации, основанные на модели предшествующего поведения пользователя и с учетом поведения пользователей со схожими характеристиками.

Наибольшее распространение в сфере электронной коммерции получили рекомендательные системы, использующие следующие варианты реализации метода коллаборативной фильтрации, а также их гибриды:

- коллаборативная фильтрация посредством анализа предпочтений групп пользователей со схожими интересами (User-User Collaborative Filtering, User-User CF);

- коллаборативная фильтрация посредством анализа взаимосвязей между объектами (Item-Item Collaborative Filtering, Item-Item CF).

Основными проблемами, связанными с реализацией и практическим использованием алгоритмов коллаборативной фильтрации, являются разреженность данных, проблема «холодного старта» и масштабируемость. Дополнительно к перечисленным проблемам можно отметить проблему ограничения разнообразия предложений. Рекомендательные системы, использующие коллаборативную фильтрацию, склонны предлагать товары уже пользующиеся популярностью, что создает проблемы для продвижения новых товаров и услуг [8].

В методе User-User CF определяется сходство между пользователями и в качестве рекомендаций пользователю выдается n самых часто покупаемых товаров k наиболее похожими на него покупателями. Для оценки степени схожести пользователей в плане их предпочтений могут использоваться различные функции сходства (метрики). Наиболее популярными среди них являются: евклидово расстояние, косинусная мера, расстояние Хэмминга, коэффициент корреляции Пирсона, коэффициент Танимото, Манхэттенское расстояние и так далее [5, 7]. Определение рекомендаций методом User-User CF предполагает построение матрицы активности пользователей, каждая строка которой описывает действия конкретного пользователя применительно к конкретному объекту (категория, товар, услуга) на сайте. Действия пользователей могут обозначаться самыми различными способами. Например, это может быть бинарная информация о посещении или не посещении заданного ресурса данным пользователем, частота (или число) использований ресурса R пользователем U , стоимость или рейтинг, предоставленный пользователем U для ресурса R и т.д. Таким образом, каждая строка матрицы активности представляет собой вектор оценок, соответствующих различным категориям товаров (тематический профиль пользователя). Профиль пользователя характеризует степень его интереса к каждой группе товаров. Для каждой пары «пользователь-объект (товар, услуга, действие)» в матрице активности вычисляется мера близости с использованием выбранной метрики [9].

Для поиска рекомендаций конкретному пользователю на основании его поведенческого профиля используются три основных подхода: основанный на соседстве (memory based), основанный на модели (model based) и гибридный подход (hybrid). В современных коммерческих системах наибольшее распространение получили гибридный подход и подход, основанный на использовании моделей (алгоритмы кластеризации, байесовские сети доверия, латентные семантические модели). Для выявления групп объектов (пользователи, товары, целевые группы) со схожими характеристиками часто используются различные алгоритмы кластеризации. В частности, в работе [10] указывается, что проблема идентификации групп пользователей по своей природе опирается на использование методов кластеризации. Кластеризация данных может быть также использована для генерации профилей пользователей на основе информации о действиях каждого пользователя, а затем для формирования групп пользователей на основе их профилей. Метод Item-Item CF исторически появился как альтернатива методу User-User CF, призванная повысить производительность рекомендательных систем для тех магазинов, где число покупателей существенно превышает количество товаров в каталоге [11]. Первоначально данный метод был предложен компанией Amazon для решения следующих основных проблем подхода User-User CF: проблема «холодного старта» и проблема частого обновления данных о пользовательской активности. Проблема «холодного старта» существенно снижает качество работы рекомендательной системы вследствие отсутствия данных о предпочтениях новых (или мало активных) пользователей. Проблема частого обновления данных о пользовательской активности (в случае компании Amazon речь идет о миллионах покупателей) резко снижает производительность рекомендательной системы в целом.

