

*ИССЛЕДОВАНИЕ РАБОТЫ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ***М.В. ЯНАЕВА, Е.В. СИНЧЕНКО**

*Кубанский государственный технологический университет,
350072, г. Краснодар, Российская Федерация, ул. Московская, 2;
электронная почта: yanaevam@mail.ru*

В статье рассматриваются проектирование и разработка интеллектуальной системы в структуре системы удалённого управления в территориально распределённых комплексах на основе использования рекомендательных систем. В качестве рекомендательной системы разработана интеллектуальная система управления, основное предназначение которой заключается в мониторинге и удалённом управлении.

Ключевые слова: интеллектуальная система (ИС), рекомендательные системы, мониторинг.

Интеллектуальная система (ИС) является системой оперативного анализа ситуационной информации и поддержки принятия решения, основанная на применении методов интеллектуального анализа данных. Такие системы предназначены для мониторинга и анализа ситуаций на основе сформированной информационной модели предметной области. В их задачи входит: управление ситуационными данными, формирование и расчет показателей для оценки обстановки, поддержка принятия решений, отображение и формирование аналитических отчетов. Поддержка принятия решений основана на специальном статистическом анализе потоков событий, формировании выборок статистических данных по складывающейся обстановке и проведение многомерного статистического анализа с применением различных технологий; выделении угроз и опасных пространственно-временных ситуаций в складывающейся обстановке; создании динамических моделей для прогнозирования изменения обстановки, основанных на процедурах машинного обучения[1,4].

Интеллектуальная система обеспечивают управление системой с графических планов и мониторинг событий. Данный модуль позволяет управлять всеми элементами системы с графических планов объекта или с помощью дерева устройств. ИС удалённого управления как системный продукт <http://ntk.kubstu.ru/file/1408>

обладает рядом возможностей, а именно: создание территориально-распределенных систем с открытой архитектурой, высокая наглядность демонстрации событий в реальном времени на планах объекта, программирование различных ответных реакций на события, выдача рекомендаций для действий охраны, повышение надежности системы защиты, событийная обработка данных и т.д [2,3].

Существует три основных подхода в рекомендательных системах:

- 1) коллаборативная фильтрация (collaborative filtering);
- 2) контентная фильтрация (content-based filtering);
- 3) гибридный.

По модели прошлых действий пользователя коллаборативная фильтрация производит рекомендации. Модель строится как по поведению текущего пользователя так и по поведению схожих между собой пользователей. Принимая во внимание, что есть пользователи, подобные текущему, коллаборативная фильтрация использует информацию об их поведении, вырабатывая рекомендации. Рекомендации строятся по принципу связности пользователей и фильтрации тех, у которых наблюдается похожее поведение.

Для примера представим, что создаётся сайт, который будет рекомендовать посетителям различные блоги. Необходимо сгруппировать пользователей на основе их предпочтений, используя информацию о подписке или чтении того или иного блога. В одну группу можно отнести пользователей, посещающих одни и те же блоги. Далее идентифицируются наиболее популярные блоги читателей выбранной группы. Следующим шагом происходит рекомендация для конкретного пользователя самого востребованного блога, который он пока не читал.

Перейдём к таблице 1. Пересечение столбца и строки информирует о том, сколько статей прочитано данным пользователем в указанном блоге. Алгоритм *ближайшего* соседа даёт возможность определить два кластера, которые разделяют пользователей на две группы, основываясь на чтении блогов. Первый кластер содержит пользователей Петя и Федя, которые прочли

определённое количество статей в тематике «Windows» и «Веб-программирование». Второй кластер объединяет Ваню и Гришу, которые читали статьи из "C#" и "C++".

