

МЕТОДЫ И АЛГОРИТМЫ КОЛЛЕКТИВНОГО РАСПОЗНАВАНИЯ: ОБЗОР

В. И. Городецкий¹, С. В. Серебряков²

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

¹<gor@iias.spb.su>, ²<sergey_s@iias.spb.su>

УДК 681.3

Городецкий В. И., Серебряков С. В. **Методы и алгоритмы коллективного распознавания: обзор** // Труды СПИИРАН. Вып. 3, т. 1. — СПб.: Наука, 2006.

Аннотация. Под коллективным распознаванием понимается задача использования множества классификаторов, каждый из которых принимает решение о классе одной и той же сущности, ситуации, образа и т.п., с последующим объединением и согласованием решений отдельных классификаторов с помощью того или иного алгоритма. В настоящее время это направление в области распознавания образов и классификации, которое, с одной стороны, зарекомендовало себя как новый шаг в данной области, и которое, с другой стороны, находит все более и более широкое применение в решении сложных крупномасштабных прикладных задач, является предметом активных теоретических и прикладных исследований. Данная работа посвящена обзору состояния исследований в упомянутой области начиная с первых работ, которые относятся еще к 1950-м годам, и заканчивая самыми новыми результатами. — Библ. 70 назв.

UDC 681.3

Gorodetsky V. I., Serebryakov S. V. **Methods and algorithms of the collective recognition: a survey** // SPIIRAS Proceedings. Issue 3, vol. 1. — SPb.: Nauka, 2006.

Abstract. Collective recognition is a task in which multiple classifiers are used for making decisions concerning the class of the same entity (object, situation, etc.) with the subsequent combining and agreement of individual decisions based on some special algorithm. Currently this research direction in the pattern recognition and classification scope is of topmost interest within information technology community. The reason of this fact is that methods and techniques of collective recognition demonstrate new capabilities in regard to pattern recognition and classification accuracy, on the one hand, and they are increasingly used in complex large scale applications. This paper surveys state of the art of the research on the scope in question covering the period from 1950-th and till now. — Bibl. 70 items.

1. Введение: история, мотивация и потенциальные преимущества коллективного распознавания

Под коллективным распознаванием понимается задача использования множества классификаторов, каждый из которых принимает решение о классе одной и той же сущности, ситуации, образа и т.п. с последующим объединением и согласованием решений отдельных классификаторов с помощью того или иного алгоритма.

Задача коллективного распознавания относится к более общей проблеме коллективного принятия решений, которая имеет большое разнообразие форм и приложений. В частности, к задачам коллективного принятия решений относятся задачи группового выбора [6, 7, 18, 37] и др., задачи теории голосования [5, 6, 7, 37] и др., методы обработки экспертных оценок [1, 3, 4, 7] и др.¹ Существование задачи коллективного принятия решений состоит в «выработке согла-

¹Здесь даны ссылки, взятые из работы [8].

сованного коллективного решения о порядке предпочтения рассматриваемых объектов на основе индивидуальных мнений членов коллектива» [8].

Как видно из приведенных ссылок¹, проблема коллективного принятия решений, в частности, проблема коллективного распознавания, исследуется уже давно. Однако «рекордно ранней» работой по данному направлению, по-видимому, является работа [25]. Хотя в современном виде проблема коллективного распознавания была поставлена впервые, по-видимому, в работах [9, 10, 11, 12], активный интерес к ней со стороны зарубежных исследователей и

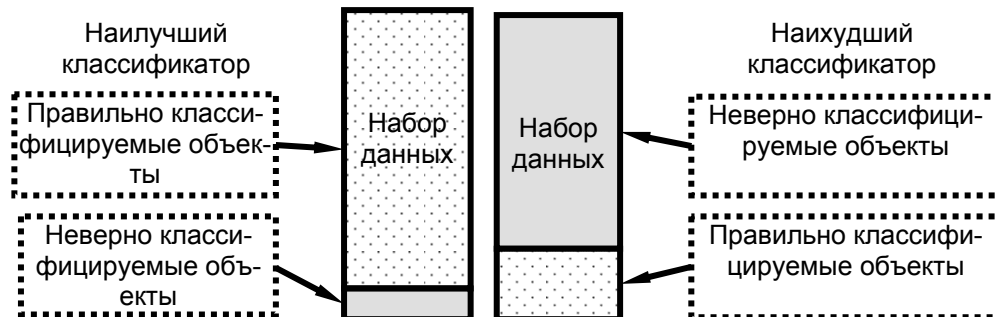


Рис. 1. Простейшая мотивация применения подхода основанного на использовании множества классификаторов: Даже наихудший классификатор может внести свой вклад в улучшение свойств системы классификации в целом.

соответственно, активные исследования по ней начались примерно лишь через 10–15 лет. Примерно с 1988 года и по настоящее время проблема коллективного распознавания является объектом все возрастающего интереса и постоянно увеличивающегося числа публикаций.

В современной зарубежной литературе задачи, методы и алгоритмы коллективного распознавания встречаются в различных работах под разными названиями [38, 43]:

- объединение множества классификаторов (combination of multiple classifiers);
- объединение классификаторов (classifier fusion);
- объединение экспертов (mixture of experts);
- комитеты (committees);
- согласованная агрегация (consensus aggregation);
- голосующее объединение классификаторов (voting pool of classifiers);
- динамический выбор классификатора (dynamic classifier selection);
- комбинированные системы классификаторов (composite classifier system);
- комбинирование решений (decision combining);
- классификаторы типа «разделяй и властвуй» (divide-and-conquer classifiers).

В действительности, приведенное разнообразие используемой терминологии отражает также и разнообразие постановок задач, предположений, типы выходов классификаторов, стратегии объединения и т.п.

Необходимость использования множества классификаторов с последующим объединением их решений мотивируется по-разному в зависимости от

¹Здесь специально выбраны самые ранние ссылки по отмеченным проблемам, хотя более новых ссылок можно указать гораздо больше.

конкретной постановки задачи, ее свойств и конкретного приложения. В одних приложениях это обуславливается эмпирически, когда таким образом установлено, что использование множества классификаторов приводит к более высокой точности распознавания и лучшим показателям вычислительной эффективности. В других приложениях это объясняется специфичностью задачи. Типичными в этом отношении являются случаи, когда данные, используемые для распознавания, собираются множеством распределенных сенсоров; когда имеется естественная декомпозиция источников данных; когда пространство признаков содержит весьма разнообразные типы признаков (булевы, целочисленные, непрерывные, текстовые, изображения, и т.п.); когда решаемая задача обладает большой размерностью пространства признаков и необходимо бороться с вычислительной сложностью и т.д.