Основная идея метода Item-Item CF заключается в группировке информационных единиц (товары, услуги, действия), имеющих сходные оценки пользователей (рейтинги). Рекомендации вырабатываются по следующему принципу: пользователю, высоко оценившему объект X , будет предложен объект Y , который высоко оценили другие пользователи, также высоко оценившие и объект X . Использование метода Item-Item CF позволяет повысить качество рекомендаций для новых пользователей (нет критической зависимости от данных о пользовательских предпочтениях), а также значительно повышает производительность рекомендательной системы в случае, когда количество пользователей существенно превышает количество объектов (характеристики объектов меняются реже). В отличие от метода User-User CF, где вычисление степени близости в парах «пользователь-объект», как правило, производится в реальном времени (так как данные о текущей транзакции становятся доступными только в момент выработки рекомендаций), для метода Item-Item CF степень близости анализируемого товара ко всем остальным товарам может быть вычислена в отложенном режиме по расписанию, так как вектора

рейтингов всех товаров доступны до момента формирования рекомендации. Таким образом, благодаря возможности проведения отложенной обработки данных, метод Item-Item CF оказывается более эффективным с точки зрения времени формирования рекомендаций. При этом качество рекомендаций в среднем не хуже, чем в случае использования подхода, основанного на анализе пользовательских профилей. Для вычисления попарной близости информационных единиц могут использоваться те же метрики, что и в случае с парами «пользователь-объект» (часто используется косинусная или модифицированная косинусная меры). Для поиска рекомендаций на основании матрицы объектов часто используются весовые функции и методы регрессионного анализа. Одним из перспективных методов решения задачи Item-Item CF является метод Item2Vec [12].

Основным недостатком метода Item-Item CF при его практическом использовании является проблема формирования достоверного рейтинга информационных единиц (товаров). Такой рейтинг может быть сформирован при наличии большого количества пользователей, которые либо регулярно покупают различные товары (что непосредственно влияет на рейтинг), либо явным образом участвуют в формировании их рейтинга (например, выставляя им оценки или заполняя анкеты). Для небольших Интернет магазинов, имеющих ограниченное число посетителей, такой подход слабо применим. В то же время метод User-User CF с точки зрения качества рекомендаций сильно зависит от наличия данных о пользовательской активности, и проблема «холодного старта» для него стоит достаточно остро. В целом получается, что по отдельности методы Item-Item CF и User-User CF используют только часть собираемой рекомендательными системами информации о пользовательской активности и предпочтениях при выработке рекомендаций. Более эффективным может быть использование гибридных подходов. Так, например, в работе [13] предлагается алгоритм комбинированной фильтрации, основная идея которого заключается в получении оценки неизвестного рейтинга как взвешенной суммы оценок на основании фильтрации по транзакциям, фильтрации по товарам и смешенной фильтрации (на основании рейтингов похожих товаров в похожих транзакциях).

Авторами по итогам проведенных прикладных научных исследований предложен собственный метод повышения pertinентности информации (персонализации предложения), сочетающий сильные стороны обоих рассмотренных выше методов коллаборативной фильтрации и позволяющий более полно использовать доступную информацию о пользователях и информационных единицах (интернет-ресурсах), в первую очередь неявные данные о пользователе.

Суть предложенного метода заключается в совместном использовании методов Item-Item CF и User-User CF для формирования персонального информационного предложения. Для новых пользователей или пользователей с нерепрезентативной историей посещения предлагается генерировать рекомендации, базируясь на данных о подобию информационных единиц (метод Item-Item CF). Таким образом, решается проблема «холодного старта» и повышается качество рекомендаций для малоактивных пользователей, а также пользователей со слабо выраженными предпочтениями. По мере накопления данных о пользовательских предпочтениях и формировании его поведенческого профиля предпочтение предлагается отдавать оценкам, полученным с использованием метода User-User CF, исходя из гипотезы, что при наличии качественных поведенческих профилей пользователей данный метод позволяет более точно предсказывать их предпочтения. При этом при необходимости рекомендации могут дополняться предложениями информационных единиц, полученными на основе их рейтинга популярности.