Т а б л и ц а 1 –Коллаборативная фильтрация

Имена Блоги	Петя	Ваня	Федя	Гриша
Windows	13	3	11	-
Робототехника	10	-	-	3
Веб-программирование	6	1	9	-
C#	-	6	-	9
C++	-	7	1	8
Статьи, прочитанные пользователем				
Кластер	1	2	1	2

Отсюда есть возможность для выделения различий в каждом кластере и формирования определённых рекомендаций. В первом кластере Петя предпочел 10 статей из блога «Робототехника», но Федя не читал подобных статей; Федя напротив выбрал статью из темы «C++», в отличии от Пети. Основываясь на таблице 1 можно смело рекомендовать блог «Робототехника» для Феди, а Пете не останется ничего делать, как выбирать самому блог для чтения, так как малая разница по чтению блога «C++» будут отфильтрована. Во втором кластере Гриша предпочёл 3 статьи из блога «Робототехника», а Ваня не читал таких статей; Ваня выбрал 3 статьи из «Windows», но Гриша напротив не выбирал из этого блога ничего. Отсюда, получаются следующие рекомендации: Грише предлагается к прочтению блог «Windows», а Ване – «Робототехника».

Диаграмма Венна, показанная на рисунке 1, рассматривает отношения на основе сходств и различий. Используя соответствующий алгоритм, определяются сходства, как необходимо объединить пользователей со схожими предпочтениями. Например, применяя фильтрацию по востребованности, можно определить различия, используемые для построения рекомендаций [5].

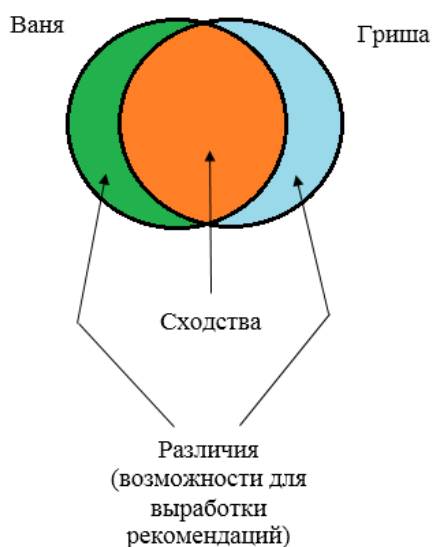


Рисунок 2 – Сходства и различия

Очень удобно использовать такое представление, как на рисунке 2, хоть оно и упрощено из-за использования малого количества образцов.

Основываясь на поведении пользователей *Контентная фильтрация* выработывает рекомендации. Этот метод предполагает использование прошлой информации о чтении блогов и их характеристик. Предположим, что пользователь предпочитает статьи о Windows или же любит ставить комментарии в статьях о Веб-программировании, тогда контентная фильтрация, использует накопленную информацию для нахождения похожих блогов и выдачи рекомендаций для пользователя (статьи из Windows или же Веб-программирования).

Перейдём снова к рассмотрению пользователя Федя из таблицы 1. Представим, что необходимо ранжировать блоги по следующему принципу: те, кто интересовался статьями по Windows, проявят интерес к Веб-программированию и робототехнике. Тогда очень легко можно предложить Феде, используя историю его чтения, блог Робототехника. Метод, показанный в таблице 2, использует информацию, связанную только с одним пользователем, не привлекая сведения о других пользователях.

Т а б л и ц а 2 – Ранжированный список различий

Имена	Федя	Сходный контент
Блоги		
Windows	11	Windows ←
Робототехника	-	Робототехника
Веб-программирование	9	Веб-программирование
C#	-	Проранжированные блоги
C++	1	

Здесь также можно использовать диаграмму Венна (рисунок 2): если левая сторона — это Федя, а правая — проранжированные блоги, то не учитываются сходства, потому что Федя уже читал эти блоги, а проранжированные различия блогов дают возможность представить рекомендацию.

Повысить эффективность и сложность рекомендательных систем позволяют *гибридные подходы*, объединяющие контентную и коллаборативную фильтрацию. Методы, которые были представлены в таблицах 1 и 2, могут быть использованы в гибридной системе. Точность рекомендации значительно повышается при связывании коллаборативной и контентной фильтрации. Полезность гибридного подхода также наблюдается если фильтрация происходит при малом количестве исходных данных, то есть наблюдается, так называемый, *холодный старт*. Гибридный подход сперва взвешивает результаты контентной фильтрации, далее смещая веса в сторону коллаборативной фильтрации (по характеру определённости контента для текущего пользователя).

