

Методика обучения базы знаний интеллектуальной системы реального времени на основе алгоритма решающих деревьев

Г.А. Благодатский¹, С.В. Вологдин¹, М.М. Горохов¹, Д.А. Переведенцев¹

¹Ижевский государственный технический университет имени М.Т. Калашникова, Студенческая 7, Ижевск, Россия, 426069

Аннотация. Работа посвящена разработке математического алгоритма самообучения интеллектуальной системы (ИС) управления наукоемкими проектами создания сложных технических объектов. Обучение ИС происходит путем оптимизации правил базы знаний, используя методы машинного обучения. Это позволит повысить эффективность поддержки принятия решений при реализации наукоемких проектов и адаптации рекомендаций ИС под запрос лица принимающего решение. На основе анализа и разметки имеющихся данных, а также сформулированных требований к алгоритму, выбран алгоритм дерева регрессии и классификации (CART) в силу высокой скорости его работы и возможностей формализации полученных зависимостей в формате продукционных правил. В соответствии с выставленными требованиями определены параметры обучения и проведена проверка эффективности работы алгоритма классификации научных проектов на заданное число классов и формулирования новых зависимостей базы знаний. Представлена общая схема обработки запроса пользователя описанным функционалом интеллектуального модуля.

1. Введение

В процессе реализации наукоемких проектов сложных технических объектов сегодня создаются и используются большие объемы разнородных и распределенных данных, подобная информационная нагрузка затрудняет не только эффективное использование накопленных данных и знаний, но и понимание характеристик развития самого проекта.

Для решения указанной проблемы учеными предлагаются различные методологические подходы к снижению уровня неопределенности наукоемких проектов, например: анализ знаний экспертов и заинтересованных сторон [1], анализ связей организационной структуры, персонала и информационных ресурсов проекта [8], планирование на основе метода критической цепи и принципах теории ограничений согласно предметной области проекта [2], анализ социальной сети и связей проекта [9].

В разное время предлагались модели анализа проектов сложных технических систем. Такенберг С., Дуквиц С. и Шлик К. предложили проектный инженерный подход на основе процессных и актер-ориентированных моделей [4]. Бухиным Е. и Розенсом Х. [5] разработана многоцелевая оптимизационная модель выбора решения о реализации проекта из набора альтернативных вариантов. В статье [6] приведены примеры решения задач оптимизации проекта в различных ситуациях на основе построенной авторами структуры дерева решений по

времени, называемой «деревом проекта». Авторами работы [7] предложены графические структуры представления знаний о проекте на основе тематического и ситуативного анализа жизненного цикла. В работе [10] представлена оптимизационная модель задачи планирования процессов проектного управления с ограничениями нескольких исполнительных режимов, многопрофильности сотрудников и приоритета задач.

Стоит отметить, что моделирование и управление наукоемкими проектами на основе знаний используется во многих сферах деятельности: разработка транспортной инфраструктуры [13], строительство [12], внедрение методов бережливого производства [11], производство электроэнергии [14], разработке стратегий гуманитарной логистики [15] и другие.

Кроме того актуальность разработки специальных систем интеллектуальной обработки данных подтверждается все новыми исследованиями в данной области. Так, в работе [3] приведена классификация систем поддержки принятия решений основанных на знаниях (KBDS) на три категории: технологии моделирования и представления знаний, технологии для рассуждений и умозаключений и веб-технологии, представлены рекомендации их применения в качестве систем обслуживания. Авторами работы [17] предложена автоматическая модель и аналитические подходы к обучению и прогнозированию характеристик проектов, в работе [18] представлен модельный метод классификации новых фактов различных инженерных систем на основе интеллектуального анализа данных оценки параметров данных систем, авторами исследования [19] описан новый подход приобретения знаний для быстрого развития экспертных систем, основанный на использовании лингвистических правила, совместимых с эвристическими экспертными знаниями.

Указанные обстоятельства значительно повлияли на процессы принятия решений, основой которых все чаще являются рекомендательные системы, разрабатываются подходы к формированию экспертных решений с коллаборативной фильтрацией решений, представляющие собой моделирование уверенности в контексте имеющихся неполных данных [21].

Учеными Сакама С. и Инуе К. показана важность абдуктивного подхода к формированию новых знаний, поскольку оно опирается на поиск объяснительных гипотез, которые подтверждаются и обосновываются новыми релевантными фактами, зарегистрированными системой [20]. В рамках данного подхода для обучения интеллектуальных систем показана эффективность алгоритма деревьев принятия решений [16].

2. Описание проблемной среды и постановка задачи исследования

В статье рассматривается онлайн-платформа регионального бизнес-инкубатора [25], которая является информационно-аналитической системой (ИАС) для оценки и отбора наукоемких проектов, представляющий собой совокупность баз данных, учетных, аналитических и интеллектуальных модулей, а также интерфейса взаимодействия с пользователем, технологическая схема данной системы представлена на рисунке 1.

