

МНОГОЭКСПЕРТНАЯ СИСТЕМА ПЛАНИРОВАНИЯ РЕМОНТНЫХ ПРОГРАММ ЭНЕРГЕТИЧЕСКОГО ОБОРУДОВАНИЯ

Бородин В.А.¹, Андришин А.В.², Проталинский О.М.², Ханова А.А.¹

¹*Астраханский государственный технический университет*

²*Московский энергетический институт*

vitkt@yandex.ru, andriushinav@mpei.ru, protalinskiy@gmail.com, akhanova@mail.ru

Аннотация: В статье рассмотрена проблема планирования ремонтной программы на крупных энергетических предприятиях, предложена многоэкспертная система поддержки принятия решений. Рассмотрены два основных подхода для получения знаний, построена модель классификации. Предложены различные подходы взаимодействия базы знаний и исторических данных. Разработан алгоритм взаимодействия между базой продукционных правил и системой предиктивного анализа, проведён эксперимент на классической задаче классификации.

Ключевые слова: многоэкспертные системы, предиктивный анализ, машинное обучение, энергетика, ремонтная программа.

Введение

В настоящее время энергетика является одной из самых высокоёмких отраслей промышленности. Отказы оборудования в этой отрасли приводят к серьёзным экономическим потерям и отрицательным социальным последствиям. Важную роль в снижении этого фактора играют системы управления производственными активами энергетических предприятий [1]. Основной целью функционирования таких систем является обеспечение заданного уровня надёжности или производственных рисков при удовлетворении экономических требований. Одной

из основных мер сокращения риска отказов оборудования является оптимальное планирование и реализация ремонтных программ. Высокая стоимость оборудования и широкое распространение энергетических технологий приводит к существенному сокращению затрат, что даёт значительный экономический эффект. Техническим последствием оптимизации ремонтных программ является обеспечение заданной надёжности, социальным – уменьшение рутинного труда и повышения квалификации персонала.

При планировании ремонтных программ источником нормативной информации являются технологические карты, представляющие организационно-технологический документ, содержащий данные о затратах на проведение работ с целью ликвидации дефектов оборудования: количество материалов, необходимость использования машин и механизмов, квалификацию и объёмы трудовых ресурсов [2].

Соответствие дефекта оборудования технологическим картам зависит от признаков оборудования, места установки и других параметров. Для принятия решения разработчиком ремонтной программы необходимо знать соответствие между признаками объекта планирования и технологическими картами. Некоторые системы управления производственными активами позволяют заносить подобные прямые соответствия между. Однако для крупных энергетических предприятий, где количество оборудования превышает 30000 единиц, технологических карт – 12000, а типовых дефектов более 900, неавтоматизированный процесс создания ремонтных программ является малоэффективным: требует реализации большого количества правил работы с организационно-технологическими документами и во многом зависит от квалификации и опыта разработчика.

Эти причины делают актуальным создание системы поддержки принятия решений (СППР) для составления ремонтной программы энергетического оборудования. Целью такой системы является повышение эффективности процесса планирования, за счет оптимизации ремонтных программ и удовлетворения экономических и технических показателей.

1 Концепция многоэкспертной системы

В качестве прототипа предлагается использовать многоэкспертную систему, включающую в себя продукционную базу знаний, построенную на основе знания и опыта экспертов, и статистическую обработку накапливаемой информации, на основе предиктивного анализа. Сочетание этих компонентов позволяет учесть возможную неполноту и некорректность экспертных правил, как результат влияния человеческого фактора, и недостаток исторических данных, собранных во время эксплуатации оборудования. Таким образом, задачу определения способа сопоставления технологических карт обнаруженным дефектам предлагается решать на основе предыдущих сопоставлений: накопленные в результате эксплуатации исторические данные, и экспертных знаний.

Достаточный набор статистических данных позволяет достичь высокой точности решения, обучение классификаторов производится существенно быстрее, чем формализация продукционных правил экспертами. При этом качество классификаторов сильно зависит от объёма выборки и достаточного количества признаков – отсутствие данных приведет к серьезному снижению точности классификации. Накопление достаточного объёма выборки исторических данных может занимать достаточно продолжительное время. Экспертная информация не требует продолжительного сбора и может быть основана на нормативных и внутрикорпоративных документах, а также на знаниях и опыте эксперта. База продукционных правил требует постоянной актуализации, в то же время статистическая выборка более гибка: переобучение может осуществляться при накоплении нового набора данных. Реализацию статистического подхода предлагается осуществить с использованием предиктивного анализа на основе машинного обучения, а экспертной информации в виде базы продукционных правил на основе известных методов обработки экспертных знаний.

