

УДК 004:83

DOI: 10.15827/0236-235X.120.625-631

Дата подачи статьи: 28.06.17

2017. Т. 30. № 4. С. 625–631

ПРИМЕНЕНИЕ МЕТОДОВ КЛАССИФИКАЦИИ И КЛАСТЕРИЗАЦИИ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ РАБОТЫ ПРЕЦЕДЕНТНЫХ СИСТЕМ

П.Р. Варшавский, к.т.н., доцент, VarshavskyPR@mpei.ru;

Ар Кар Мьо, аспирант, arkar2011@gmail.com

*(Национальный исследовательский университет «Московский энергетический институт»,
ул. Красноказарменная, 14, г. Москва, 111250, Россия);*

Д.В. Шункевич, ассистент, shunkevichdv@gmail.com

*(Белорусский государственный университет информатики и радиоэлектроники (БГУИР),
ул. П. Бровки, 6, г. Минск, 220013, Республика Беларусь)*

В статье рассматриваются актуальные вопросы повышения эффективности работы систем, использующих рассуждения на основе прецедентов (CBR – Case-Based Reasoning). Прецедентные методы и системы (CBR-системы) активно применяются для решения целого ряда задач в области искусственного интеллекта (например, для моделирования правдоподобных рассуждений (рассуждений здравого смысла), машинного обучения, интеллектуальной поддержки принятия решений, интеллектуального поиска информации, интеллектуального анализа данных и др.). Следует отметить, что современные инструменты для интеллектуального анализа данных, широко используемые сегодня в интеллектуальных системах, системах управления базами данных и знаний, бизнес-приложениях, системах машинного обучения, системах электронного документооборота и др., не обладают развитыми CBR-средствами.

Для повышения эффективности работы CBR-систем предлагается использовать модифицированный CBR-цикл, позволяющий сформировать базу удачных и неудачных прецедентов на основе имеющейся экспертной информации (тестовых выборок), а также модификацию алгоритма *k* ближайших соседей (*k*-NN) для извлечения прецедентов. Предложенные модификации позволяют повысить качество решения задач интеллектуального анализа данных (в частности, задачи классификации данных). Кроме того, в работе для повышения быстродействия CBR-систем рассматривается возможность сокращения количества прецедентов в базе удачных прецедентов за счет применения методов классификации и кластеризации.

С использованием разработанного в среде MS Visual Studio на языке C# прототипа CBR-системы проведены вычислительные эксперименты по оценке эффективности предлагаемых в работе решений на наборе данных из UCI Machine Learning Repository.

Ключевые слова: интеллектуальная система, прецедентный подход, интеллектуальный анализ данных, классификация, кластеризация.

В настоящее время актуальной задачей в области *искусственного интеллекта* (ИИ) является разработка методов *интеллектуального анализа данных* (ИАД) и соответствующих программных средств [1–3]. Методы ИАД активно применяются в *интеллектуальных системах* (ИС), *системах управления БД* (СУБД) и *системах управления базами знаний* (СУБЗ), бизнес-приложениях, системах машинного обучения, системах электронного документооборота и др. В ИАД для извлечения новых знаний из имеющихся данных применяются статистические и индуктивные методы, генетические алгоритмы, *искусственные нейронные сети* (ИНС), кластерный анализ, методы правдоподобных рассуждений на основе прецедентов (CBR – Case-Based Reasoning) и т.п. [2, 3].

Следует отметить, что в современных средствах ИАД для СУБД и СУБЗ, в том числе ориентированных на платформу *Business Intelligence* (BI) [4], практически отсутствуют развитые средства и инструменты для анализа данных на основе прецедентов. По этой причине остро стоит вопрос разработки эффективных CBR-средств для расширения инструментов ИАД на платформе BI для современных СУБД и СУБЗ.

Прецедентный подход

Прецедентный подход базируется на понятии прецедента, определяемого как случай, имевший место ранее и служащий примером или оправданием для последующих случаев подобного рода, и довольно простом принципе, что подобные задачи имеют подобное решение.

В общем случае модель представления прецедента включает описание ситуации, решение для данной ситуации и результат применения решения: $CASE = (Situation, Solution, Result)$, где *Situation* – ситуация, описывающая данный прецедент; *Solution* – решение (например диагноз и рекомендация); *Result* – результат применения решения, который может включать список выполненных действий, дополнительные комментарии и ссылки на другие прецеденты, а также в некоторых случаях обоснование выбора данного решения и возможные альтернативы. В большинстве случаев для представления прецедентов используется простое параметрическое представление [5].