Обозначение на рисунке 1: АРМ1 – автоматизированное место пользователя, АРМ2 – автоматизированное место эксперта, БД1 – база данных пользователей, БД2 – база данных проектов, БПр – база правил нечеткой логики, БЗн – база знаний для экспертных рекомендаций, ХД – хранилище данных, АМ – модуль анализа, ЭМ – модуль генерации экспертных рекомендаций, МВФД – модуль валидации и фильтрации данных, МРО – модуль расчета оценки проектов, ММОб – модуль машинного обучения системы.

Интеллектуальным ядром системы является модуль экспертных рекомендаций, на основе которого в системе формируются рекомендации о возможности финансирования проекта по результатам анализа его перспективности и текущих рисков реализации.

Таким образом, архитектура приложения объединяет в единое информационное пространство подсистемы оперативной работы с проектами, их многомерного анализа и экспертный модуль, реализующие широкий набор решений и процедур, необходимых для повышения качества организации, продвижения и реализации наукоемких проектов.

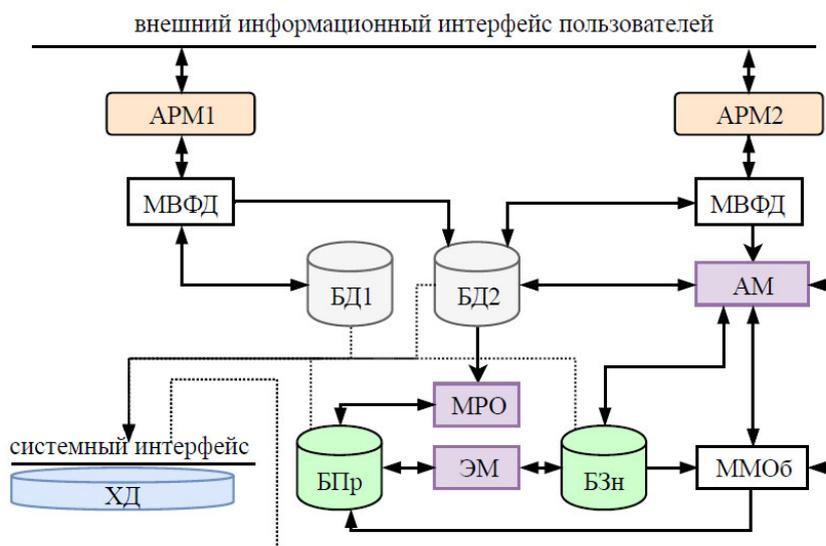


Рисунок 1. Технологическая схема АКС поддержки наукоемких проектов.

Качество данной системы, как и любой ИАС, определяется размером и актуальностью используемой базы знаний для вывода рекомендаций, поэтому поддержка принятия решений в динамичных условиях, в которых действуют научно-технические системы, предъявляет высокие требования к адаптивности и функциональному развитию системы поддержки принятия решений. Однако привлечение экспертов для расширения и корректировки базы знаний ИС в соответствии с новыми данным в данной случае нецелесообразно, сегодня данная задача эффективно решается использованием алгоритмов машинного обучения.

Таким образом, предметом рассмотрения данной статьи является исследование и решение научно-технической проблемы актуализации базы правил интеллектуальной системы с применением современных технологий математического моделирования и вычислительного эксперимента.

3. Описание методики решения поставленной задачи

В рамках данной статьи представим подход к моделированию экспертных знаний в процессе анализа и выбора наукоемких проектов бизнес-инкубатора. На основе экспертного опроса и описания онтологии предметной области, согласно разработанной ранее методики [22, 23], сформулированы критерии оценки проектов и построены их математические модели. Таким образом, получен следующий список критериев проекта: оценка руководителя (К1), оценка признания проекта (К2), оценка наукоемкости (К3), оценка ресурсов проекта (К4), оценка коммерческой реализуемости (К5), оценка текущих результатов проекта (К6), новизна (К7), оценка знания проблемной области (К8), рентабельность (К9), оценка аналогов (К10), Оценка стратегии (К11), оценка рисков (К12) и общая оценка проекта (Y). Все значения критерии нормализованы и представлены в интервале от 0 до 10.

Пример рассчитанной оценки проекта представлен на рисунке 2.

Таким образом, формализовано понятие общей оценки проекта и оценки структуры его критериев для выявления потенциала развития проекта. Разработанный алгоритм оценки проекта на основе совокупности данных расчетных характеристик позволяет наиболее точно определить его состояние.

В рамках описанной онлайн-платформы задача отбора проектов сводится к выбору наиболее перспективного для инвестирования проекта в определенный момент времени. В данном случае разработанная методика формализации проектов позволяет объединить алгоритм отбора и анализ данных.