Задача сопоставления технологических карт дефектам может быть описана в виде концептуальной модели:

$$(1) \quad Tcrd = f(Dfs, Attrs_1 \dots Attrs_n),$$

где $Dfs \subseteq \{Df_1, Df_2 \dots Df_m\}$ – подмножество дефектов, выявленных у единицы оборудования, принадлежит множеству типовых дефектов; $Tcrd = f(Dfs, Eq)$ – подмножество технологических карт для устранения данных дефектов; $\forall Attrs_n \in \{Attr_n\}$ – аргументы функции, соответствующие значимым для сопоставления атрибутам. Каждый аргумент принадлежит своему множеству

допустимых значений: $Eqt \in \{Eqt_1, Eqt_2 \dots Eqt_n\}$ – тип оборудования принадлежит множеству типов оборудования. В качестве атрибутов используются характеристики оборудования или дефекта: географические координаты расположения оборудования, время обнаружения дефекта и другие.

2 Подсистема предиктивного анализа

Сформулированная выше задача сопоставления технологических карт может быть представлена как задача классификации с пересекающимися классами (multi-label classification), для решения которой используют подход «один против всех» (one-vs-rest). При таком подходе для каждого класса создаётся свой бинарный классификатор (рис. 1).

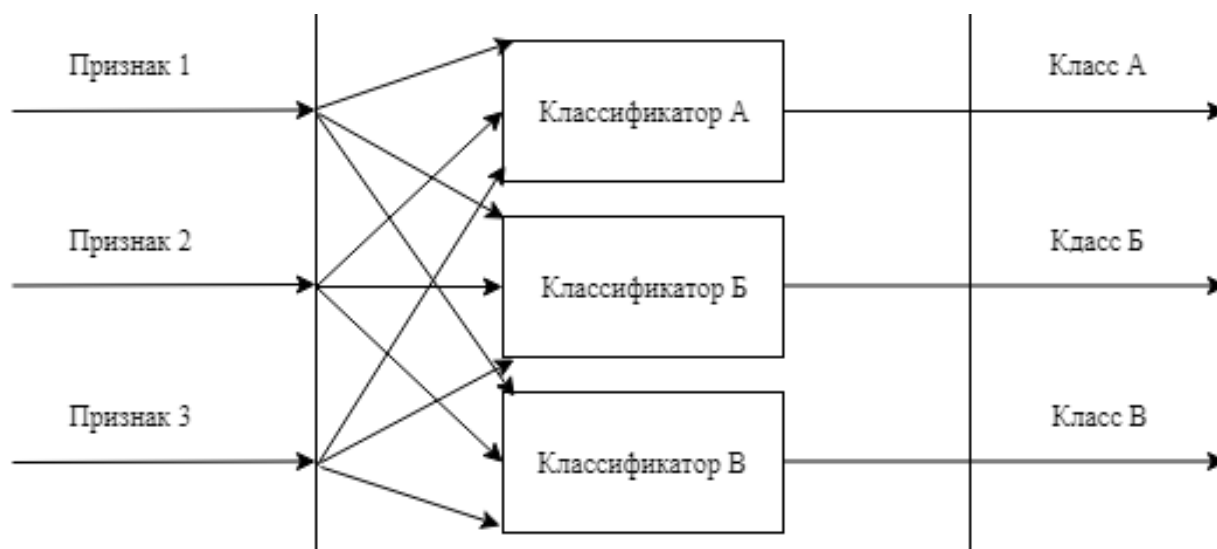


Рис.1. Задача классификации с непересекающимися классами

Идентификация математической модели осуществлялась на основе данных о плановых показателях ремонтной программы одного из филиалов «МРСК Северного Кавказа» - 7537 заказа, ремонтной программы, 137 уникальных дефектов, 730 уникальных технологических карт, 53 типа оборудования. Данные включали в себя информацию о единице оборудования, наборе дефектов в заказе и назначенных технологических картах. В первом эксперименте использованы только наборы дефектов и технологических карт. Множества дефектов и технологических карт упорядочены, а заказы представлены в виде логической матрицы. Каждая строка матрицы содержит информацию о заказе: первые N столбцов матрицы соответствуют дефектам, следующие M столбцов - технологическим картам. Значение «истина» соответствует наличию данного дефекта или технологической карты в