Как правило, CBR-методы основываются на так называемом CBR-цикле [6], включающем в себя четыре основных этапа:

- извлечение наиболее соответствующего (подобного) прецедента (или прецедентов) для сложившейся ситуации из *базы прецедентов* (БП);
- повторное использование извлеченного прецедента для попытки решения текущей проблемы;
- адаптация и применение полученного решения для решения текущей проблемы;
- сохранение вновь принятого решения как части нового прецедента.

Для извлечения прецедентов из БП системы могут применяться различные методы [5]:

- метод ближайшего соседа (NN – Nearest Neighbor) и его модификации (например метод *k* ближайших соседей (*k*-NN));
- метод поиска на деревьях решений;
- метод извлечения на основе знаний;
- метод извлечения с учетом применимости прецедентов и др.

В данной работе используются параметрическое представление прецедентов и модификация алгоритма *k*-NN.

Модификация алгоритма *k*-NN для извлечения прецедентов

Самым распространенным методом сравнения и извлечения прецедентов является метод ближайшего соседа [1, 5]. Он позволяет довольно легко вычислить степень сходства текущей проблемной ситуации и прецедентов из БП. Для определения степени сходства на множестве параметров, используемых для описания прецедентов и текущей ситуации, вводится некоторая метрика. Далее в соответствии с выбранной метрикой определяется расстояние от целевой точки, соответствующей текущей проблемной ситуации, до точек, представляющих прецеденты из БП, и выбирается ближайшая к целевой точка.

Метод ближайшего соседа широко применяется для решения задач классификации, кластеризации, регрессии и распознавания образов. Обычно решение выбирается на основе нескольких ближайших точек (соседей), а не одной (метод *k*-NN) [7]. Основными преимуществами указанного метода являются простота реализации и универсальность в смысле независимости от специфики конкретной проблемной области, а к недостаткам можно отнести сложность выбора метрики для определения степени сходства, прямую зависимость требуемых вычислительных ресурсов от размера БП и низкую эффективность при работе с неполными и плохо определенными (так называемыми «зашумленными») исходными данными.

В работе предложена модификация алгоритма *k*-NN, которая заключается в том, что *k* изменяется в зависимости от изменения размера БП. Чем больше прецедентов в БП, тем большее значение можно выбрать для *k* (от 1 до *k*_{max}). *k*_{max} соответствует максимальному количеству прецедентов из

БП, имеющих одинаковое решение (то есть принадлежащих одному классу или кластеру). В результате вычислительных экспериментов было установлено, что целесообразно выбирать *k* как ближайшее целое число к среднему арифметическому значению между 1 и *k*_{min}, то есть $k_{avg} = (1 + k_{min})/2$, где *k*_{min} – минимальное количество прецедентов, имеющих одинаковое решение (то есть принадлежащих одному классу или кластеру).

Работа модифицированного алгоритма *k*-NN для извлечения прецедентов была рассмотрена на примере решения одной из задач ИАД – задачи классификации данных (рис. 1) с использованием