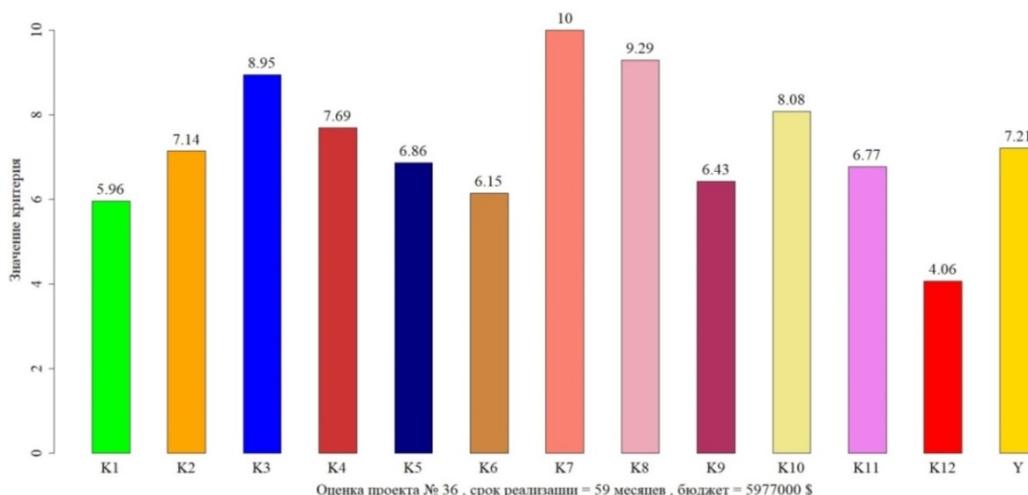


Рисунок 2. Пример оценки проекта.

Далее опишем методику разработки алгоритма оптимизации правил отбора проектов на основе построенных моделей его оценки. Однако, прежде чем перейти к описанию правил базы знаний, необходимо провести классификацию проектов и определиться с набором признаков, на основе которых можно отнести проект к тому или иному классу. Если формализовать условия на основе представленных 12 критериев и общей оценки (Y) проекта, получим довольно сложную структуру правил, поэтому проведем анализ значимости признаков с целью сокращения их числа.

Разработанный алгоритм оценки подразумевает расчет нескольких комплексных признаков («оценка руководителя», «новизна», «рентабельность», «оценка рисков» и «общая оценка»), поэтому в качестве гипотезы выдвинем предположение о большей значимости данных критериев при классификации проектов. С помощью специально разработанной анкеты собраны данные 2570 проектов, поданных на регистрацию в бизнес-инкубатор. Представленным проектам эксперты присвоили 4 класса согласно следующей иерархии (таблица 1).

Таблица 1. Описание структуры исследуемой выборки проектов.

Класс	Описание	Количество наблюдений
A	Проект имеет высокий потенциал реализации	310
B	Проект является перспективным	845
C	Проект имеет коммерческий потенциал	759
D	Проект имеет низкий потенциал реализации	656
Итого:		2570

Метод бэггинга над алгоритмом CART позволяет оценить важность признаков путем построения множества деревьев и последующего усреднения их результатов. Другим названием данного подхода является *Random forest* (с англ. «случайный лес»). На рисунке 3 представлены результаты 1000 независимых раундов перекрестной проверки признаков.

В алгоритме *Random forest* для оценки важности переменных используются метрики Mean Decrease Accuracy (RMSE) и Mean DecreaseGini.

MeanDecreaseAccuracy измеряет, насколько включение этого предиктора в модель уменьшает ошибку классификации. RMSE можно вычислить, используя формулу (1):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{f}(x))^2}{n}}, \tag{1}$$

где y_i – i -е наблюдение переменной, $\hat{f}(x)$ – i -е предсказание для i -го наблюдения и n_i – количество выборов.

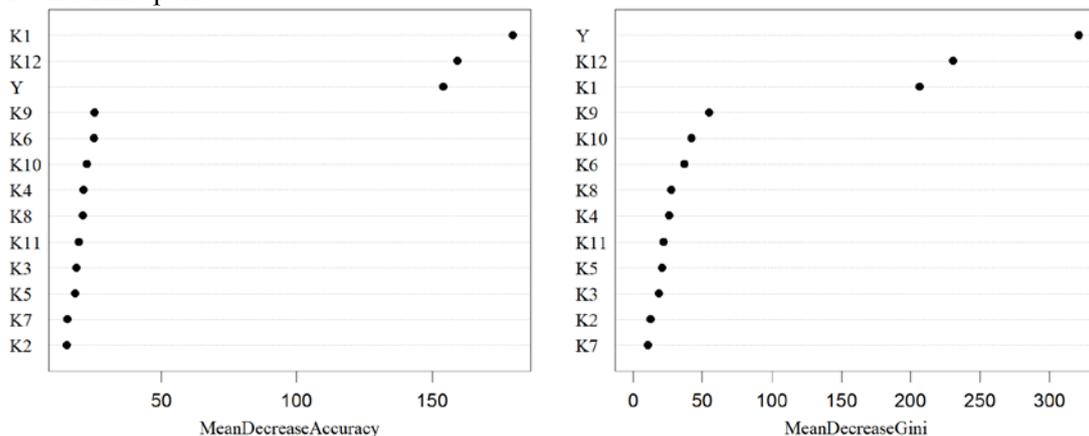


Рисунок 3. Уровень значимости признаков классификации проектов.