Идея объединения классификаторов для улучшения точности и качества работы системы классификации имеет глубокую аналогию с историческим опытом человека: с древних времен, любой политический лидер независимо от режима и политической системы привлекал различные группы людей, чтобы разделить обязанности и области ответственности. В этой аналогии мотивацией является желание уменьшить сложность управления государственными делами и увеличить компетентность благодаря специализации конкретных административных групп. Эти две идеи, а именно, *уменьшение сложности* решаемой задачи и *увеличение компетентности* принятия решений, являются основными движущими силами и главной мотивацией использования согласованного объединения решений множества классификаторов, специализированных в том или ином смысле. «Даже в том случае, когда один из классификаторов обладает существенно лучшими свойствами по сравнению с другими, множества неправильно классифицированных объектов у различных классификаторов не обязательно будут пересекаться. ... По этой причине разные классификаторы потенциально могут обеспечивать различной информацией о классифицируемом объекте, что может оказаться существенным для улучшения свойств системы в целом» [38]. Графическое объяснение последнего факта представлено на рис. 1.

Наиболее ранним результатом в теории объединения решений является уже ранее упоминавшаяся работа [25]¹, в которой сформулирована так называемая теорема Кондорсе о присяжных. Эта теорема утверждает, что если каждый член жюри присяжных обладает независимым мнением по поводу рассматриваемой проблемы, и если вероятность того, что любой из них принимает правильное решение, будет больше 0.5, то тогда вероятность того, что решение присяжных в целом окажется правильным, возрастает с увеличением количества членов жюри, и стремится к единице. Если же вероятность быть правым у каждого из членов жюри меньше 0.5, то вероятность принятия правильного решения присяжными в целом монотонно уменьшается и стремится к нулю с увеличением количества присяжных.

В 1980-х годах, когда распознавание образов и классификация достигли зрелости, были обнаружены дополнительные преимущества использования множества классификаторов [8, 19]. В частности, к этому времени было разработано и практически проверено много различных методов распознавания и классификации: статистические методы; синтаксические методы, включающие формализацию модели в терминах из теории автоматов; методы теории фор-

¹ Ссылка и содержание теоремы были взяты из работы [34].

мальных грамматик; структурные методы, использующие модели теории графов; нейросетевые подходы; методы, построенные на основе теории машинного обучения и т.п. Разработчики, обладая множеством различных подходов, столкнулись с трудностью выбора среди них в ситуациях, когда нет очевидного лидера, когда каждый подход должен быть оценен в контексте конкретного класса приложений со множеством ограничений. Можно утверждать, что уже в 1980-х годах в области методов распознавания объектов и классификации наметился определенный кризис. Этот кризис был обусловлен тем, что к тому времени отсутствовали новые конструктивные идеи, которые могли бы существенно улучшить качество работы систем классификации и распознавания, а практические потребности в таких идеях постоянно возрастали. По существу был осознан простой, однако очень важный факт, суть которого состоит в том, что не существует модели классификации, которая была бы пригодна для всех приложений. Эта проблема была одной из причин, по которой исследователи пытались экспериментальным путем выбирать наиболее подходящий подход. В конце концов это привело к пониманию преимущества одновременного использования множества разнообразных классификаторов и к необходимости согласованного объединения их решений по некоторому алгоритму. Что касается алгоритма объединения решений, то оказалось, что наилучшего варианта не существует и здесь. Последний вывод стимулировал исследования в области методов и алгоритмов коллективного распознавания, обзор которых составляет основное содержание данной работы.

Данный обзор описывает текущее состояние исследований и сравнительный анализ известных авторам результатов в области объединения решений классификаторов. В разделах 2 рассматривается общая схема коллективного распознавания и ранние работы в этой области, которые к настоящему времени «почти забыты», но которые заложили основу современного понимания этой проблемы. Разделы 3, 4, 5, 6 и 7, которые формируют основное содержание данной работы — обзор разработанных к настоящему времени подходов, методов и алгоритмов объединения решений, В частности, описываются четыре группы подходов:

- 1) вероятностные подходы, основанные на классическом правиле Байеса и его различных упрощениях (раздел 4);
- 2) группа методов объединения решений, использующих кооперацию классификаторов на уровне мета-данных. В зарубежной литературе они называются методами, построенными на основе «многоярусного обобщения» (*“stacked generalization”*) (раздел 5);
- 3) группа методов, основанных на понятии областей компетентности классификаторов и использовании процедур, позволяющих оценивать компетентность классификаторов по отношению к каждому входу системы классификации(раздел 6);
- 4) методы объединения решений, основанные на применении нейросетевых технологий.

В разделе 7 анализируются нейросетевые методы объединения решений, которые базируются, в основном, на тех идеях, которые представлены методами, описанными в разделах 3, 4, 5 и 6. Раздел 8 посвящен краткому описанию структур объединения решений, которые могут быть использованы в системах коллективного распознавания и классификации. В разделе 9 обсуждается проблема разнообразия классификаторов и ее роль в повышении качества систем классификации, основанных на объединении и согласовании решений индиви-

дуальных классификаторов. Заключение резюмирует главные аспекты и основные направления исследований в рамках теории коллективного принятия решений множеством классификаторов.

В данном обзоре используются два термина, распознавание и классификация, которые имеют, вообще говоря, неодинаковый смысл. Однако те методы, которые описываются далее, в равной мере используются в задачах обоих типов, и при распознавании образов (например, изображений) и при решении задач классификации, когда решением является метка класса для некоторой сущности. Именно это является причиной равноправного использования терминов распознавание и классификация в рамках данного обзора.

2. Общая схема коллективного распознавания и ее простейшие примеры

Большинство известных к настоящему времени моделей объединения решений классификаторов можно охарактеризовать иерархической схемой, представленной на рис. 2. В ней на нижнем уровне рассматриваются источники данных об объекте классификации, на втором уровне помещается множество отдельных классификаторов, решения которых объединяются с помощью некоторого алгоритма; последний представляет третий уровень. По-видимому, впервые такая модель объединения решений была предложена в работе Фикса и Ходжеса еще в 1952 году [27]. В этой модели в качестве исходной информации рассматривалось некоторое множество интерпретированных примеров данных (каждое такое данное представлялось множеством значений признаков с указанием класса, к которому объект принадлежит). В пространстве примеров задавалась некоторая мера близости, например, евклидова. Для нового объекта, подлежащего классификации, отыскивалось априори заданное число соседей, ближайших по выбранной мере близости. Решающее правило сводилось к схеме голосования: новый объект относился к тому классу, к которому принадлежало наибольшее число ближайших соседей. В этом алгоритме легко узнать ныне широко используемый метод k ближайших соседей (на первоисточник давно уже никто не ссылается), а с другой стороны, широко известный и часто используемый метод голосования¹. Можно легко видеть, что метод k ближайших соседей является примером иерархической схемы классификации, представленной на рис. 2.