Подход, который был использован в конкурсе Netflix, показал, что в рекомендательные системы можно привлекать множество разных алгоритмов. Конечные данные, в зависимости от проблем и отношений, заданных для алгоритма, могут разниться. Корни большинства алгоритмов произрастают из области машинного обучения, являющейся подобластью искусственного

интеллекта, в которой алгоритмы предназначены для прогнозирования, обучения и принятия решений.

С помощью *корреляции Пирсона* возможно чёткое вычисление сходств между двумя читателями блогов. Данный алгоритм определяет линейную зависимость между двумя переменными на основе их атрибутов. Но он не сможет вычислить зависимость по всем пользователям. Всех пользователей вначале нужно отфильтровать, основываясь на сходстве, как, например, предпочтение блогов [6].

В коллаборативной фильтрации корреляция Пирсона – это один из самых востребованных алгоритмов.

В случайных данных дают возможность определить структуру *алгоритмы кластеризации*. Эти алгоритмы основаны на определении сходства элементов путём просчитывания удалённости элементов друг от друга в признаковом пространстве (на примере таблиц признаков – количество статей). На основе объёма пространства признаков выявляются независимые признаки. Например, если читатели схожи между собой, то они соединяются в единый кластер.

На данный момент распространено большое количество алгоритмов кластеризации. Например, алгоритм *k-средних*, разбивающий переменные на *k* кластеров, является одним из простейших. Сперва переменные расходятся по кластерам случайно. Далее высчитывается *центр масс* кластера. Следующим шагом происходит проверка удалённости каждой переменной от центра кластера. После проверки происходит перемещение переменных к более подходящим кластерам. После повторения предыдущего шага для кластеров центры просчитываются снова. Если при повторе всех предыдущих действий переменные больше не перемещаются, то алгоритм подходит к концу, а данные являются кластеризованными.

Вычисление удалённости одного объекта от другого является нетривиальной задачей при визуализировании. Самый часто встречающийся подход к решению этой проблемы заключается в рассмотрении каждой

переменной кластера в виде многомерного вектора и определении для каждого евклидового расстояния.

Встречается много разных видов кластеризации, таких как:

- a) Теория адаптивного резонанса;
- b) Нечеткая кластеризация методом С-средних;
- c) Вероятностная кластеризация с помощью EM-алгоритма.

Далее перечислю некоторые успешные алгоритмы, используемые в рекомендательных системах:

– Классификация методом Роккио — для улучшения точности основывается на релевантности переменных.

– Байесовские сети доверия – можно представить в виде ориентированного ациклического графа, в котором ребра - это объединяемые вероятности элементов.

– Цепи Маркова – похожи на предыдущий алгоритм за исключением того, что они определяют создание рекомендаций в виде оптимизации, которая двигается последовательно, шаг за шагом.

Отсюда важно при создании территориально-распределённых систем учесть огромное число данных. Они нужны для выдачи рекомендаций на основе мониторинга.

В работе проектируемой интеллектуальной системе будут реализованы, отслежены и программно обработаны следующие типы действий [7,8]:

- 1) Камера, взаимодействующая с датчиком движения.
- 2) Контроль света (выкл и вкл по голосу, автоотключение, если помещение закрыто).
- 3) Контроль температуры (взаимодействие с вентиляционным, отопительными системами).
- 4) Мониторинг внешней среды.
- 5) Управление дверями и входом/выходом.
- 6) Система оповещения об утечках, задымлении.
- 7) Управление техникой.