заказе, «ложь» - отсутствию. Данные заказов на ремонт разбиты на тестовую и обучающую выборки. Для проверки возможности автоматического обучения протестировано несколько классификаторов – деревья принятия решений, байесовский классификатор, многослойный персептрон (число скрытых слоёв – 50). Для каждой технологической карты создан классификатор, обученный на заказах для определения необходимости использования технологической карты при данном наборе дефектов. По результатам классификации вычислены различные метрики – доля, верно, классифицированных объектов, точность, полнота, F-мера (гармоническое среднее между точностью и полнотой). Для обобщения результатов всех классификаторов выбрана концепция микро-усреднения (значение метрик всех классификаторов усредняются).

Все классификаторы показали высокий процент правильно классифицированных объектов – около 80-90%. Значение остальных метрик оказалось низким (рис. 2).

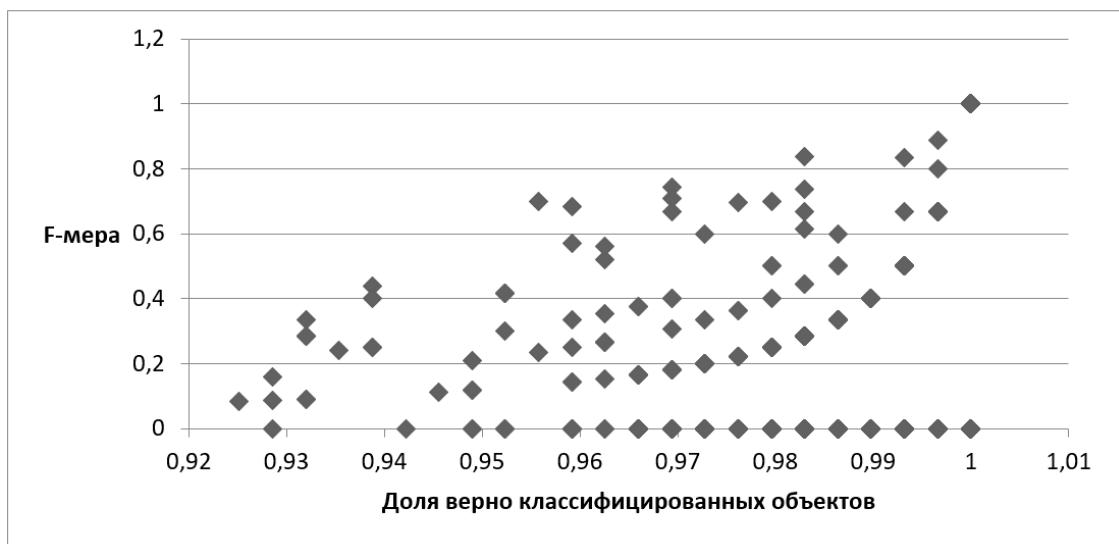


Рис 2. Доля верно классифицированных объектов и F-мера для различных технологических карт в выборке (деревья принятия решений)

На следующем этапе в обучающий набор добавлена информация о типе оборудования. Значения метрик возросли, но тенденция сохранилась: высокое значение верно классифицированных объектов при низком значении точности, полноты и F-меры. Для более подробного анализа произведено обучение на всей выборке. Обучение по всей выборке не позволило оценить точность классификатора из-за эффекта переобучения, но в данном случае необходимо выяснить причину разницы между метриками. Средние значения возросли, но присутствовали классификаторы некоторых технологических карт, где точность и полнота остались нулевыми. При подробном анализе найден 441 заказ, для которых одинаковый набор дефектов приводит к разным наборам технологических карт.

Исторических данных, зарегистрированных системой учёта, оказалось недостаточно для эффективного решения задачи. Помимо проблемы с конфликтующими результатами прослеживается зависимость качества классификации технологической карты от количества заказов, в которые она входит (рис 3).