Другим примером такой же схемы является бинарная схема распознавания, предложенная Ф. Розенблатом [13], известная под названием «трехслойный перцептрон». В нем на первом уровне в качестве источников данных используется сенсорное поле² (S элементы в терминологии Ф. Розенблата) и каждый сенсор соединяется с некоторыми ассоциативными элементами (A -элементами), образующими второй уровень. Сенсоры могут выдавать только два сигнала, 0 или 1. Число ассоциативных элементов обычно меньше числа сенсоров. Связи сенсоров с ассоциативными элементами задаются матрицей связи A . Эта матрица может задаваться, например, случайным образом, хотя вариант ее задания существенно влияет на качество распознавания объектов,

¹В настоящее время автор этой работы забыт, а авторство как метода голосования, так и метода ближайших соседей приписывается авторам работ, опубликованных ближе к концу 1980-х годов.

²Ф. Розенблат пытался таким образом моделировать зрение живых организмов.

реализуемое перцептроном. Ассоциативные элементы A_j осуществляют простое суммирование входных сигналов. Каждый ассоциативный элемент имеет только один выход, который после некоторого преобразования подается на вход элемента R , который называется эффектором. Преобразование выходных

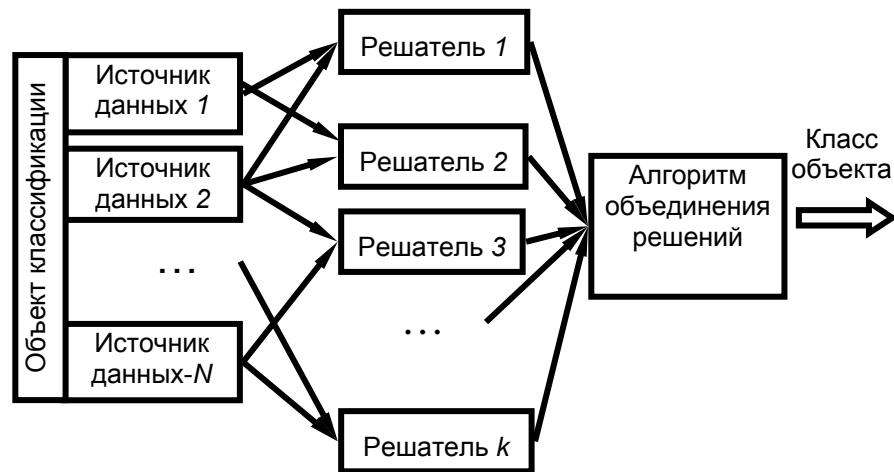


Рис. 2. Иерархическая схема классификации.

сигналов ассоциативных элементов выполняется дважды. Сначала ассоциативные элементы осуществляют суммирование входных сигналов и сравнивают его с некоторым порогом, и выходной сигнал ассоциативного элемента зависит от того, превышает ли полученная сумма заданный порог:

$$a_j = \begin{cases} 1, \text{ если } \sum_{i=1}^N b_{ij} s_i \geq 0; \\ 0, \text{ если } \sum_{i=1}^N b_{ij} s_i < 0, \end{cases} \quad (1)$$

где s_i — сигнал i -го сенсора, $b_{ij} = \pm 1$ — знак связи i -го сенсора с j -м ассоциативным элементом, $b_{ij} = 0$, если i -й сенсор не соединен с j -м ассоциативным элементом, N — число сенсоров, $j=1, \dots, m$, где m — число ассоциативных элементов.

Сигнал a_j поступает на вход элемента R (он образует следующий уровень классификации), где все входы от ассоциативных элементов суммируются с некоторыми весами:

$$r = \sum_{j=1}^m \beta_j a_j \quad (2)$$

Этот сигнал уже используется для принятия решения следующим образом:

$$\text{Объект} \in \begin{cases} \text{класс 1, если } r \geq 0; \\ \text{класс 2, если } r < 0. \end{cases} \quad (3)$$

Обучение перцептрона сводится к выбору весов β_j , $j=1, \dots, m$. Эта процедура хорошо известна и потому здесь ее описание опускается.

Легко видеть, что если выходы ассоциативных элементов (1) рассматривать как голосование за первый или второй класс, то принятие решений сводится к взвешенному голосованию.

Другой хорошо известной схемой иерархической классификации, предложенной на заре развития теории классификации и распознавания образов, является схема, предложенная М. М. Бонгардом [2]. В этой схеме также рассматривается бинарная схема распознавания. В ней роль индивидуальных решателей играют логические функции, принятые в форме конъюнкций, которые ставятся в соответствие решателям, представленным в иерархической схеме на втором уровне. При этом конъюнкции выбираются таким образом, что они принимают значение «истина» на объектах только одного класса. Качество каждой конъюнкции как признака класса определяется числом объектов «своего» класса обучающей выборки, на которых эта конъюнкция принимает значение «истина». Алгоритм объединения решений реализуется в форме дизъюнкции истинностных значений конъюнкций, поступающих на вход алгоритма объединения решений, и их отрицаний¹. Следует заметить, что более поздние алгоритмы такого класса, которые были разработаны в 1980–90-х годах (например, [22, 30, 51] и другие) по существу отличаются только более совершенными алгоритмами обучения и алгоритмами объединения решений. Частными случаями подхода М. М. Бонгарда являются, например, некоторые предложенные схемы аргументации.