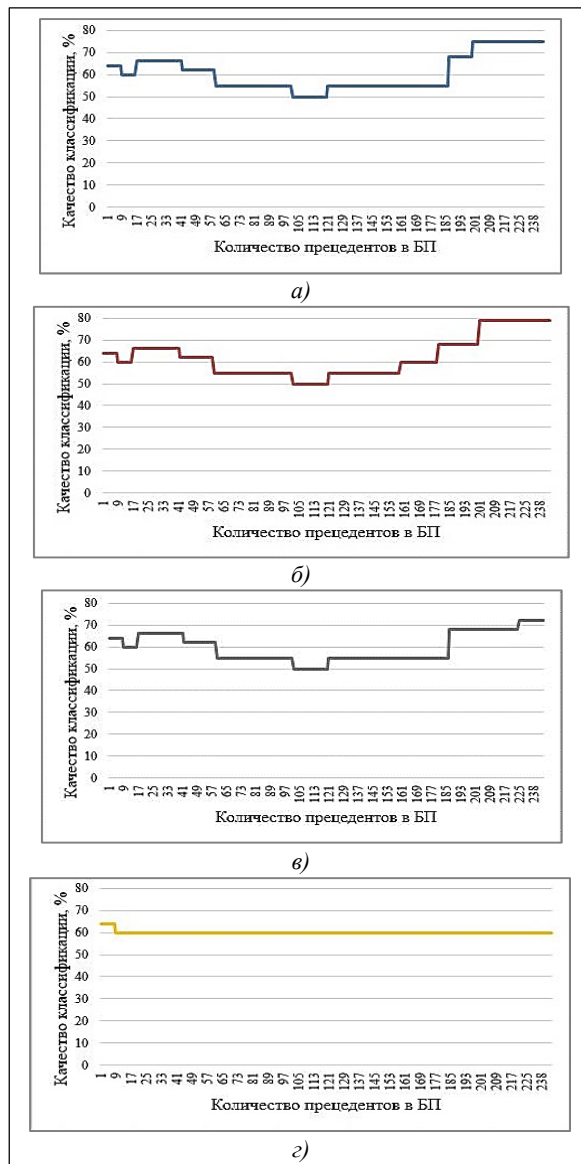


Рис. 1. График зависимости качества классификации от количества прецедентов в БП при различных значениях *k*: а) *k*=1; б) *k*_{avg}; в) *k*_{min}; г) *k*_{max}

Fig. 1. A dependency diagram of classification quality on the number of use cases in a use case database for different values of *k*: а) *k*=1; б) *k*_{avg}; в) *k*_{min}; г) *k*_{max}

набора данных из хранилища UCI Machine Learning Repository, включающего тестовую выборку и БД с информацией об уровне знаний студентов по дисциплине «Электрические машины постоянного тока», которая содержит 258 записей [8]. Начальная БП была сформирована на основе первых 20 записей из БД в UCI Machine Learning Repository, а остальные 238 записей были включены в обучающую выборку.

Из графиков, представленных на рисунке 1, видно, что показатели качества классификации при увеличении БП (накоплении опыта – прецедентов) выше при выборе в алгоритме k -NN для извлечения прецедентов значения k , равного k_{AVG} . Следует отметить, что при каждом добавлении нового прецедента в БП вычисляется новое значение для k_{AVG} .

Модифицированный CBR-цикл

В работе предложен модифицированный CBR-цикл (рис. 2), в котором могут применяться тестовые выборки с примерами для проверки корректности найденного решения.

При наличии экспертных знаний (тестовых выборок) перед сохранением в CBR-цикле должна выполняться проверка корректности решения на

тестовых наборах. Если решение проходит проверку и принимается пользователем, оно сохраняется в БП как новый прецедент. Если проверка корректности решения на тестовых наборах завершается неудачно, прецедент сохраняется в базе *неудачных прецедентов* (БНП).

Удачным назовем прецедент, который не ухудшает качество работы CBR-системы после его добавления в БП, а *неудачным* – прецедент, который ухудшает качество работы CBR-системы после его добавления в БП. Таким образом, предлагается использовать тестовую (экспертную) выборку на последнем этапе CBR-цикла для формирования БП и БНП.

Для предыдущего примера решения задачи классификации данных с использованием начальной БП из 20 прецедентов был применен модифицированный CBR-цикл. В результате его работы осуществлялось пополнение начальной БП новыми прецедентами, формируемыми на основе обучающей выборки из 238 примеров, и при этом выполнялась проверка на заданной тестовой выборке. Если вновь добавленный прецедент ухудшал качество классификации, он удалялся из БП и сохранялся в БНП. Таким образом, был накоплен 171 прецедент в БП и 87 неудачных прецедентов в