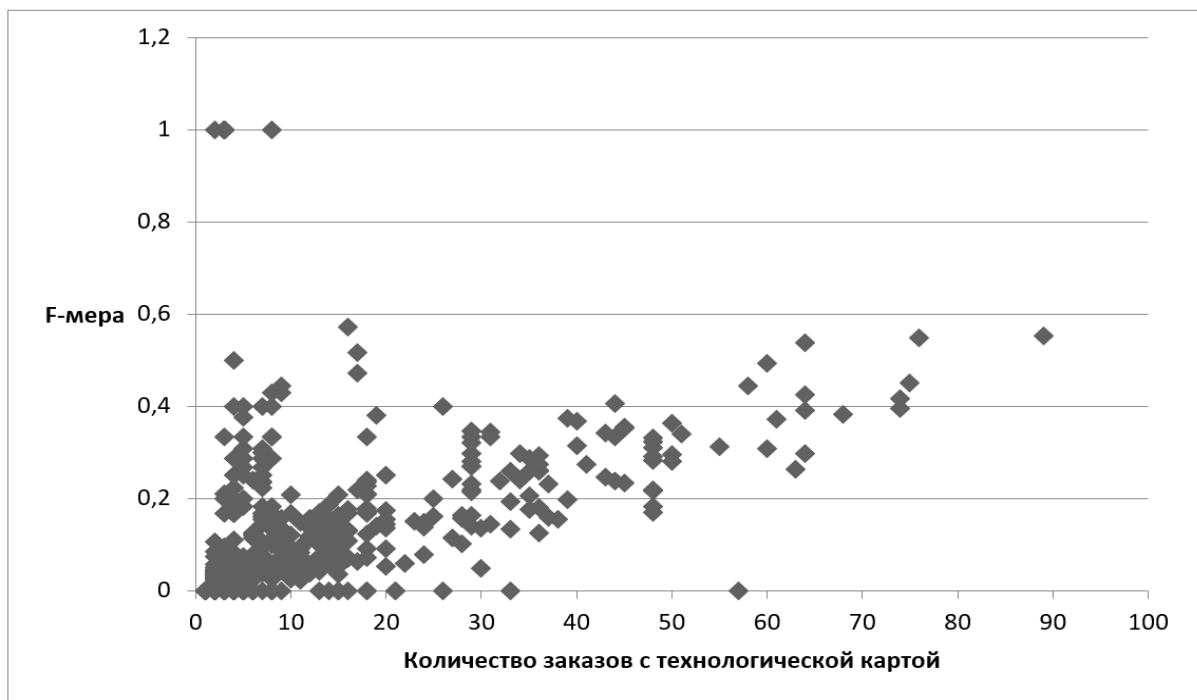


Рис. 3. Зависимость между F-мерой и количеством заказов с технологической картой (байесовский классификатор)

3 Взаимодействие продукционной базы и подсистемы предиктивного анализа

Можно предположить, что для конфликтующих заказов существует какой-то атрибут, позволяющий верно классифицировать набор дефектов. Например, в зависимости от особенностей установки оборудования могут быть необходимы дополнительные работы по демонтажу\монтажу устройства. С другой стороны, для некоторых технологических карт значения метрик оказались высокими. Для решения проблемы предложена концепция объединения двух подходов – машинного обучения и базы продукционных правил. Объединение дополняет подходы – с одной стороны, в случае недостатка статистических данных система может пользоваться экспертными данными, с другой – неточности в правилах могут быть исправлены машинным обучением. Возможны разные способы взаимодействия подходов: использование результата работы классификаторов в качестве части продукционной базы знаний, импорт знаний из классификаторов в базу, расширения обучающей выборки классификаторов с помощью экспертных знаний, представление базы знаний как классификатора (многоэкспертная система).

В разработанной системе поддержки принятия решений использован смешанный подход. Взаимодействие базы продукционных правил и машинного обучения регламентируется ещё одной экспертной системой (метаправила). В этом случае возникает проблема приоритизации результатов. Если пользователь системы создал базу из утверждённых регламентных документов, тогда в метаправилах должен быть приоритет экспертных знаний перед машинным обучением. При отсутствии правил для определённого набора дефектов система использует результаты работы классификаторов. Если приоритетна статистика, то в первую очередь могут вызываться классификаторы. В этом случае при обучении необходимо находить конфликтующие наборы данных. При попадании таких данных на вход системы используется информация из базы знаний, в остальном случае – из классификаторов. Этот вариант может использоваться при риске недостаточной компетентности эксперта. Так же возможна стратегия, при которой база продукционных правил изначально не содержит знаний, система выдаёт решения из модуля машинного обучения. В случае конфликта данных пользователь добавляет новое правило в базу знаний.