Дальнейшее развитие иерархической схемы классификации в том простом варианте, как это описано в данном разделе, получили в 1980-х годах под названием *схем голосования* [15, 23]. Благодаря своей простоте и, в основном, удовлетворительной точности они активно используются и в настоящее время. Достаточно часто используется самый простой из них, известный под названием метода *простого голосования* (“*unweighted vote*”, или “*majority vote*”) [23, 26, 69]. Обычно в новейшей зарубежной литературе авторство приписывает работе [23], хотя правильно было бы сослаться ну уже ранее упоминавшуюся работу [27]. Получили распространение также различные схемы *взвешенного голосования* (“*weighted vote*”). Если результатом работы множества классификаторов являются условные вероятности классов для заданного вектора атрибутов, то простое суммирование вероятностей в пользу каждого из классов с последующей нормировкой результатов практически реализует такую схему голосования. Разработано также много других методов взвешенного голосования [15, 21, 31, 36, 56], и др.), которые по своим качествам в среднем мало отличаются, однако для конкретных приложений одни методы могут оказаться предпочтительнее других.

3. Парадигмы коллективного распознавания

Как уже отмечалось выше, проблема классификации уже к началу 1980-ых годов достигла своей зрелости. К тому времени было разработано большое разнообразие частных подходов, методов и алгоритмов классификации, которые уже упоминались выше: статистические, синтаксические с использованием теории автоматов, формальных языков и грамматик, структурные, фокусированные на спецификации отношений между элементами классифицируемых объектов и ряд других. К тому времени уже были исследованы и поняты достоинства, недостатки и потенциальные области применения различных методов и алгоритмов классификации.

¹ Этот алгоритм известен как алгоритм «Кора».

В то время преобладающим мнением было то, что основной задачей является *оптимизация классификаторов*, чьи решения предполагается объединять, и именно такой выбор должен обеспечить наиболее точную классификацию системы в целом. При такой парадигме задача объединения решений рассматривалась просто как следующий шаг в решении проблемы. В работе [33] такой подход называется «*моделью оптимизации решения*». К сожалению, такой подход не гарантировал, что система в целом будет способна достичь, по меньшей мере, точности наилучшего индивидуального классификатора, входящего в систему. Важнейший вопрос коллективного распознавания «*Возможно ли создание множества классификаторов для заданной проблемы таким образом, чтобы для каждого входного образа существовал, по крайней мере, один классификатор, формирующий правильное решение?*» при таком подходе вообще не рассматривался [33]. Однако позднее именно этот вопрос оказался основной движущей силой при исследованиях в области объединения решений. По существу ответ на поставленный вопрос формирует новую парадигму коллективного распознавания. Эта парадигма, которая ставит целью использовать новые свойства, которые могут появиться при взаимодействии классификаторов, в работе [33] названа “*coverage optimization methods*”, что может быть примерно переведено как «*методы оптимизации в целом*» (в отличие от оптимизации индивидуальных классификаторов). Впервые подход такого рода был предложен в уже ранее упомянутых работах Л. А. Растригина и Р. Х. Эренштейна [8, 9, 10, 11, 12]. Несколько позднее, к концу 1980-х годов и позже, было предложено несколько подходов, которые реализуют парадигму «*оптимизации в целом*». Например, такие хорошо известные методы как «*бутстрапинг*» (bootstrap aggregation), или *багинг* (bagging) [19], и *бустинг* (boosting) [28] были также ранними представителями подходов на основе парадигмы «*оптимизации в целом*». В первых двух методах обучающая выборка делится на несколько под-выборок фиксированного размера, и далее поочередно все они вместе, кроме одной, используются в процедуре обучения, а оставшаяся под-выборка используется для тестирования (эта процедура широко используется на практике и в зарубежной литературе называется “*cross validation*” — «*перекрестная оценка пригодности*»). В процедурах типа *бустинг* используется последовательное обучение на подмножествах обучающей выборки (эти подмножества могут формироваться различным образом, что и определяет разнообразие процедур типа *бустинга*), которые обязательно включают в себя объекты обучающей и тестовой выборок, которые ранее построенными решателями были классифицированы неправильно.

«Методы оптимизации в целом», в свою очередь, могут быть разделены на две группы: методы выбора компетентного классификатора и методы объединения решений классификаторов. Методы первой группы используют понятие компетентности классификатора. Впервые методы такого типа были предложены в работах [8, 9, 10, 11, 12]. Подходы данной группы различаются тем, как именно выбирается наиболее компетентный классификатор. Во второй группе, решения всех классификаторов рассматриваются как одинаково компетентные, и финальное решение определяется с помощью специальных правил, которые получаются также на основе обучающей процедуры. Эти правила, как и методы формирования правил могут быть различными, что приводит также к наличию разнообразия подходов во второй группе.

Рассмотрим некоторые из наиболее известных подходов к построения алгоритмов коллективного распознавания. уделяя основное внимание базовым идеям подходов и формальной основе алгоритмов.

4. Вероятностные подходы на основе правила Байеса и его упрощений

Необходимым предусловием использования вероятностных подходов для объединения решений состоит в том, что все классификаторы должны иметь однородные выходные данные, а именно, их выходы должны быть или бинарными (верное/неверное решение), или категориальными (метки классов), или векторами с мерами неопределенности одного типа («мягкие метки» классов).

При описании алгоритмов коллективного распознавания в данном разделе будут использоваться следующие обозначения:

- $(\omega_1, \omega_2, \dots, \omega_m)$ — метки классов решений;
- x_i — вектор измерений (он формирует пространств признаков i -го классификатора). Пространства признаков различных классификаторов могут быть различными, одинаковыми или пересекаться;
- $D_i(x_i)$ — решение i -го классификатора, D — имя класса;
- $p(x_i | \omega_k)$ — плотность распределения вероятности вектора измерений в классе ω_k ;
- $P(\omega_k)$ — априорная вероятность появления объектов класса ω_k ;
- R — общее число классификаторов, решения которых объединяются.

4.1. Правила объединения решений, получаемые упрощением правила Байеса

В данном и последующих подразделах данного раздела предполагается, что каждый классификатор может представлять выход в одной из следующих форм:

1. Верное/неверное решение:

$$D_i(x) \in \{0, 1\} \quad (4)$$

2. Метку класса:

$$D_i(x) \in \{\omega_1, \omega_2, \dots, \omega_m\} = \Omega \quad (5)$$

3. Вектор решений размерности m с мерой неопределенности:

$$\mu^i = \langle d_{i,1}(x_i), d_{i,2}(x_i), \dots, d_{i,m}(x_i) \rangle, \quad d_{i,j} \in [0, 1], \quad i \text{ — индекс классификатора} \quad (6)$$

Каждая компонента этого вектора выдает решение о принадлежности объекта соответствующему классу с некоторым значением уверенности $\mu_k^i = d_{i,k}(x_i)$, $k=1, 2, m$, которое иногда называется «мягкой меткой» (Это может быть апостериорная вероятность, нечеткость, возможность, мера Демпстера–Шафера, субъективная оценка уверенности и т.п.).