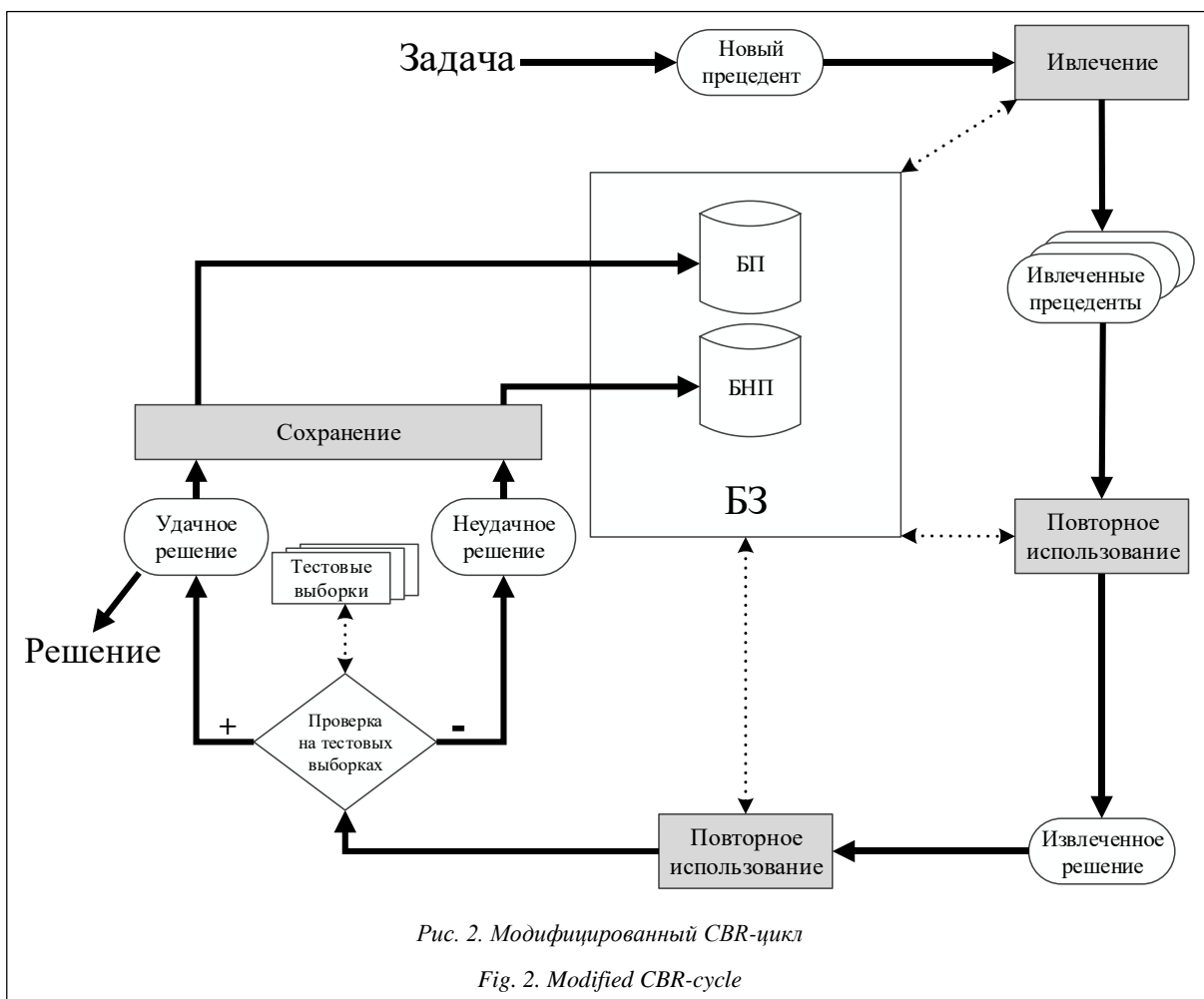


Рис. 2. Модифицированный CBR-цикл

Fig. 2. Modified CBR-cycle

БП. Результаты оценки качества классификации CBR-системой приведены на рисунке 3.



Рис. 3. Оценка качества классификации для БП, сформированной на основе модифицированного CBR-цикла

Fig. 3. Classification quality estimation for a use case database formed on the basis of a modified CBR cycle

Использование модифицированного CBR-цикла и формирование БП на основе удачных прецедентов (171 прецедент) позволили улучшить качество классификации CBR-системой до 87 %.

Повышение эффективности работы CBR-систем

В процессе функционирования CBR-систем увеличивается количество прецедентов в БП, то есть CBR-системы обучаются (накапливают опыт) и качественно решают поставленные перед ними задачи, но при этом значительно увеличиваются временные затраты и затраты на организацию хранения прецедентов, требуемые для поддержания CBR-цикла. По этой причине для CBR-систем актуальной является задача повышения эффективности их работы за счет увеличения быстродействия CBR-систем и сокращения объема памяти, необходимого для хранения прецедентов в БП. Достичь этого можно несколькими путями, например, за счет индексации БП или сокращения количества прецедентов в БП системы.