Если существует достаточное количество статистики для работы классификаторов и база продукционных правил обладает большим объёмом знаний, результат её работы представляется как результат классификатора. В машинном обучении известен ряд алгоритмов (комитет машин), объединяющих результаты нескольких классификаторов [3-8]. Они делятся на статические и динамические. В объединении результатов статические алгоритмы не участвуют входные данные в отличие от динамических. К статическим алгоритмам относят ансамблевые методы и бустинг. Ансамблевые методы предполагают независимое обучение классификаторов и последующее объединение результатов. Результаты могут быть усреднены, выбраны голосованием. Существуют различные вариации таких алгоритмов, например, включающие присваивание классификаторам весовых коэффициентов. Алгоритм бустинга основан на идее последовательного построения композиции слабых классификаторов. В этом случае каждый следующий алгоритм стремится компенсировать недостатки всех предыдущих. Динамические методы, такие как mixture of experts, предполагают использование входных данных при формировании решения об окончательном результате. Mixture of experts использует нейронную сеть, которая обучается одновременно с классификаторами. Результаты работы этой сети используются при формировании результата. В разработанной системе поддержки принятия решений база знаний объединена с классификаторами взвешенным голосованием.

Предиктивный анализ и базу знаний можно использовать не только в качестве двух изолированных систем. Знания классификаторов, возникшие в результате обучения, можно импортировать в базу правил. Некоторые классификаторы имеют объяснительную способность. Так деревья принятия решений хранят информацию в виде дерева логических правил, сформированных на основе данных по определённому критерию (энтропия, индекс Джини). На рис. 4 изображена часть дерева принятия решений, обученного для одной из технологических карт. Каждому узлу дерева соответствует условие, индекс Джини и количество сэмплов. Основное условие в данной задаче – наличие или отсутствие какого-либо дефекта в заказе.