Оптимальное классификационное правило Байеса, получает на вход вектор измерений $x = \langle x_1, x_2, \dots, x_R \rangle$ и формирует решение следующим образом:

$$D \rightarrow \omega_j \text{ if } p(\omega_j | x_1, x_2, \dots, x_R) = \max_{k=1}^m p(\omega_k | x_1, x_2, \dots, x_R) \quad (7)$$

Здесь и далее выражение $D \rightarrow \omega_j$ означает, что результирующее решение D по множеству классификаторов принимается в пользу класса ω_j .

К сожалению, это правило сложно использовать на практике, т.к. необходимые распределения вероятностей обычно либо неизвестны, либо приводят к сложным вычислениям, либо распределение сложно оценивается. Однако, были предложены и обоснованы некоторые дополнительные предположения, которые существенно упрощают правило Байеса (7) (хотя при этом они перестают быть оптимальными). Рассмотрим их [38] и поясним, почему и как они могут быть использованы для объединения решений.

Первое упрощение исходит из предположения, которое известно как предположение об «условной независимости». Формально последнее представляется следующим образом:

$$p(x_1, x_2, \dots, x_R | \omega_k) = \prod_{i=1}^R p(x_i | \omega_k) \quad (8)$$

Если распределение вероятности, используемое в правиле Байеса (7), может быть представлена в такой форме, то тогда правило Байеса преобразуется к следующему виду [38]:

$$D \rightarrow \omega_j \text{ if } p^{(R-1)}(\omega_j) \prod_{i=1}^R p(\omega_j | x_i) = \max_k p^{(R-1)}(\omega_k) \prod_{i=1}^R p(\omega_k | x_i) \quad (9)$$

Это правило называется «*правилом умножения*» (*Production rule*). В соответствии с этим правилом, среди решений, представленных различными классификаторами, выбирается решение с максимальным значением вероятности. В соответствии с этим правилом, этот классификатор полагается наиболее компетентным для имеющегося входного вектора (другими словами, для данной точки в пространстве признаков).

Правило умножения часто используется на практике. Оправданием (но не более, чем оправданием) его использования в качестве метода объединения решений множества классификаторов являются следующие аргументы [38]:

- иногда на практике условная независимость на самом деле имеет место;
- в некоторых приложениях предположение об условной независимости может обеспечить более или менее адекватную аппроксимацию весьма сложной реальности;
- этот подход часто используется на практике и часто приводит к неплохим результатам;
- теоретический анализ показал, что ошибки, вызванные использованием данного правила, обычно компенсируются низкой чувствительностью решения к ошибке такой аппроксимации.

На практике многие из существующих стратегий объединения классификаторов основаны на этом правиле.

Другое упрощение исходит из предположения, что апостериорные вероятности, вычисленные соответствующими классификаторами, не слишком отли-

чаются от априорных [38]. Следует отметить, что данное предположение является, мягко говоря, «слишком сильным», и оно, как правило, не выполняется в большей части приложений. Правило, получаемое на основе такого предположения, называется *правилом суммы (Sum rule)*. Выглядит оно следующим образом [38]:

$$D \rightarrow \omega_j \quad \text{if} \quad (1-R)P(\omega_j) + \sum_{i=1}^R p(\omega_j | x_i) = \max_k [(1-R)P(\omega_k) + \sum_{i=1}^R p(\omega_k | x_i)] \quad (10)$$

Несмотря на то, что это правило основано практически на невыполнимом предположении, его использование мотивируется тем фактом, что оно обладает низкой чувствительностью к ошибке оценки распределения [38].

Другие правила, основанные на упрощении правила Байеса, также используются. Они базируются на следующих неравенствах [38] (эти неравенства всегда имеют место):

$$\prod_{i=1}^R p(\omega_k | x_i) \leq \min_{i=1}^R p(\omega_k | x_i) \leq (1/R) \sum_{i=1}^R p(\omega_k | x_i) \leq \max_{i=1}^R p(\omega_k | x_i) \quad (11)$$

Одним из таких правил является так называемое *правило максимума (Max rule)*, которое предполагает, что априорные вероятности объектов различных классов примерно равны:

$$D \rightarrow \omega_j \quad \text{if} \quad \max_{i=1}^R p(\omega_j | x_i) \leq \max_{k=1}^m \max_{i=1}^R p(\omega_k | x_i) \quad (12)$$

Еще одно правило, которое также предполагает равенство априорных вероятностей, называется *правилом минимума (Min rule)* [38]:

$$D \rightarrow \omega_j \quad \text{if} \quad \min_{i=1}^R p(\omega_j | x_i) \leq \max_{k=1}^m \min_{i=1}^R p(\omega_k | x_i) \quad (13)$$

Интересно отметить, что, если преобразовать апостериорные вероятности к бинарной форме в соответствии с уравнением

$$\Delta_{ki} = \begin{cases} 1, & \text{if } P(\omega_k | x_i) = \max_{j=1}^m P(\omega_j | x_i) \\ 0, & \text{в противном случае} \end{cases}, \quad (14)$$

то тогда *правило суммы* преобразуется в хорошо известную стратегию простого голосования (простого большинства), рассмотренное ранее в разделе 2:

$$D \rightarrow \omega_j \quad \text{if} \quad \sum_{i=1}^R \Delta_{ji} = \max_k \Delta_{ki} \quad (15)$$

4.2. Вероятностный подход: шаблоны решений

Рассматриваемый ниже подход предложен в работе [43]. Он основан на использовании R матриц (R равно общему числу классов), названных авторами «Шаблонами решений». Этот подход также базируется на вероятностном подходе к объединению решений классификаторов. Он использует сравнение выходов классификаторов с шаблонами эталонов, определенным для каждого класса. Данный подход не разделяет классификаторы по их компетентности, а использует все решения всех классификаторов для того чтобы вычислить итоговую «поддержку» в пользу каждого из классов.

Предположением является то, что каждый классификатор использует полный вектор признаков, а не его части, как это предполагалось в предыдущем подразделе.

Пусть $d^i = \langle d_{i,1}(x), d_{i,2}(x), \dots, d_{i,m}(x) \rangle$, $d_{i,j} \in [0,1]$, является выходом i -ого классификатора, где компоненты вектора d^i , $d_{i,j} \in [0,1]$, интерпретируются как значения меры уверенности (она может иметь смысл вероятности, меры нечеткости, возможности и т.п.) i -ого классификатора в том, что входной вектор x представляет объект класса j .