MeanDecreaseGini является мерой качества разделения для каждой переменной дерева с помощью индекса Gini и показывает как каждая переменная способствует однородности узлов и листьев в полученном случайном лесу, рассчитывается как (2):

$$\sum_{i \in left} (y_{left} - y_i)^2 + \sum_{i \in right} (y_{right} - y_i)^2, \tag{2}$$

где y_{left} и y_{right} – ошибка классификации наблюдения при расщеплении соответственно на левую и правую ветвь дерева.

Более высокие значения Mean Decrease Accuracy и Mean DecreaseGini, таким образом, указывает на большее прогностическое значение переменной [24].

Из рисунка 3 видим, что наиболее значимыми можно выделить 3 признака: «оценка руководителя», «общая оценка» и «оценка рисков», что подтверждает выдвинутую ранее гипотезу.

Определение оптимального уровня описания данных является актуальным для всех эвристических алгоритмов, в частности для дерева решений это размер конечного дерева. Так, мало разветвленное дерево может не отразить часть важной для исследования информацию о выборочном пространстве, с другой стороны – слишком детальное дерево может нагрузить базу знаний излишне подробными и незначительными правилами. В связи с этим довольно сложно определить, когда алгоритм должен прекратить наращивание внутренних узлов, потому что невозможно спрогнозировать, добавление какого узла позволит наиболее значительно уменьшить неопределенность в структуре данных. Поэтому далее оценим влияние различной совокупности признаков и значения критерия останова обучения на качество классификации проектов, для этого разобьем набор данных на обучающую и тестовую выборку в отношении 80% на 20%. Для оценки качества модели используем F -меру – метрику, которая представляет собой среднее гармоническое точности (precision) и полноты (recall) алгоритма классификации проектов (3):

$$F = (\beta^2 + 1) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times (\text{Precision} + \text{Recall})}. \tag{3}$$

Если приоритет отдается точности, β принимает значения в диапазоне $\{0;1\}$, а если полноте – то при $\beta > 1$. При $\beta = 1$ формула придает одинаковый вес точности и полноте, так называемая сбалансированная F -мера. В рамках данной работы примем $\beta = 1$.

Результаты расчета эффективности обучения модели при различном наборе признаков и уровне критерия остановки b приведены на графике (рисунок 4). Здесь b – это минимальное количество наблюдений в листе дерева, рассчитывается в процентах от общего числа наблюдений.

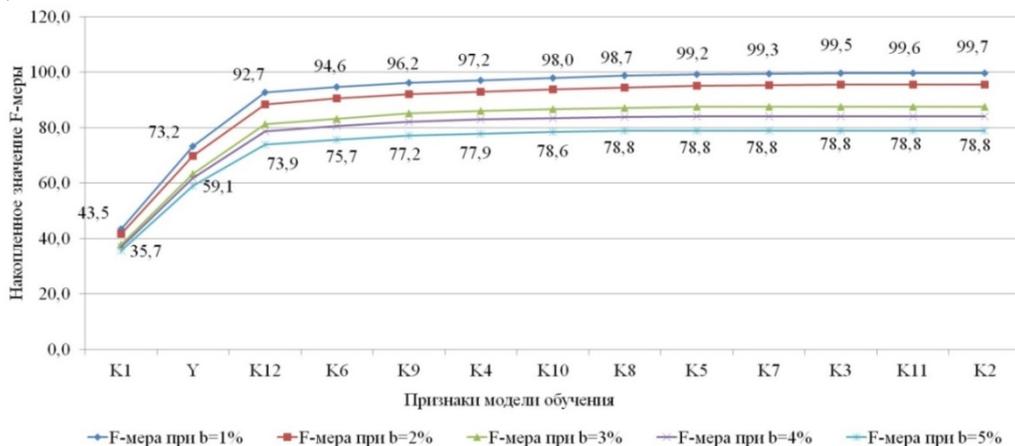


Рисунок 4. Оценка качества модели в зависимости от уровня критерия остановки обучения.

Из рисунка 4 видим, что критерии «Общая оценка», «Оценка рисков» и «Оценка руководителя» в совокупности позволяют на уровне 90-процентной эффективности классифицировать проекты в терминах предметной области, что является приемлемым уровнем качества модели. Выбор уровня критерия остановки равный наличию не менее 2% объектов в каждом классе при незначительном снижении точности классификации позволяет получить оптимальную структуру дерева решения для его формализации, таким образом, можно принять данное значение для критерия остановки обучения. Таким образом, выбираем указанные параметры модели для формирования правил базы знаний и ее обучения.

4. Разработка алгоритма оптимизации базы экспертных рекомендаций

Обозначим исходный набор данных T , наблюдения которого классифицированы на n классов, тогда критерий расщепления дерева, который в алгоритме CART представлен индексом $Gini$, определяется как (4):

$$Gini(T) = 1 - \sum_{i=1}^n p_i^2, \tag{4}$$

где p_i – вероятность (относительная частота) класса n_i в T . Если набор T разбивается на две части T_1 и T_2 с числом примеров в каждом N_1 и N_2 соответственно, тогда показатель качества разбиения будет равен (5):

$$Gini_{split}(T) = \frac{N_1}{N} Gini(T_1) + \frac{N_2}{N} Gini(T_2). \tag{5}$$

Таким образом, при построении дерева решений по методу CART выявляется такой вариант ветвления, при котором максимально уменьшается значение показателя $Gini_{split}(T)$.