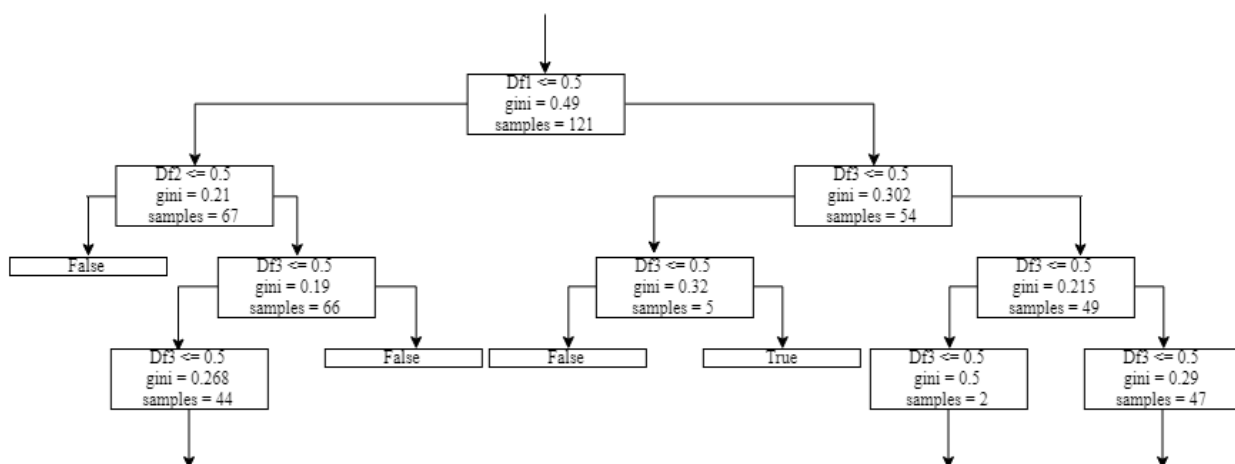


Рис. 4. Дерево принятия решений

При данном подходе возможен импорт правил из дерева принятия решений в базу продукционных правил. Импорт может происходить в ручном или автоматическом режиме. При ручном режиме пользователю предлагается набор правил. Он может переносить их из дерева принятия решений в базу знаний. Возможны алгоритмы автоматического переноса либо слияния правил. Список технологических карт формирует экспертная система, обновлённая правилами из классификатора. Использование метода затрудняет то, что данные в деревьях принятия решений не всегда имеют понятный пользователю вид, так как вычислены с помощью статистики. Правила могут быть ясны в случае категориальных признаков в отличие от признаков, выраженных целыми или действительными числами. Поэтому данный подход реализован в системе поддержки принятия решений как аналитический инструмент – пользователь может просматривать обученные деревья принятия решений и изменять продукционные правила в соответствии с ними.

Кроме того, реализована проверка соответствия экспертных знаний и машинного обучения. В этом случае весь набор исторических данных проходит через правила базы знаний экспертной системы. Пользователь информируется о несовпадении результатов, после чего он может вносить коррективы в правила либо пометить исторические данные как неактуальные.

Пользователь системы может редактировать правила базы знаний, редактировать базу со статистикой (в том числе пометить данные как неактуальные и добавлять новые данные, полученные в результате работы системы), запускать процесс переобучения, просматривать структуру деревьев принятия решений.

Таблица 1. Метаправила системы, объединяющей базу продукционных правил и машинное обучение

Условие	Решение
При обучении классификаторов обнаружены конфликтующие данные	Данные исключаются из обучающей выборки, при поступлении на вход выполняется вызов продукционных правил
База знаний не содержит правил для входных данных	Используется результат классификаторов
База статистики не содержит достаточного количества данных	Используется результат работы продукционных правил
Обе подсистемы имеют решение для входного набора	Используется ансамблевый метод для объединения результатов
Решение одобрено пользователем либо сохранено скорректированное решение	Результат сопоставления добавляется в обучающий набор.

4 Постановка вычислительного эксперимента

Ввиду отсутствия базы знаний для набора данных заказов принято решение проверить модель на известной задаче классификации. Одна из классических задач – ирисы Фишера. Это набор данных для задачи классификации, 50 экземпляров из трёх видов — Ирис щетинистый (*Iris setosa*), Ирис виргинский (*Iris virginica*) и Ирис разноцветный (*Iris versicolor*). Для каждого экземпляра измерялись четыре характеристики (в сантиметрах): длина наружной доли околоцветника (англ.

sepal length); (англ. sepal width); длина внутренней доли околоцветника (англ. petal length); ширина внутренней доли околоцветника (англ. petal width). Этот набор данных часто используется для тестирования классификаторов. На рис.5 изображён набор данных, оси X и Y соответствуют первым двум признакам – ширине наружной доли околоцветника и длине наружной доли околоцветника. Различные фигуры – круг, треугольник и квадрат – трём видам цветов.