Выходы классификаторов организуются в соответствии со структурой в виде матрицы, называемой «профилем решения» (“decision profile”), $DP(x)$, в форме, представленной на рис. 3:

Предложено несколько вариантов рассматриваемого подхода, которые различаются тем, как используются профили решений. В частности, рассмотренные выше подходы, основанные на упрощении правила Байеса, легко интерпретируются в терминах профиля решений. Все правила, а именно, правило произведения, суммы, минимума, максимума, рассмотренные в подразделе 4.1, соответствуют некоторым операциям с колонками профиля решений с последующим выбором решения в пользу того или иного класса в соответствии с названными выше правилами. Например, правило суммы реализуется путем суммирования значений в колонках в соответствии с формулой (5) и выбора колонки, которой соответствует максимальное значение результата суммирования [43]. Заметим, что в этом и других правилах, рассматриваемых далее, индекс выбранной колонки указывает на метку класса, которой соответствует максимальное значение поддержки, выдаваемое классификаторами.

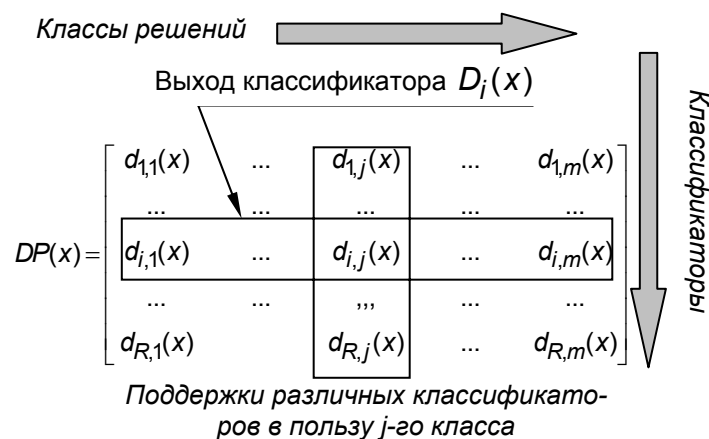


Рис. 3. Профиль решения

Второй вариант использования профиля решений базируется на интерпретации его компонент в качестве нового пространства признаков $D(x)$, которое строится как конкатенация строк профиля решений. Методы объединения этого типа названы авторами работы [43] «безразличными к классу» (“class indifferent”). Входом классификатора верхнего уровня $D(x)$ (рис. 3) является строка, полученная путем объединения строк профиля решений, а его выходом является метка класса.

Для слияния информации на верхнем уровне (мета-уровне) могут быть использованы многие из известных подходов. Однако, к сожалению, подход предложенный авторами работы [43], слишком прямолинеен, и он может не работать, если классификаторы первого уровня предоставляют бинарные решения из множества $\{0,1\}$, а объединение классификаторов основывается на использовании обычной линейной разделяющей функции. Причина в том, что такие подходы обычно базируются на использовании ковариационных матриц классов или на использовании ковариационной матрицы, построенной для всех классов. В таких случаях, если классификаторы нижнего уровня обладают высокой точностью, то названные ковариационные матрицы, как правило, будут сингулярными или плохо-обусловленными, что приведет к неустойчивости вычислений. Важно отметить, что рассматриваемый далее подход, называемый мета-классификацией, который основан на принципе «многоярусного обобщения» (“stacked generalization”) и использовании механизма классификации на основе правил, получаемых путем обучения, свободен от указанного недостатка.

Понимая и принимая этот недостаток, авторы статьи [43] предлагают подход, который является компромиссом между подходом, имеющим дело со столбцами профиля решений и использующим правила наподобие правил произведения, суммы и т.п. (“class-different” — в терминологии цитируемой здесь статьи), и подходом, предполагающим использование всех столбцов для слияние решений (“class-indifferent”). В этом подходе авторы используют понятие «шаблона решений». Не вдаваясь в детали этого подхода (он достаточно громоздок как в описании, так и в использовании), дадим лишь его краткую характеристику. В нем обучающая процедура имеет целью вычисление профиля решения для каждого класса объектов по отдельности путем усреднения его по множеству обучающих данных, относящихся к каждому классу. Процедура классификации базируется на поиске минимуму евклидовой меры профиля решений входного описания на множестве эталонных профилей классов. Возможно также использование других, кроме евклидовой, метрик близости. Детали алгоритмической реализации данного подхода могут быть найдены в работе [43].

5. Объединение решений: «многоярусное обобщение» и мета классификация

Идея метода «многоярусного обобщения» (stacked generalization) достаточно проста, однако она оказалась весьма продуктивной для задач объединения решений классификаторов. Исторически, она идет от работы [70]. Идея использования этого подхода задачах объединения решений впервые встречается в работе [49]. Независимо от работы [70] аналогичная идея была предложена в работе [61] под названием «*мета-классификация*». Этот вариант многоярусного обобщения был реализован применительно к задаче обнаружения вторжений в компьютерную сеть [57, 61]. Рассмотрим этот метод, поскольку в нем идея многоярусного обобщения была алгоритмизирована в той форме, в какой она сейчас наиболее часто используется.

Пусть имеется обучающая выборка L , и выбрано множество алгоритмов классификации A_1, \dots, A_K . Эти алгоритмы называются базовыми классификаторами. Для объединения решений используется специальный алгоритм, который

называется мета-классификатором. Входными данными мета-классификатора являются решения базовых классификаторов, т.е. его входом является множество меток классов, к которым базовые классификаторы отнесли описание входного объекта (Это может быть и множество «мягких» меток). Множество меток на входе мета-классификатора интерпретируется как множество признаков нового признакового пространства. Соответственно, обучающие и тестирующие данные для мета-классификатора (их принято называть мета-данными обучения) формируются базовыми классификаторами на основании тех данных, которыми располагают базовые классификаторы. Таким образом, мета-данные состоят из множества кортежей («мягких») меток классов, полученных в качестве решений алгоритмами A_1, \dots, A_K при их тестировании на множестве входных описаний экземпляров объектов. Каждому такому кортежу решений ставится в соответствие метка (кортеж «мягких» меток) класса, к которому этот объект относится на самом деле. Таким образом, обучающие данные мета-классификатора представляются векторами вида $(x_{1j}, \dots, x_{Rj}, y_j)$, где j — индекс экземпляра объекта из тестовой выборки, $1, \dots, R$ — индексы базовых классификаторов, x_{ki} — метки классов (или строки мягких меток), полученных базовыми классификаторами A_k для тестируемого примера с номером i , y_j — истинное значение метки тестируемого примера. Обобщенная схема объединения решений, основанная на мета-классификации, представлена на рис. 4. Метод мета-классификации был неоднократно использован авторами данной статьи в различных приложениях, и везде он приводил к хорошим результатам. В частности, он использовался в таких приложениях, как обнаружение вторжений в компьютерные сети (в различных постановках этой задачи), распознавание наземных объектов по данным инфракрасных изображений, полученных с борта летательных аппаратов и в ряде других.