Представленный механизм для анализа базы данных наукоемких проектов можно описать следующей процедурой:

1. Вычисляем энтропию исходного множества $Gini(S)$:

1.1. если $Gini(S) = 0$ значит все объекты исходного набора, принадлежат к одному классу, сохраняем этот класс в качестве листа дерева;

1.2. если $Gini(S) \neq 0$, значит перебираем все элементы исходного множества, при этом выбирается k -ый признак f_k с множеством значений $X(k)$, на основе каждого элемента генерируем предикат вида $x_{i1} \leq c < x_{i2}$, где c – некоторый порог, который чаще всего выбирается как среднее арифметическое двух соседних упорядоченных значений переменной обучающей выборки и разбивающий исходное множество на два подмножества.

2. Рассчитываем значение прироста определенности, при этом определяется такое значение $x_0(k) \in X(k)$ для всех признаков $f_k, k = 1, \dots, m$, чтобы мера неопределенности *Ginisplit* была минимальной, т.е.: $x_0^{(k)} = \arg \min_{f_k, x^{(k)} \in X^{(k)}} Gini_{split}(T, X^{(k)})$.

Найденный предикат является частью дерева принятия решений, он сохраняется, а исходное множество разбивается на два подмножества, согласно выбранному полученному условию.

3. Указанная процедура выполняется рекурсивно для каждого полученного подмножества до тех пор, пока не будет достигнуто условие критерия останова.

Приемлемым уровнем останова алгоритма обучения также выберем тот, при котором возможно формализовать полный набор правил при незначительном отклонении от максимального уровня точности. Из рисунка 4 видим, что наиболее критичное падение эффективности алгоритма происходит при уровне критерия останова обучения в 3% объектов каждого класса.

Анализ структуры деревьев решений при уровне критерия останова обучения равном 1%, 2% и 3% показал, что при уровне 3% модель обладает меньше объяснительной способностью, поэтому выбор уровня критерия останова равный наличию не менее 2% объектов в каждом классе при незначительном снижении точности классификации позволяет получить оптимальную структуру дерева решения для его формализации, таким образом, можно принять данное значение для критерия останова обучения.

В этом случае в базу знаний добавятся следующие правила:

- 1) если $Y \geq 5,7$ и $K1 \geq 6,1$, то проект принадлежит классу А;
- 2) если $Y \geq 5,7$ и $K1 < 6,1$ и $K12 \geq 4,5$, то проект принадлежит классу В;
- 3) если $Y < 5,7$ и $K12 \geq 4,5$, то проекта принадлежит классу С;
- 4) если $Y \geq 5,7$ и $K1 < 6,1$ и $K12 < 4,5$, то проект принадлежит классу D;
- 5) если $Y < 5,7$ и $K12 < 4,5$, то проект принадлежит классу D.

По результатам проведенного анализа реализован алгоритм обучения интеллектуальной системы, который позволяет расширять базу знаний на основе новых данных о динамике развития проектов, каждое новое правила хранится в базе знаний до тех пор, пока не будет добавлено правило исключаящее его. Схема обработки данных наукоемкого проекта с учетом алгоритмов расчета оценки проекта выглядит следующим образом (рисунок 5):

Таким образом, разработанный алгоритм позволяет в реальном времени формировать рекомендации, актуальные накопленным данным о проектах и предпочтениям инвесторов.

5. Описание формата правил базы знаний

Как было показано ранее, окончательное решение при выборе проекта для финансирования зависит от общей оценки состояния проекта и уровня принятия рисков инвестором.



Рисунок 5. Схема обработки запроса на подбор проекта.

Пусть x – проект, $S(x)$ – состояние проекта. На основе выбранных переменных опишем алгоритмы базы правил для формирования экспертной рекомендации и условия их срабатывания, полученные данные представим в виде продукционной модели, отражающей

текущее состояние проекта и заданная функцией (6):

$$S(x) = Y \wedge K1 \wedge K12, \tag{6}$$

где: $S(x)$ – вывод о состоянии проекта x и его возможных рисках, $S_x = \{вывод\}$; Y – общая оценка проекта, $Y_x \in (0;10)$; $K1$ – оценка опыта руководителя проекта, $K1_x \in (0;10)$; $K12$ – оценка рисков проекта, $K12_x \in (0;10)$.

Итак, вывод о состоянии проекта строится на основе сопоставления общей оценки, уровня оценки опыта руководителя и уровня оценки рисков проекта. Правило базы знаний для данного блока выглядит следующим образом:

«Формирование вывода» = <«Параметры проекта», блок актуализирован, $Y \wedge K1 \wedge K12 \rightarrow$ решение = $\{вывод\}$, иначе = «недостаточно данных», решение о состоянии проекта и его перспективности $R(x_i)$ >

Формирование начальной базы правил происходит путем описания множества правил исходя из возможных сочетаний входных и выходных параметров, отметим, что использование указанного подхода целесообразно при небольшом числе переменных. Кроме того указанный подход также позволит данной базе знаний соответствовать критерию избыточности и полноты, поскольку в этом случае для любых входных значений существует хотя бы одно правило вывода, а разработанный алгоритм оптимизации позволит исключить правила с одинаковыми функциями вывода.