В целях проверки результатов теоретических исследований составные части метода повышения pertinентности информации были применены к реальным данным действующего интернет-магазина Thaisoap. Магазин ориентирован на продажу натуральной тайской косметики и кокосового масла. Каталог товаров магазина содержит более 1 500 наименований товаров, разбитых на 180 классов (44 корневых класса, 136 подклассов). Ежедневно магазин посещают в среднем около 1 500 посетителей и проводят на нем (в среднем) порядка 11 минут каждый (на каждого посетителя приходится в среднем 28 переходов по ссылкам). Для исследований использовались тестовая выборка данных за одно полугодие (IV полугодие 2015 года), в котором каталог товаров был неизменен, а также статистические данные, полученные при помощи аналитической системы Яндекс.Метрика (<http://metrika.yandex.ru>). Исследования подтвердили адекватность и непротиворечивость результатов кластеризации информационных единиц и пользовательских профилей [14, 15].

Тем не менее указанные исследования не могли определить эффективность предложенного метода в силу отсутствия информации о воздействии на пользователя персонального информационного предложения, формируемого в результате применения метода.

Здесь необходимо ответить, что одним из наиболее распространённых подходов к определению эффективности методов, применяемых в рекомендательных системах интернет-магазинов, является оценка средней наполненности покупательской корзины до и после использования рекомендательной системы. Существующие исследования дают довольно большой разброс результатов измерений. Но сходятся все они в том, что в любых случаях наблюдается рост средней наполненности покупательской корзины, который может составлять 12-60% [16]. При этом наблюдается значительная дифференциация по типам товаров. Так для бытовой техники и электроники эффект использования рекомендательных систем, как правило, минимален. Для книг и спортивных товаров, напротив, эффект близок к максимальному. Непосредственно на эффективность влияет такой параметр как покрытие, определяющий насколько полно имеющийся ассортимент товаров и услуг охватывается при выработке рекомендаций. Косвенное влияние на эффективность оказывает качество работы рекомендательной системы. Чем более полезные рекомендации вырабатываются, тем больше доверие и лояльность пользователей интернет-магазину. Доверие пользователей является фактором, который может играть существенное значение в долгосрочной перспективе.

В силу отсутствия возможностей оценить метод в долгосрочном периоде было принято решение о проведении сплит-тестирования, т.е. организовать сравнение поведения пользователей интернет-магазина Thaisoap на базе четырёх сценариев: при отсутствии персонального информационного предложения, при наличии только предложений Item-Item CF, только User-User CF и при наличии персонального информационного предложения, полностью соответствующего разработанному авторами методу повышения пертинентности информации.

Исследования были проведены на основе экспериментального программного комплекса, реализующего каждый из четырёх сценариев. На тестирование каждого сценария был выделен месяц в период с февраля по май 2016 года. При этом каталог товаров оставался неизменным (т.е. не изменялась номенклатура товаров, предлагаемых посетителям магазина). Одним из условий проведения сравнительного анализа являлась работа «с чистого листа», т.е. в начале каждого месячного отрезка времени не использовались данные о пользовательских предпочтениях или рейтингах информационных единиц, накопленные за предыдущие периоды времени.

В первом сценарии (без персонального информационного предложения) было выявлено 1860 посетителей, из которых 140 человек совершили покупки (конверсия 7,5%). Средняя наполненность корзины при этом составила 1,5 товара. При навигации по сайту посетители совершали в среднем 36 кликов в рамках одной сессии и проводили на сайте в среднем порядка 16 минут.

Во втором сценарии (метод Item-Item CF) было выявлено 1908 посетителей, из которых покупки совершили 164 человека (конверсия 8,6%). Средняя наполненность корзины при этом составила 2,1 товара. При навигации по сайту посетители совершали в среднем 31 клик в рамках одной сессии и проводили на сайте в среднем около 12 минут.