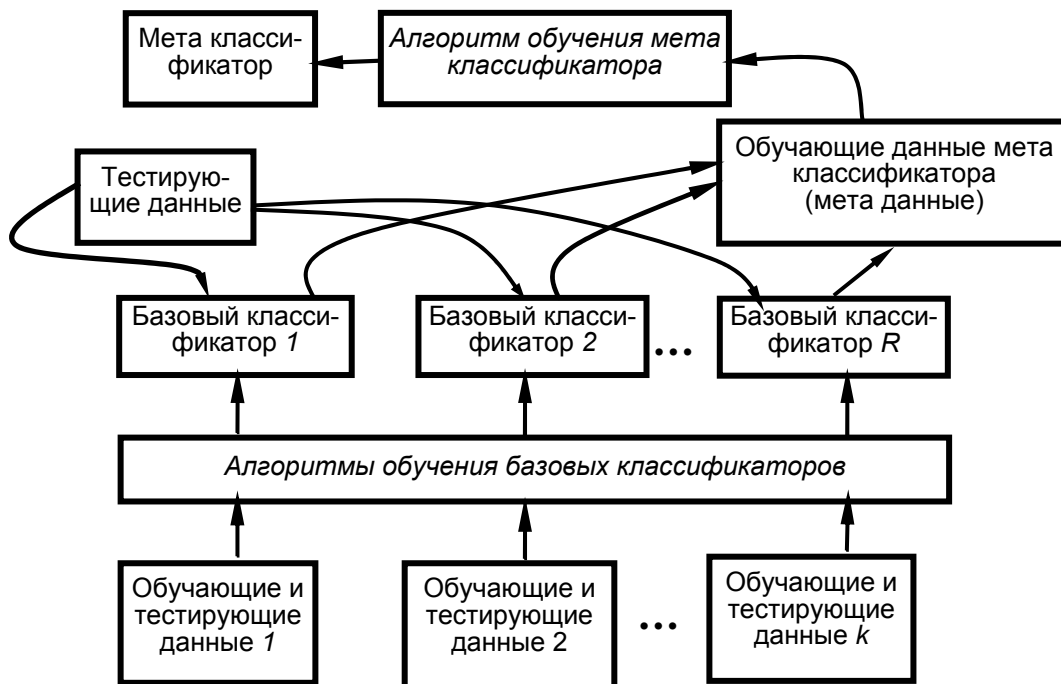


Рис. 4. Схема генерации мета-данных для обучения и схема обучения мета-классификатора.

Рассмотрим другие варианты объединения решений, которые используют идеи метода многоярусного обобщения.

В работе [63] рассматривается вариант алгоритма многоярусного обобщения для случая, когда решение каждого из базовых классификаторов представляется в виде вектора вероятностей принадлежности объекта обучающей выборки к одному из классов (т.е. когда используется один из вариантов «мягких» меток, упомянутых в предыдущем разделе). В этом случае мета-данные представляются в числовой шкале, и для мета-обучения авторы работы [19] предлагают использовать регрессионную модель (stacked regression model). Авторы экспериментально обнаружили (на основании тестирования этого варианта многоярусного обобщения на десяти различных наборах данных из UCI репозитория [50], что такой подход приводит к уменьшению ошибки классификации по сравнению со случаем, когда в качестве мета-данных используются метки классов бинарного или категориального типов. Однако следует отметить, что авторы сравнивают точность предложенного ими метода с методом простого голосования, который среди известных методов объединения решений не является наилучшим, а не с методом мета-классификации. Возможно, что другой метод, в частности, метод мета-классификации, мог бы дать лучшие результаты.

Как было отмечено ранее, одной из ранних работ, использующих идеи многоярусного обобщения для объединения решений, является работа [49]. Многие авторы, работающие над проблемой объединения решений [17, 29 49], и т.д.), придают особое значение целесообразности выбора на базовом уровне таких классификаторов, которые обеспечивают некоррелированность ошибок классификации (иначе говоря, отмечают важность такого выбора базовых классификаторов, которые «ошибаются» на различных входных данных или в разных подобластях пространства признаков). Заметим, что целесообразность такого требования ранее демонстрировалась на рис. 1. Например, в работе [49] исходное множество классификаторов строится на базе *различных* хорошо известных алгоритмов, а именно CN2 [22], C4.5 [58], OC1 [52], PEBLS [24], алгоритма k ближайших соседей, наивного байесовского метода и алгоритма обратного распространения ошибки [59]. Исследования показали, что в случае использования столь разнообразных классификаторов¹ достигается лучшее качество классификации, чем в том случае, когда для различных классификаторов используется один и тот же алгоритм, обученный на разных подмножествах обучающих данных. Заметим, что такие широко известные методы как багинг, бустинг и ряд других, в принципе, используют классификаторы, построенные именно таким образом, т.е. на базе одного и того же алгоритма, которые различаются выборками, на которых они обучены и тестированы, что, как следует из ранее сказанного, не является наилучшим вариантом выбора базовых классификаторов.

В работе [29] рассматривается другая версия алгоритма многоярусного обобщения. Он назван методом каскадного обобщения. Предложенный в данной работе подход содержит ряд оригинальных идей, в частности, он расширяет некоторые из хорошо известных методов, основанных на многоярусном обобщении. Идея каскадного обобщения заключается в последовательном использовании нескольких различных классификационных алгоритмов. Также как и в методе, описанном в работе [63], выходы базовых классификаторов представляются в виде кортежей вероятностей, («мягких» меток), каждая компонен-

¹Проблема разнообразия классификаторов и меры оценки разнообразия рассматриваются далее детально в разделе 9.