Таким образом, создана база знаний, выраженная на естественном языке в терминах предметной области, данная база является ядром модуля принятия решений интеллектуальной системы бизнес-инкубатора.

6. Эксперимент

В качестве примера отберем 10 проектов, участвующих в одном конкурсе инновационных проектов [25], и проведем их анализ средствами разработанной ИАС. Результаты сравнения методик по оценке проектов приведены в таблице 2.

Таблица 2. Описание наукоемких проектов.

№	Название проекта	Решение экспертной комиссии	Рекомендация системы
1	Эксцентриковый механизм перекося для колонн направленного бурения по нефти и газу	3 место	Класс С
2	Разработка интегрированной технологической системы связи предприятия с применением технологии RoIP	Не выиграл	Класс D
3	Разработка программно-аппаратного комплекса для снижения шума установок вентиляции	2 место	Класс В
4	Автоматизированный диагностический комплекс для контроля электрических двигателей самолета	3 место	Класс В
5	Автоматизированная система проектирования локально-вычислительных сетей в жилых и офисных помещениях	1 место	Класс С
6	Двигатель внутреннего сгорания, работающий от газогенератора с принудительным воспламенением	Не выиграл	Класс С
7	Информационно-измерительная система для определения параметров движения механизмов автоматики стрелкового оружия бесконтактным способом	2 место	Класс В
8	Устройство для вибрационного сверления	Не выиграл	Класс D
9	Разработка частотного преобразователя для основного привода лифта	2 место	Класс В
10	Широкополосный коротковолновый SDR модем	1 место	Класс А

На рисунке 6 представлено визуальное сравнение представленных данных решения, по оси Y отложены варианты решений экспертной комиссии и автоматические рекомендации ИАС по каждому проекту в порядке возрастания их уровня.

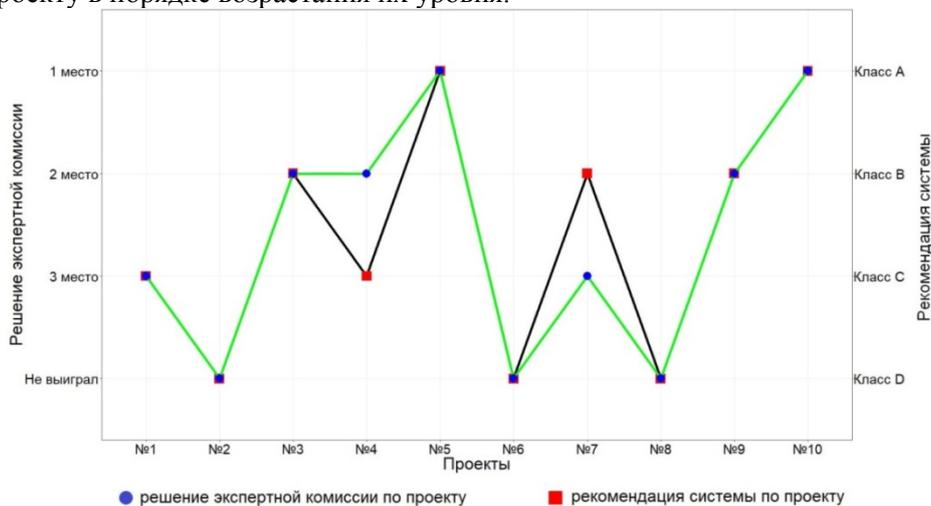


Рисунок 6. Сравнение рекомендаций конкурсной комиссии и экспертного модуля.

Из рисунка 6 видим, что разработанный алгоритм из выборки проектов только в двух случаях и не более чем на один ранг отклонился от результатов оценки экспертной комиссии. В этом случае коэффициент ранговой корреляции Спирмена для приведенного примера равен 91,4%, что говорит о высокой сопоставимости результатов оценки проектов комиссией конкурса и автоматических рекомендаций разработанного аналитического модуля онлайн-платформы.

На сайте бизнес-инкубатора рекомендации модуля интеллектуальной поддержки принятия решения отображаются в форме визуальной подсветки выбранного проекта в списке.

7. Сравнение с аналогами

К наукоемким проектам относятся проекты с высокой степенью научной новизны в применяемых подходах к решению определенной задачи, поэтому методики оценки и отбора таких проектов разрабатываются и используются в различных организационных системах: научных фондах (1), вузах (2), промышленных предприятиях (3), специальных программах поддержки проектов (4). Представленную в данной статье методику обозначим (5).

Сравним различные методики оценки и отбора наукоемких проектов с точки зрения аспектов представленной статьи, а именно по критериям функциональных возможностей методики, а также масштабируемости и автоматизации процессов оценки. Примем следующую шкалу степени соответствия рассматриваемого аналога критерию сравнения: 3 – высокая, 2 – средняя, 1 – низкая. Результаты сравнения аналогов приведены в таблице 3.