Индексирование БП. Можно повысить эффективность (быстродействие) работы CBR-системы за счет формирования индекса по БП. Индекс – объект БД, создаваемый с целью повышения производительности поиска данных. Таблицы в БД могут иметь большое количество строк, хранящихся в произвольном порядке, и их поиск по заданному критерию путем последовательного просмотра таблицы строка за строкой может занимать много времени. Индекс формируется из значений одного или нескольких столбцов таблицы и указателей на соответствующие строки таблицы и, таким образом, позволяет искать строки, удовлетворяющие критерию поиска. Ускорение работы с использованием индексов достигается в первую очередь за счет того, что индекс имеет структуру, оптимизированную под поиск, например сбалансированного

дерева. Возможно сокращение количества прецедентов в БП за счет применения методов классификации и кластеризации.

Сокращение количества прецедентов в БП с использованием метода классификации. Одним из способов сокращения количества прецедентов в БП является классификация (например, с применением алгоритма k -NN) с последующим удалением всех прецедентов, которые определяются как ближайшие к выбранному. Это позволяет представить БП в компактном виде для пользователя с сохранением распределения по классам, снизив объем памяти, необходимый для хранения прецедентов, и существенно повысить быстродействие CBR-системы, но при этом возможно ощутимое снижение качества решения задач CBR-системой.

Алгоритм 1. Алгоритм сокращения количества прецедентов в БП на основе классификации.

Входные данные: CB – непустое множество прецедентов (БП); n – количество параметров в прецедентах; m – количество прецедентов в БП; $S(C_i, C_j)$ – заданная метрика для вычисления расстояния между прецедентами C_i и C_j ; H – пороговое значение степени сходства.

Выходные данные: CB – множество прецедентов после сокращения БП.

Промежуточные данные: SC – множество ближайших прецедентов к выбранному; i, j – параметры циклов.

Шаг 1. $j = 1$, $SC = \emptyset$, переход к следующему шагу.

Шаг 2. Если $j \leq m$, выбираем прецедент C_j из множества CB ($C_j \in CB$); переход к шагу 3, иначе все прецеденты из БП рассмотрены и переход к шагу 5.

Шаг 3. Определяем SC – множество ближайших прецедентов C_i ($C_i \in CB, i \neq j$) к прецеденту C_j в соответствии с заданной метрикой $S(C_i, C_j)$ и переходим к следующему шагу.

Шаг 4. Если $SC \neq \emptyset$, удаляем из CB все прецеденты, входящие в множество SC . Далее $m = m - |SC|$, $SC = \emptyset$, $j = j + 1$ и переход к шагу 2.

Шаг 5. Конец (завершение алгоритма).

Сокращение количества прецедентов в БП на основе кластерных методов. Кластерный анализ (кластеризация) – это способ группировки многомерных объектов, основанный на представлении результатов отдельных наблюдений точками подходящего геометрического пространства с последующим выделением групп как «сгустков» этих точек (кластеров, таксонов). В последние годы данный метод исследования получил развитие в связи с возможностью компьютерной обработки больших БД. Кластеризация предполагает выделение компактных, удаленных друг от друга групп объектов, характеризующихся внутренней однородностью и внешней изолированностью. Применительно к CBR-системам алгоритмы кластеризации

позволяют получить множество прецедентов – «центров» кластеров, по отношению к которым остальные прецеденты в кластере близки. То есть множество «центров» кластеров в БП можно использовать вместо самой БП, так как ситуации, близкие к какому-либо прецеденту из кластера, будут близки и к центру этого кластера.

В работе предлагается подход к сокращению количества прецедентов в БП на основе кластеризации (алгоритм k -средних) с последующим удалением всех прецедентов в каждом кластере, кроме его центрального представителя. Таким образом, кластеризация позволяет представить БП в более прозрачном виде для пользователя с сохранением распределения по кластерам и сократить время поиска решения и объем памяти, необходимый для хранения прецедентов в БП. Недостатком этого способа является потеря точности на границах кластеров.

Алгоритм 2. Алгоритм k -средних для сокращения количества прецедентов в БП.

Входные данные: CB – БП; m – количество прецедентов в БП; k – количество кластеров; n – количество параметров в описании прецедентов; $S(C_i, C_j)$ – заданная метрика для вычисления расстояния между претендентами C_i и C_j .

Выходные данные: CB – множество прецедентов после сокращения БП.

Промежуточные данные: CL – полученные кластеры; μ – центры масс кластеров; V – суммарное квадратичное отклонение объектов (прецедентов) в кластерах от их центров; i, j – параметры циклов.