Все перечисленные виды взаимодействия с интеллектуальной системой требуют мониторинга данных, а также его анализа для принятия решений. Наиболее подходящими алгоритмами для реализации рекомендательных воздействий являются гибридный подход и цепи Маркова. При использовании гибридного подхода есть перспективы, например создание такого набора команд, чтобы после того, как вы ложились спать, температура медленно понижалась (спать в глубокой фазе комфортнее в прохладной комнате), а после пробуждения (сигнала будильника), наоборот, повышалась (просыпаться приятнее в комнате более теплой).

Касательно цепей Маркова – они будут выполнять последовательную оптимизацию данных, для выдачи вывода о том или ином управляющем воздействии на двери, технику и т.п. Например, если на улице похолодало, а двери или окна открыты, то необходимо проанализировать температуру внешней среды, внутренней, построить прогноз дальнейшей температуры и принять соответствующее решение:

- a) закрыть дверь и окно;
- b) закрыть только дверь или окно;
- c) включить отопление после закрытия двери и окна;
- d) ничего не делать.

Подобная разработка будет весьма интересна в структуре территориально- распределённых комплексов и позволит решить круг задач удаленного управления [9].

ЛИТЕРАТУРА

1. Цыгикало Т.И., Янаева М.В., Цыгикало Д.В., Руденко М.В., Автоматизация процесса управления экологическим мониторингом строительной площадки // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар КубГАУ , 2012 . - №77. Режим доступа: <http://ej.kubagro.ru/2012/03/pdf/70.pdf>.

2. Янаева М.В., Мурлин А.Г., Мурлина В.А., Системы экологического мониторинга в строительных организациях // Научный журнал КубГАУ <http://ntk.kubstu.ru/file/1408>

[Электронный ресурс]. – Краснодар КубГАУ, 2012. - №84. Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/35.pdf>

3. Янаева М.В., Мурлин А.Г., Мурлина В.А., Методы прогнозирования в информационной системе экологического мониторинга // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар КубГАУ, 2012. - №84. Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/34.pdf>

4. Янаева М.В., Мурлин А.Г., Мурлина В.А., Управление эффективностью пространственно распределённых промышленных предприятий // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар КубГАУ, 2014. - №102(08). Режим доступа: <http://ej.kubagro.ru/2014/08/pdf/65.pdf>

5. Видовский Л.А., Янаева М.В., Мурлин А.Г., Мурлина В.А. Стратегическое управление территориально – распределёнными комплексами // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар КубГАУ, 2015. - №112. Режим доступа: <http://ej.kubagro.ru/2015/08/pdf/86.pdf>

6. Янаева М.В. Интеллектуальная система комплексного мониторинга экологической обстановки стройплощадки // Монография, Краснодар, 2012.

7. Янаева М.В., Адамова А.В. Разработка информационной подсистемы стратегического управления территориально – распределённым строительным комплексом // Научные труды Кубанского государственного технологического университета. [Электронный ресурс]. – Научные труды КубГТУ, №10, 2015 г. Режим доступа: <http://ntk.kubstu.ru/file/574>

8. Янаева М.В., Коломбет А.В. Разработка информационной системы управления и мониторинга территориально – распределёнными комплексами // Научные труды Кубанского государственного технологического университета. [Электронный ресурс]. – Научные труды КубГТУ, №10, 2015 г. Режим доступа: <http://ntk.kubstu.ru/file/601>

9. Янаева М.В., Аваков Д.В. Подсистема удаленного обмена информацией на основе использования облачных технологий для информационной системы управления территориально – распределённым

строительным комплексом // Научные труды Кубанского государственного технологического университета. [Электронный ресурс]. – Научные труды КубГТУ, №10, 2015 г. Режим доступа: <http://ntk.kubstu.ru/file/602>

REFERENCES

1. Tsygikalo T.I., Yanaeva M.V., Tsygikalo D.V., Rudenko M.V., Avtomatizatsiya protsessa upravleniya ekologicheskim monitoringom stroitelnoy ploshchadki // Nauchnyy zhurnal KubGAU [Elektronnyy resurs]. – Krasnodar KubGAU , 2012 . - №77. Rezhim dostupa: <http://ej.kubagro.ru/2012/03/pdf/70.pdf>.