Таблица 3. Сравнение методик оценки проектов.

№	Критерий сравнения	Аналог				
		(1)	(2)	(3)	(4)	(5)
1	Возможность масштабировать методику на предметные области за рамками основной сферы деятельности	2	1	1	3	3
2	Возможность одновременной оценки большого количества проектов	1	2	2	3	3
3	Автоматическая оптимизация условий методики на основе накопленных данных	1	1	1	1	3
4	Простота внесения новых условий и параметров в методику	1	2	2	1	3
5	Автоматизация принятия итогового решения	1	1	1	2	3
<i>Сумма баллов:</i>		6	7	7	10	15

Таким образом, разработанные алгоритмы аналитического модуля онлайн-платформы бизнес-инкубатора обеспечивают большую эффективность процессов оценки и отбора наукоемких проектов в сравнении с аналогами.

8. Заключение

В представленной работе описан алгоритм самообучения и оптимизации базы знаний интеллектуальной системы управления наукоемкими проектами, что способствует увеличению скорости обработки данных наукоемких проектов на основе статистического анализа накопленной информации о проекте. Повышение эффективности обработки информации о состоянии проекта на определенный момент времени в условиях слабо формализованной проблемной ситуации и выбор форм ее представления лицу принимающему решение приводит к росту определенности выбора управленческого решения, а также расширению горизонта прогнозирования научного потенциала и реализуемости проекта.

В качестве примера была представлена реализация алгоритма на выборке проектов для целей поддержки принятия решения о финансировании, однако данный подход применим и к обучению интеллектуальной системы под множество других целей, поскольку с учетом скорости работы алгоритма позволяет на каждый запрос формировать рекомендации актуальные заданным условиям и накопленным данным.

9. Литература

- [1] Caron, F. Data management in project planning and control // *Int. J. of Data Science*. – 2015. – Vol. 1(1). – P. 42-57. DOI: 10.1504/IJDS.2015.069050.
- [2] Yuan, Q. The development of a robust resource constrained project scheduling framework / Q. Yuan, Y.E. Polychronakis // *Int. J. of Project Organisation and Management*. – 2012. – Vol.4(4). – P. 339-367. DOI: 10.1504/IJPOM.2012.050329.
- [3] Zarate, P. A new trend for knowledge-based decision support systems design / P. Zarate, Sh. Liu // *International Journal of Information and Decision Sciences (IJIDS)*. – 2016. – Vol. 8(3). – P. 305-324. DOI: 10.1504/IJIDS.2016.078586.
- [4] Tackenberg, S. Activity- and actor-oriented simulation approach for the management of development projects / S. Tackenberg, S. Duckwitz, Ch. Schlick // *Int. J. of Computer Aided Engineering and Technology*. – 2010. – Vol. 2(4). – P. 414-435. DOI: 10.1504/IJCAET.2010.035395.
- [5] Bukchin, Y. A multi-objective approach for decision making during the project life cycle / Y. Bukchin, Sh. Rozenes // *Int. J. of Project Organisation and Management*. – 2011. – Vol. 3(2). – P. 184-203. DOI: 10.1504/IJPOM.2011.039820.
- [6] Muhlemann, A.P. Portfolio modeling in multiple-criteria situations under uncertainty / A.P. Muhlemann, A.G. Lockett, A.E. Gear // *Decision Sciences*. – 1978. – Vol. 9. – P. 612-626. DOI: 10.1111/j.1540-5915.1978.tb00749.x.
- [7] Daradkeh, M. Information visualisation for decision support under risk // *Int. J. of Information and Decision Sciences*. – 2017. – Vol. 9(3). – P. 276-296. DOI: 10.1504/IJIDS.2017.10007807.
- [8] Burke, R. Project Organization Structures / R. Burke, S. Barron // *Project Management Leadership*. – 2015. DOI: 10.1002/9781119207986.ch4.
- [9] Chinowsky, P. Networks in engineering: an emerging approach to project organization studies / P. Chinowsky, J.E. Taylor // *Engineering Project Organization Journal*. – 2012. – Vol. 2(1). – P. 15-26. DOI: 10.1080/21573727.2011.635647.
- [10] Liu, Q. Scheduling optimization of design stream line for production research and development projects / Q. Liu, X. Geng, M. Dong, Ch. Ye // *Engineering Optimization*. – 2017. – Vol. 49(5). – P. 896-914. DOI: 10.1080/0305215X.2016.1216113.
- [11] Ballard, G. Lean management methods for complex projects / G. Ballard, I. Tommelein // *Engineering Project Organization Journal*. – 2012. – Vol. 2(1). – P. 85-96. DOI: 10.1080/21573727.2011.641117.
- [12] Fengjie, T. Shape optimization based design of arch-type dams under uncertainties / T. Fengjie,