Шаг 1. Выбираются k случайных центров кластеров (прецедентов) из CB .

Шаг 2. Для каждого центра кластера μ_i ($i=1, \dots, k$) определяется множество прецедентов, для которых соответствующий центр является ближайшим CL_i , то есть расстояние $S(C_j, \mu_i)$ (метрика) от прецедента C_j ($C_j \in CB$) до центра кластера μ_i минимально.

Шаг 3. Если для одного из центров соответствующее ему множество ближайших прецедентов является пустым, вместо него выбирается новый случайный центр (прецедент).

Шаг 4. В каждом полученном множестве ближайших прецедентов (кластере CL_i) определяется новый центр масс кластеров, который используется на следующей итерации.

Шаг 5. Оценивается суммарное расстояние прецедентов в БП до полученных центров:

$$V = \sum_{i=1}^k \sum_{C_j \in CL_i} (C_j - \mu_i)^2.$$

Если суммарное расстояние не уменьшилось, осуществляется переход к шагу 6, иначе – возврат к шагу 2.

Шаг 6. Из CB удаляются все прецеденты, за исключением полученных центров кластеров.

Шаг 7. Завершение алгоритма (конец).

В таблице приведены результаты вычислительных экспериментов по оценке эффективности предложенных алгоритмов для сокращения количества прецедентов в БП на примере рассмотренной выше задачи классификации данных из хранилища UCI Machine Learning Repository.

Результаты исследования Research results

БП	Количество прецедентов	Качество классификации, %
Начальная	20	64
Накопленная	171	87
Полученная в результате применения алгоритма 1	25	72
Полученная в результате применения алгоритма 2	4	80

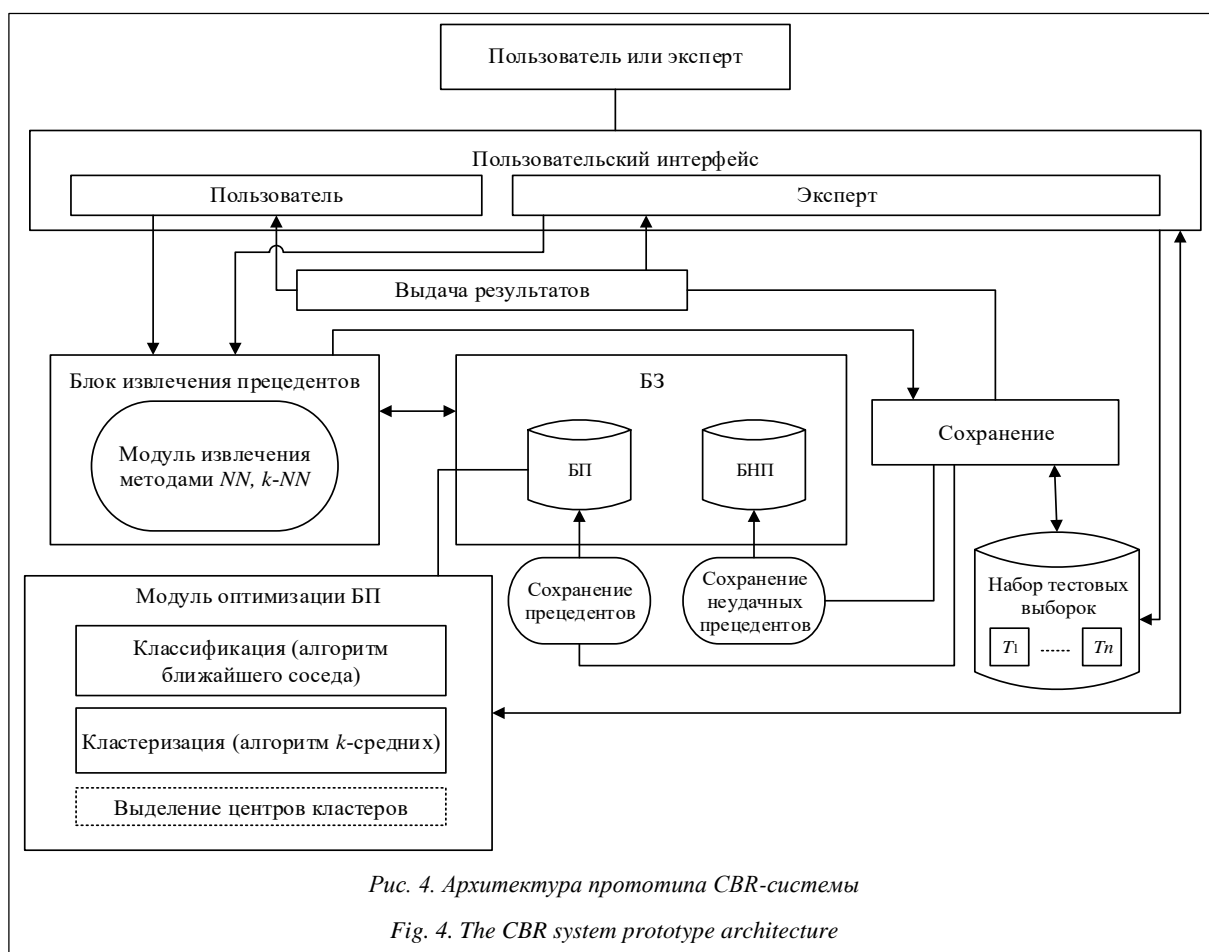
Как видно из таблицы, при сокращении количества прецедентов в накопленной БП (со 171 до 25 прецедентов) с использованием классификационного алгоритма 1 качество классификации на тестовом наборе снижается на 15 %, а при сокращении количества прецедентов в накопленной БП (со 171 до 4 прецедентов) с использованием кластерного алгоритма 2 качество классификации на тестовом наборе снижается только на 7 %, что является допустимым значением при существенном повышении быстродействия СВР-системы и уменьшении объема памяти, необходимого для хранения прецедентов в БП, за счет сокращения количества прецедентов в БП (более чем в 40 раз).