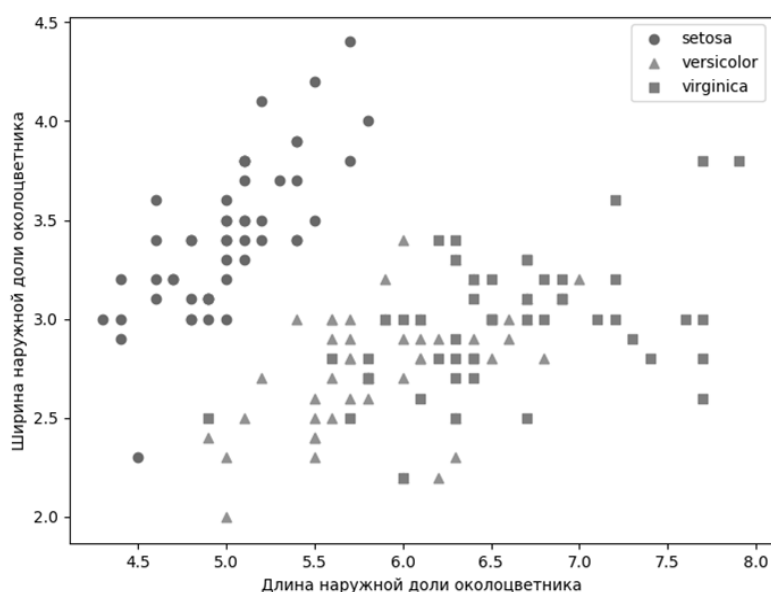


Рис 5. Визуализация набора данных об ирисах

Входные данные разбиты на обучающий и тестовый наборы. Для имитации конфликтующих данных при обучении классификаторов использованы только первые два признака. В данных присутствуют сэмплы, где при одинаковых длине и ширине наружной доли околоцветника результатом являются разные виды ириса. Конфликтующие данные исключены из обучающей и тестовой выборки, учитываются в метриках как верно рассчитанные (имитация передачи конфликтующих данных в экспертную систему). Важным вопросом является имитация базы продукционных правил. Наиболее близкий к продукционной модели классификатор – дерево принятия решений. Для воссоздания неполных, но точных данных, дерево обучено на всех четырёх признаках. Для трети сэмплов обучающей выборки решения от дерева не принимались (имитация отсутствия правил для части данных). Весами классификаторов являлись значения доли верно классифицированных объектов, для «базы знаний» вес принимался за единицу. Эксперимент повторён 100 раз, результаты усреднены (табл. 2.).

Таблица 2. Средний процент, верно, классифицированных сэмплов при имитации системы

Классификатор	Значение
Байесовский классификатор (вместе с конфликтующими сэмплами)	0.7694
Многослойный перцептрон (вместе с конфликтующими сэмплами)	0.7044
Байесовский классификатор (обучение без конфликтующих сэмплов)	0.8210
Многослойный перцептрон (обучение без конфликтующих сэмплов)	0.8151
«База знаний» (обучение с конфликтующими сэмплами с использованием 4 признаков)	0.9566
«База знаний» (с исключением трети выборки)	0.6973
Байесовский классификатор (конфликтующие сэмплы включены в результат как, верно, классифицированные)	0.8489
Многослойный перцептрон (конфликтующие сэмплы включены в результат как, верно, классифицированные)	0.8447
Вся система	0.8884

Заключение

Исследование показало высокую эффективность объединения методов машинного обучения и базы продукционных правил. Недостатки предиктивного анализа при наличии конфликтующих

сэмплров или отсутствии достаточного объёма данных восполняются использованием правил продукционной модели. В процессе эксплуатации системы происходит накопление исторических данных, что улучшает эффективность классификаторов. Возможные ошибки в правилах экспертной системы выявляются с помощью процесса сопоставления исторических данных результатам работы подсистемы логического вывода. Одновременное использование базы знаний и классификаторов в ансамблевом методе повышает эффективность классификации. В дальнейшем функционал системы может быть расширен другими способами взаимодействия – автоматическим импортом правил из деревьев принятия решений в экспертную систему или более эффективным алгоритмом объединения результатов.

Литература

1. *Проталинский О.М., Проталинский И.О., Кладов О.Н.* Система оптимального управления производственными активами энергетических предприятий // Автоматизация и ИТ в энергетике. 2017. №4 (93).
2. *Яцура А.И.* Система технического обслуживания и ремонта энергетического оборудования. Справочник. М.: НЦ ЭНАС, 2006.
3. *Т.К. Ho, J.J. Hull, and S.N. Srihari,* “Decision Combination in Multiple Classifier Systems,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 1, pp. 66-75, Jan. 1994.
4. *Kittler J., Alkoot F.M.* Sum versus vote fusion in multiple classifier systems // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003. V. 25, N 1. P. 110-115.
5. *Abuomman A.A., Reaz M.B.J.* A survey of intrusion detection systems based on ensemble and hybrid classifiers // Computers k, Security. 2016. P. 1-45.
6. *Xu L., Amari Shun-ichi.* Combining classifiers and learning mixture-of-experts // Encyclopedia of Artificial Intelligence. 2009. P. 319-326.
7. *Rutkowska D.* (2004) Multi-expert Systems. In: Wyrzykowski R., Dongarra J., Paprzycki M., Waśniewski J. (eds) Parallel Processing and Applied Mathematics. PPAМ 2003. Lecture Notes in Computer Science, vol 3019. Springer, Berlin, Heidelberg
8. *Аведьян Э.Д., Ле Тхи Чанг Линь* Процедуры оптимального голосования в многоэкспертных бинарных системах // Труды МФТИ. 2017. №4 (36).