- T. Lahmer // *Engineering Optimization*. – 2018. – Vol. 50(9). – P. 1470-1482. DOI: 10.1080/0305215X.2017.1409348.
- [13] Jia, J. Designing train-speed trajectory with energy efficiency and service quality / J. Jia, K. Yang, L. Yang, Y. Gao, Sh. Li // *Engineering Optimization*. – 2018. – Vol. 50(5). – P. 797-818. DOI: 10.1080/0305215X.2017.1358712.
- [14] Malekpoor, H. Integrated grey relational analysis and multi objective grey linear programming for sustainable electricity generation planning / H. Malekpoor, K. Chalvatzis, N. Mishra, M.K. Mehlawat, D. Zafirakis, M. Song // *Annals of Operations Research*. – 2018. – Vol. 269(1). – P. 475-503. DOI: 10.1007/s10479-017-2566-4.
- [15] Roben, S. A design approach for asset supply logistics / S. Roben, J.A. Ottjes, G. Lodewijks, A. Dool // *Int. J. of Computer Aided Engineering and Technology*. – 2010. – Vol. 2(4). – P. 311-323. DOI: 10.1504/IJCAET.2010.035388.
- [16] Kumar, A. A case study on machine learning and classification / A. Kumar, B.K. Sarkar // *Int. J. of Information and Decision Sciences*. – 2017. – Vol. 9(2). – P. 179-208. DOI: 10.1504/IJIDS.2017.10005873.
- [17] Shi, L. Learning to predict characteristics for engineering service projects / L. Shi, L. Newnes, S. Culley, B. Allen // *AI EDAM*. – 2017. – Vol. 31(3). – P. 313-326. DOI: 10.1017/S0890060417000129.
- [18] Quaglio, M. A model-based data mining approach for determining the domain of validity of approximated models / M. Quaglio, E.S. Fraga, E. Cao, A. Gavriilidis, F. Galvanin // *Chemometrics and Intelligent Laboratory Systems*. – 2018. – Vol. 172. – P. 58-67. DOI: 10.1016/j.chemolab.2017.11.010.
- [19] Xing, H. Rapid development of knowledge-based systems via integrated knowledge acquisition / H. Xing, S.H. Huang, J. Shi // *AI EDAM*. – 2003. – Vol. 17(3). – P. 221-234. DOI: 10.1017/S0890060403173052.
- [20] Sakama, C. An abductive framework for computing knowledge base updates / C. Sakama, K. Inoue // *Theory and Practice of Logic Programming*. – 2003. – Vol. 31(3). – P. 671-715. DOI: 10.1017/S1471068403001716.
- [21] Gohari, F.S. A new confidence-based recommendation approach: Combining trust and certainty / F.S. Gohari, F.S. Aliee, H. Haghighi // *Information Sciences*. – 2018. – Vol. 422. – P. 21-50. DOI: 10.1016/j.ins.2017.09.001.
- [22] Blagodatsky, G.A. Modelling the system of fuzzy logical inference for evaluating science-intensive projects / G.A. Blagodatsky, M.M. Gorokhov, D.A. Perevedentsev // *Automation of control processes*. – 2017. – Vol. 2(48). – P. 84-91. (in Russian).
- [23] Gorokhov, M.M. Designing the knowledge base of the expert system for managing scientific projects / M.M. Gorokhov, D.A. Perevedentsev // *Bulletin of Tomsk State University. Management, Computer Science and Informatics*. – 2017. – Vol. 1(38). – P. 47-51. (in Russian).
- [24] Han, H. Variable selection using mean decrease accuracy and mean decrease Gini based on random forest / H. Han, X. Guo, H. Yu // *Software Engineering and Service Science (ICSESS): 7th IEEE International Conference, 2016*. – P. 219-224.
- [25] Официальный сайт бизнес-инкубатора «ФГБОУ ВПО ИжГТУ имени М.Т. Калашникова» [Electronic resource]. – Access mode: <http://innobinc.ru>.

Благодарности

Работа выполнена в рамках гранта ФГБОУ ВО «ИжГТУ имени М.Т. Калашникова» № 09.04.02/18ГММ.

Method of knowledge base training of intellectual real - time system based on the algorithm of decision tree

G.A. Blagodatsky¹, S.V. Vologdin¹, M.M. Gorohov¹, D.A. Perevedencev¹

¹Izhevsk State Technical University, Studencheskaja 7, Izhevsk, Russia, 426069

Abstract. The development of a mathematical algorithm for the self-learning of the intellectual system (IS) for the management of science-intensive projects for the creation of complex technical objects is considered. IS training occurs by optimizing the knowledge base rules using machine learning methods. This will increase the effectiveness of decision support in the implementation of science-intensive projects and the adaptation of the recommendations of the IS on the request of the decision-maker. Based on the analysis and markup of the available data, as well as the formulated requirements for the algorithm, the algorithm of regression and classification (CART) trees is chosen because of its high speed and the possibilities of formalizing the obtained dependencies in the format of production rules. In accordance with the requirements set, the training parameters were determined and the efficiency of the algorithm for classifying scientific projects for a given number of classes and for formulating new dependencies of the knowledge base was checked. A general scheme for processing the user's request with the described functionality of the intelligent module is presented.