Архитектура прототипа СВР-системы

Архитектура прототипа СВР-системы состоит из следующих основных компонентов (рис. 4):

- *пользовательский интерфейс* – интерфейс для взаимодействия с экспертом или пользователем и отображения результатов работы;
- *блок извлечения прецедентов* – для поиска прецедентов на основе модифицированного метода k -NN;
- *база знаний* – содержит базу удачных и базу неудачных прецедентов (соответственно БП и БНП);
- *набор тестовых выборок* – экспертная информация, содержащая данные (примеры) с корректными решениями, для ее учета при извлечении прецедентов и формирования БП и БНП;
- *модуль оптимизации БП* – для выполнения сокращения количества прецедентов в БП с использованием предложенных в работе алгоритмов.

Программная реализация прототипа СВР-системы выполнена с использованием языка C# и среды программирования MS Visual Studio, а также



СУБД Microsoft SQL Server, в том числе Microsoft SQL Server Analysis Services. Модуль оптимизации базы прецедентов для CBR-систем зарегистрирован в государственном реестре программ для ЭВМ (свидетельство № 2016617638 от 13 июля 2016 г.).

Заключение

В статье обсуждается перспективная возможность расширения существующих инструментов ИАД для современных СУБД и СУБЗ средствами поиска решения (рассуждения) на основе прецедентов (CBR-средствами), а также рассматриваются актуальные вопросы повышения эффективности работы CBR-систем. Для повышения качества решения, получаемого CBR-системой, разработана модификация алгоритма извлечения прецедентов на основе k -NN, заключающаяся в изменении значения k в зависимости от размера БП, и предложен модифицированный CBR-цикл, использующий экспертную информацию (тестовые наборы данных) для проверки корректности полученного решения и формирования БП и БНП. Для повышения эффективности работы CBR-систем предложены алгоритмы сокращения количества прецедентов в БП на основе методов классификации и кластеризации. С использованием разработанного прототипа CBR-системы проведены вычислительные

эксперименты по оценке эффективности предлагаемых в работе решений на наборе данных из хранилища UCI Machine Learning Repository.

Работа выполнена при финансовой поддержке РФФИ, проекты №№ 15-07-04574, 16-51-00058, 17-07-00553.

Литература

1. Рассел С., Норвиг П. Искусственный интеллект: современный подход; [пер. с англ.]. М.: Вильямс, 2006. 1410 с.
2. Финн В.К. Об интеллектуальном анализе данных // Новостной искусственного интеллекта. 2004. № 3. С. 3–19.
3. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. М.: Физматлит, 2008. 712 с.
4. Барсегян А.А., Холод И.И., Тесс М.Д., Куприянов М.С., Елизаров С.И. Анализ данных и процессов. СПб: БХВ-Петербург, 2009. 512 с.
5. Варшавский П.Р., Еремеев А.П. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений. 2009. № 2. С. 45–57.
6. Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. Artificial Intelligence Communications. IOS Press. 1994, vol. 7, no. 1, pp. 39–59.
7. Алахин Р.В., Ар Кар Мью, Варшавский П.Р., Зо Лин Кханг. Реализация прецедентного модуля для интеллектуальных систем // Программные продукты и системы. 2015. № 2. С. 26–31.
8. User knowledge modeling data set. URL: <http://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling>; (дата обращения: 10.03.2017).