2. Yanaeva M.V., Murlin A.G., Murlina V.A., Sistemy ekologicheskogo monitoringa v stroitelnykh organizatsiyakh // Nauchnyy zhurnal KubGAU [Elektronnyy resurs]. – Krasnodar KubGAU , 2012 . - №84. Rezhim dostupa: <http://ej.kubagro.ru/2012/10/pdf/35.pdf>

3. Yanaeva M.V., Murlin A.G., Murlina V.A., Metody prognozirovaniya v informatsionnoy sisteme ekologicheskogo monitoringa // Nauchnyy zhurnal KubGAU [Elektronnyy resurs]. – Krasnodar KubGAU , 2012 . - №84. Rezhim dostupa: <http://ej.kubagro.ru/2012/10/pdf/34.pdf>

4. Yanaeva M.V., Murlin A.G., Murlina V.A., Upravlenie effektivnostyu prostranstvenno raspredelennykh promyshlennykh predpriyatiy // Nauchnyy zhurnal KubGAU [Elektronnyy resurs]. – Krasnodar KubGAU, 2014. - №102(08). Rezhim dostupa: <http://ej.kubagro.ru/2014/08/pdf/65.pdf>

5. Vidovskiy L.A., Yanaeva M.V., Murlin A.G., Murlina V.A. Strategicheskoe upravlenie territorialno – raspredelennymi kompleksami // Nauchnyy zhurnal KubGAU [Elektronnyy resurs]. – Krasnodar KubGAU, 2015. - №112. Rezhim dostupa: <http://ej.kubagro.ru/2015/08/pdf/86.pdf>

6. Yanaeva M.V. Intellektualnaya sistema kompleksnogo monitoringa ekologicheskoy obstanovki stroyploshchadki // Monografiya, Krasnodar, 2012.

7. Yanaeva M.V., Adamova A.V. Razrabotka informatsionnoy podsistemy strategicheskogo upravleniya territorialno – raspredelennym stroitelnyim kompleksom // Nauchnye trudy Kubanskogo gosudarstvennogo tekhnologicheskogo universiteta.

[Elektronnyy resurs]. – Nauchnye trudy KubGTU, №10, 2015 g. Rezhim dostupa: <http://ntk.kubstu.ru/file/574>

8. Yanaeva M.V., Kolombet A.V. Razrabotka informatsionnoy sistemy upravleniya i monitoringa territorialno – raspredelennymi kompleksami // Nauchnye trudy Kubanskogo gosudarstvennogo tekhnologicheskogo universiteta. [Elektronnyy resurs]. – Nauchnye trudy KubGTU, №10, 2015 g. Rezhim dostupa: <http://ntk.kubstu.ru/file/601>

9. Yanaeva M.V., Avakov D.V. Podсистема udalennogo obmena informatsiey na osnove ispolzovaniya oblachnykh tekhnologiy dlya informatsionnoy sistemy upravleniya territorialno – raspredelennym stroitelnyim kompleksom // Nauchnye trudy Kubanskogo gosudarstvennogo tekhnologicheskogo universiteta. [Elektronnyy resurs]. – Nauchnye trudy KubGTU, №10, 2015 g. Rezhim dostupa: <http://ntk.kubstu.ru/file/602>

RESEARCH OF OPERATION OF RECOMMENDATORY SYSTEMS

M.V. YANAIEVA, E.V. SINCHENKO

*Kuban State Technological University,
2, Moskovskaya st., Krasnodar, Russian Federation, 350072,
e-mail: yanaevam@mail.ru*

The article deals with the design and development of intelligent system in the structure of the system remote control complexes in geographically distributed through the use of reference systems. As an advisory system intelligent control system will be developed, the main purpose of which will be to remote monitoring and management.

Key words: intellectual system, recommendatory systems, monitoring.