В третьем сценарии (метод User-User CF) было выявлено 1873 посетителя, из которых 145 человек совершил покупки (конверсия 7,7%). Средняя наполненность корзины при этом составила 1,9 товара. При навигации по сайту посетители совершали в среднем 32 клика в рамках одной сессии и проводили на сайте в среднем около 14 минут.

В последнем сценарии (метод повышения пертинентности информации) было выявлено 2011 посетителей, из которых 183 человека совершили покупки (конверсия 9,1%). Средняя наполненность корзины при этом составила 2,4 товара. При навигации по сайту посетители совершали в среднем 28 кликов в рамках одной сессии и проводили на сайте в среднем около 11 минут.

Анализ полученных результатов исследования позволяет сделать следующие основные выводы:

Использование рекомендательной системы в рассматриваемых сценариях позволило повысить среднюю наполненность корзины и сократить время, затрачиваемое посетителями на поиск интересующих их товаров. Так среднее время пребывания посетителей на сайте сокращалось в различных сценариях от 16 до 14-11 минут. При этом в сценариях с использованием рекомендательной системы увеличивается средняя наполненность корзины, что напрямую влияет на среднюю величину чека и непосредственные доходы магазина.

Метод User-User CF показал себя малоэффективным на рассматриваемом интервале времени (один месяц). Основной причиной этого является низкое качество рекомендаций, связанное с недостаточностью данных о пользовательских предпочтениях.

Метод Item-Item CF показал себя достаточно эффективным за счет успешного решения проблемы «холодного старта». По сравнению с первым сценарием конверсия увеличилась более чем на 1%, средняя наполненность корзины увеличилась с 1,5 до 2,1.

Рекомендательная система на базе комбинированного подхода позволила получить наибольший прирост рассматриваемых показателей. Так конверсия выросла более чем на 1,5% по сравнению с первым сценарием, а средняя наполненность корзины выросла с 1,5 до 2,4 (т.е. более чем на 35%).

В целом необходимо отметить, что эффективность рекомендательной системы сильно зависит как от типов продаваемых товаров, так и от особенностей используемых алгоритмов и качества реализации рекомендательной системы. В определенных ситуациях рекомендательная система может не давать заметное повышение показателей конверсии и средней наполненности товарной корзины. Тем не менее, она остается полезной, так как предлагает более удобный и простой для посетителей способ поиска интересующих их товаров, тем самым создавая удобное окружение и повышая лояльность посетителей интернет-ресурсу, что уже в среднесрочной перспективе может обеспечить возвраты пользователей, формирование пула постоянных покупателей, обеспечивающих стабильный рост выручки интернет-магазина.