**CLASSIFICATION AND CLUSTERING METHODS FOR IMPROVING EFFICIENCY
OF CASE-BASED SYSTEMS**

P.R. Varshavsky¹, Ph.D. (Engineering), Associate Professor, VarshavskyPR@mpei.ru
Ar Kar Myo¹, Postgraduate Student, arkar2011@gmail.com
D.V. Shunkevich², Assistant, shunkevichdv@gmail.com

¹ National Research University "MPEI",
Krasnokazarmennaya St. 14, Moscow, 111250, Russian Federation

² Belarusian State University of Informatics and Radioelectronics (BSUIR),
P. Brovki St. 6, Minsk, 220013, Belarus

Abstract. The article examines topical issues of improving efficiency of case-based reasoning (CBR) systems. Case-based methods and systems (CBR systems) are actively used to solve a number of problems in the field of artificial intelligence (for example, for modeling plausible reasoning (common sense reasoning), machine learning, intellectual decision support, intelligent information search, data mining (DM) and etc.). It should be noted that modern tools for DM, which are widely used in intelligent systems, database and knowledgebase management systems, business applications, machine learning systems, electronic document management systems, etc., do not have advanced CBR tools.

The paper proposes to use the modified CBR cycle to increase the efficiency of CBR systems. This cycle allows creating a base of successful (CB) and unsuccessful cases (UCB) based on available expert information (test samples), as well as the k -nearest neighbors (k -NN) modification algorithm for case retrieval. The proposed modifications allow improving the quality of solving DM tasks (in particular, data classification task). In addition, the authors consider the possibility of reducing the number of cases in CB using classification and clustering methods to improve performance of CBR systems.

The paper shows computational experiments to estimate the effectiveness of the solutions offered while working on a data set from the UCI Machine Learning Repository. They use CBR system prototype developed in MS Visual Studio in C# language.

Keywords: intelligent system, case-based approach, data mining, classification, clustering.

Acknowledgements. This work has been supported by RFBR projects no. 15-07-04574, 16-51-00058, 17-07-00553.

References

1. Russell S., Norvig P. *Artificial Intelligence: a Modern Approach*. Pearson Publ., 2009, 3rd ed., 1159 p. (Russ. ed.: Moscow, Vilyams Publ., 2010, 1410 p.).
2. Finn V.K. On intelligent data analysis. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News]. 2004, no. 3, pp. 3–19 (in Russ.).
3. Vagin V.N., Golovina E.Yu., Zagoryanskaya A.A., Fomina M.V. *Dostoverny i pravdopodobny vyvod v intelektualnykh sistemakh* [Reliable and Believable Reasoning in Intelligent Systems]. 2nd ed. V.N. Vagin, D.A. Pospelov (Eds.). Moscow, Fizmatlit Publ., 2008.
4. Barsegyan A.A., Kholod I.I., Tess M.D., Kupriyanov M.S., Elizarov S.I. *Analiz dannykh i protsessov* [Data and Process Analysis]. 3rd ed. St. Petersburg, BHV-Peterburg Publ., 2009, 512 p.
5. Varshavsky P.R., Ereemeev A.P. Modeling of case-based reasoning in intelligent decision support systems. *Iskusstvenny intellekt i prinyatie resheny* [Artificial intelligence and decision-making]. 2009, no. 2, pp. 45–47 (in Russ.).
6. Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *Artificial Intelligence Communications*. IOS Press, 1994, vol. 7, no. 1, pp. 39–59.
7. Alekhin R.V., Ar Kar Myo, Varshavsky P.R., Zo Lin Khaing. Implementation of a cases-based module for intelligent systems. *Programmnye produkty i sistemy* [Software & Systems]. 2015, no. 2, pp. 26–31 (in Russ.).
8. *User Knowledge Modeling Data Set*. Available at: <http://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling> (accessed March 10, 2017).