Литература

1. Walker L. Amazon Gets Personal With E-Commerce // Washington Post Staff Writer Sunday, November 8, 1998; Page H1. URL: <http://www.washingtonpost.com/wp-srv/washtech/daily/nov98/amazon110898.htm>
2. С.А.Филиппов, В.Н.Захаров, С.А.Ступников, Д.Ю.Ковалев Подходы к повышению pertinентности информационного предложения в медиасервисах на основе обработки больших объемов данных // Ceur workshop proceedings, Vol-1536, Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015) Obninsk, Russia, October 13-16, 2015, p. 114-118.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining //СПб.: БХВ-Петербург, 2004. — 336 с.
4. Лекция: Методы и стадии Data Mining // Статья в сети Интернет, URL: <http://www.intuit.ru/studies/courses/6/6/lecture/162?page=1>.
5. С.А.Филиппов, В.Н.Захаров, С.А.Ступников, Д.Ю.Ковалев Организация больших объемов данных в рекомендательных системах поддержки жизнеобеспечения, входящих в состав глобальных платформ электронной коммерции // Ceur workshop proceedings, Vol-1536, Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015) Obninsk, Russia, October 13-16, 2015, p. 119-124.
6. М. Тим Джонс Рекомендательные системы: Часть 1. Введение в подходы и алгоритмы // Библиотека IBM, 2013. URL: <http://www.ibm.com/developerworks/ru/library/os-recommender1/>.
7. Xiao Yuan Su, Taghi M. Khoshgoftaar A survey of collaborative filtering techniques // Advances in Artificial Intelligence, Volume 2009 (2009), Article ID 421425, 19p.
8. Fleder D., Hosanagar K. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity // Management Science, Vol. 55, No. 5, May 2009, pp. 697-712.
9. Брейкин Е. А. Рекомендательная система на основе коллаборативной фильтрации // Молодой ученый. — 2015. — №13. — С. 31-33.
10. Марманис Х., Бабенко Д. Алгоритмы интеллектуального Интернета // СПб.-М.: Символ, 2011. – 466 с.
11. Greg Linden, Brent Smith and Jeremy York Amazon.com recommendations: Item-to-Item Collaborative Filtering // Industry Report, IEEE INTERNET COMPUTING, 2003.
12. Barkan O., Koenigstein N. Item2Vec: Neural Item Embedding for Collaborative Filtering // arXiv preprint arXiv:1603.04259, Mar 2016.
13. Гончаров Максим Системы выработки рекомендаций // Статья в сети Интернет. URL: <http://www.businessdataanalytics.ru/RecommendationSystems.htm>.
14. Филиппов С.А., Захаров В.Н., Ступников С.А., Ковалев Д.Ю. Метод определения подобия информационных единиц по неявным пользовательским предпочтениям в рекомендательных системах поддержки жизнеобеспечения // Аналитика и управление данными в областях с интенсивным использованием данных: XVIII Международная конференция DAMDID/RCDL'2016 (Ершово, Московская обл., 11 – 14 октября 2016 года, Россия): Труды конференции / Под. Ред. Л.А. Калиниченко, Я. Манолопулоса, С.О. Кузнецова. – М. ФИЦ ИУ РАН, 2016, с. 169 – 174. ISBN 978-5-94558-206-5.
15. Филиппов С.А., Захаров В.Н., Ступников С.А., Ковалев Д.Ю. Кластеризация профилей пользователей в рекомендательных системах поддержки жизнеобеспечения на основе реальных неявных данных // Аналитика и управление данными в областях с интенсивным использованием данных: XVIII Международная конференция DAMDID/RCDL'2016 (Ершово, Московская обл., 11 – 14 октября 2016 года, Россия): Труды конференции / Под. Ред. Л.А. Калиниченко, Я. Манолопулоса, С.О. Кузнецова. – М. ФИЦ ИУ РАН, 2016, с. 163 – 168. ISBN 978-5-94558-206-5.
16. Валерий Дьяченко Сервисы рекомендаций: как с их помощью увеличить продажи на 60% // Статья в сети Интернет, URL: <http://www.kom-dir.ru/article/51-servisy-rekomendatsiy>.

References

1. Walker L. Amazon Gets Personal With E-Commerce // Washington Post Staff Writer Sunday, November 8, 1998; Page H1. URL: <http://www.washingtonpost.com/wp-srv/washtech/daily/nov98/amazon110898.htm>
2. S. Philippov, V. Zakharov, S. Stupnikov, D. Kovalev, 2015. Approaches to Improve the Pertinence of Information in the Media Services on the Basis of Big Data Processing. Ceur workshop proceedings, Vol-1536 [Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015)] Obninsk, Russia, October 13-16, 2015, pp. 114-118.
3. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Metody i modeli analiza dannykh: OLAP i Data Mining //SPb.: BKhV-Peterburg, 2004. — 336 p.
4. Lektsiya: Metody i stadii Data Mining . URL: <http://www.intuit.ru/studies/courses/6/6/lecture/162?page=1>.

5. S. Philippov, V. Zakharov, S. Stupnikov, D. Kovalev, 2015. Organization of Big Data in the Global E-commerce Platforms Ceur workshop proceedings, Vol-1536 [Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015)] Obninsk, Russia, October 13-16, 2015, pp. 119-124.
6. M. Tim Dzhons Rekomendatel'nye sistemy: Chast' 1. Vvedenie v podkhody i algoritmy. URL: <http://www.ibm.com/developerworks/ru/library/os-recommender1/>.
7. Xiaoyuan Su, Taghi M. Khoshgoftaar A survey of collaborative filtering techniques // Advances in Artificial Intelligence, Volume 2009 (2009), Article ID 421425, 19p.
8. Fleder D., Hosanagar K. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity // Management Science, Vol. 55, No. 5, May 2009, pp. 697-712.
9. Breykin E. A. Rekomendatel'naya sistema na osnove kollaborativnoy fil'tratsii // Molodoy uchenyy. — 2015. — №13. — p. 31-33.
10. Marmanis Kh., Babenko D. Algoritmy intellektual'nogo Interneta // SPb.-M.: Simvol, 2011. – 466 p.
11. Greg Linden, Brent Smith and Jeremy York Amazon.com recommendations: Item-to-Item Collaborative Filtering // Industry Report, IEEE INTERNET COMPUTING, 2003.
12. Barkan O., Koenigstein N. Item2Vec: Neural Item Embedding for Collaborative Filtering // arXiv preprint arXiv:1603.04259, Mar 2016.
13. Goncharov M., 2010. Sistemy vyrabotki rekomendatsiy. URL: <http://www.businessdataanalytics.ru/RecommendationSystems.htm>.
14. Philippov S.A., Zakharov V.N., Stupnikov S.A., Kovalev D.Yu. Metod opredeleniya podobiya informatsionnykh edinit po neyavnyim pol'zovatel'skim predpochteniyam v rekomendatel'nykh sistemakh podderzhki zhizneobespecheniya // Analitika i upravlenie dannymi v oblastiakh s intensivnym ispol'zovaniem dannykh: XVIII Mezhdunarodnaya konferentsiya DAMDID/RCDL'2016 (Ershovo, Moskovskaya obl., 11 – 14 oktyabrya 2016 goda, Rossiya): Trudy konferentsii / Pod. Red. L.A. Kalinichenko, Ya. Manolopulosa, S.O. Kuznetsova. – M. FITs IU RAN, 2016, s. 169 – 174. ISBN 978-5-94558-206-5.
15. Philippov S.A., Zakharov V.N., Stupnikov S.A., Kovalev D.Yu. Klasterizatsiya profiley pol'zovateley v rekomendatel'nykh sistemakh podderzhki zhizneobespecheniya na osnove real'nykh neyavnykh dannykh // Analitika i upravlenie dannymi v oblastiakh s intensivnym ispol'zovaniem dannykh: XVIII Mezhdunarodnaya konferentsiya DAMDID/RCDL'2016 (Ershovo, Moskovskaya obl., 11 – 14 oktyabrya 2016 goda, Rossiya): Trudy konferentsii / Pod. Red. L.A. Kalinichenko, Ya. Manolopulosa, S.O. Kuznetsova. – M. FITs IU RAN, 2016, s. 163 – 168. ISBN 978-5-94558-206-5.
16. D'yachenko V, 2016. Servisy rekomendatsiy: kak s ikh pomoshch'yu uvelichit' prodazhi na 60%. URL: <http://www.kom-dir.ru/article/51-servisy-rekomendatsiy>.

Поступила 21.10.2016

Об авторах:

Филиппов Станислав Александрович, старший научный сотрудник лаборатории Института проблем информатики Федерального исследовательского центра "Информатика и Управление" Российской академии наук, кандидат технических наук, stanislav@philippov.ru;

Захаров Виктор Николаевич, учёный секретарь Федерального исследовательского центра "Информатика и Управление" Российской академии наук, доктор технических наук, VZakharov@ipiran.ru.