та которого отвечает мере уверенности классификатора в том, что классифицируемый объект относится к соответствующему классу. Далее, в методе каскадного обобщения, создаются два множества мета-данных, обозначенных L и T , которые соответствуют двум группам атрибутов мета-данных, генерируемых на основе выборок обучающих и тестовых данных, которые используются базовыми классификаторами. Однако, в отличие от метода, предложенного в работе [63], в множествах L и T используются не только кортежи вероятностных меток, полученные классификаторами по отношению к каждому из классов, но также и все множество атрибутов базового уровня. Заметим, что в базовой версии метода каскадного обобщения используется единственный базовый классификатор, однако общая модель, предложенная в работе [29], позволяет использовать несколько таких классификаторов. Построенные множества L и T рассматриваются как обучающее и тестовые данные для мета-классификаторов уровня 1. Для этих мета-классификаторов решается задача обучения и тестирования аналогично тому, как это описано для мета-классификаторов первого уровня. Одним из основных требований к алгоритмам мета-классификации в данной работе является то, что они должны быть иными чем те, которые используются в качестве алгоритмов базовых классификаторов. В общем случае схема каскадного обобщения может включать в себя несколько уровней с использованием на каждом из них процесса обучения и тестирования.

В представленных выше алгоритмах мета-классификации алгоритм мета-уровня, выполняющий объединение решений, может интерпретироваться как арбитр, который анализирует решения базовых классификаторов и принимает окончательное решение. Он может согласиться с одним из уже предложенных решений, а может принять иное решение, отличное от всех решений, предложенных базовыми классификаторами. Заметим, что возможность такого выбора является существенным отличием методов объединения решений, основанных на многоярусном обобщении, от других подходов, например, вероятностных методов голосования, в которых финальное решение всегда выбирается из множества решений, предложенных базовыми классификаторами. Простейшим примером алгоритма, в котором явно используется понятие «арбитража», является алгоритм, описанный в работе [54]. В нем рассматривается задача объединения решений двух классификаторов, что выполняется третьим классификатором, который явно называется арбитром и вовлекается в процесс принятия решений лишь тогда, когда два классификатора нижнего уровня расходятся во мнениях. Такой же термин иногда используется в работах по мета-классификации [24, 57].

Вообще говоря, группа методов объединения решений классификаторов, основанных на многоярусном обобщении, весьма популярна среди исследователей, и все больше работ появляется в этой области. Практика показывает (в том числе, обширная практика авторов данного обзора) что методы многоярусного обобщения достаточно хорошо работают на практике, в частности, в варианте мета-классификации. Отрицательной стороной методов этой группы является необходимость их переобучения в том случае, когда на базовом уровне в систему добавляется новый классификатор. В противоположность ему, группа методов, которая рассматриваемые в последующем разделе, свободна от этого, иногда весьма существенного недостатка.

6. Выбор классификатора: подходы основанные на оценке компетентности классификаторов

Методы данной группы используют понятие *компетентности классификаторов*. Идея таких методов базируется на том, что каждый базовый классификатор может работать хорошо в некоторой области пространства признаков (эта область интерпретируется как область компетентности классификатора), превосходя в этой области остальные классификаторы по точности и достоверности решений. Область компетентности каждого базового классификатора должна каким-либо образом оцениваться. Соответствующая программа в работе [55] называется «рефери» или «арбитром». Соответственно, задача классификации в рассматриваемой группе методов решается таким образом, что каждый алгоритм используется только в области своей компетентности, т.е. там, где он дает наилучшие результаты по сравнению с другими классификаторами. При этом в каждой области принимается во внимание решение одного и только одного классификатора. Однако, очевидно, что дополнительно к множеству базовых классификаторов в рассматриваемом случае необходимо иметь какой-то метод, алгоритм, который для любого входа определяет, какой из классификаторов наиболее компетентен. Поскольку в рассматриваемом подходе основная задача — это выбор решения одного из нескольких классификаторов, наиболее компетентного, поэтому такая парадигма коллективного распознавания называется «выбор решения».

В западной литературе считается, что этот метод впервые был предложен в работах [62] и [49]. Однако, еще за более чем два десятилетия до этого, подход к объединению решений на основе оценки компетентности классификаторов был предложен в работах Л. А. Растригина и Р. Х. Эренштейна [8, 9, 10, 11, 12]. Причем важно отметить, что предложенный ими подход к коллективному распознаванию был назван тем же именем, а именно «подходом на основе компетентности классификаторов». Однако далее обзор методов такого типа дается на основе работ современных зарубежных авторов, в которых используется более современная терминология, а оценка различных вариантов алгоритмов базируется на более обширном компьютерном эксперименте и разнообразных обучающих и тестовых данных.

Ряд авторов, предлагающих свои версии алгоритма объединения решений на основе оценок компетентности классификаторов, ссылаются на результаты работы [17]. Эта работа представляет интерес с различных точек зрения. Основной ее результат — это модель оценивания ошибки конкретного классификатора, а также модель для оценки различия между классификаторами. Как указывают авторы цитируемой работы, первая модель может использоваться для таких целей как (1) определение областей в пространстве признаков, где классификатору нельзя доверять, (2) определение «трудных» областей данных для конкретного приложения, (3) отыскание областей, в которых тот или иной классификатор нужно улучшать или «латать», (4) определение влияния изменений параметров классификатора на качество его работы.

Очевидно, что для реализации любого метода объединения решений, основанного на оценках компетентности отдельных классификаторов по отношению ко входному описанию объекта распознавания или классификации, необходимо решить следующие ключевые вопросы: (1) Что понимать под компетентностью алгоритма (автор работы [62] называет это свойство «точностью предсказания в данной области» — *“predictive accuracy in this region”*), и (2) Как

вычислять этот показатель компетентности для конкретной области и конкретного алгоритма? В работе [62] этот вопрос решается для частных случаев алгоритмов классификации, а именно, для таких алгоритмов как наивный байесовский метод и одной из вариаций метода ближайшего соседа, который в зарубежной литературе известен под названием *Instance-based learning*.

Достаточно детальное описание и математически корректная алгоритмизация метода объединения решений на основе оценок компетентности базовых классификаторов дано в работе [55]¹. Описанный здесь подход предполагает, что вместе с каждым классификатором используется специальный алгоритм, называемый «рефери», который предназначен для оценки компетентности классификатора. Под компетентностью классификатора в данной области пространства представления объектов классификации авторы понимают его точность, т.е. вероятность правильной классификации объектов, чье описание принадлежит этой области. Так как авторы исследуют данный подход экспериментально применительно к алгоритму C4.5 (алгоритм построения деревьев решений, предложенный в работе [58]), то проблема определения компетентности в этом случае решается тривиально: в процессе поиска по дереву решений классифицируемый объект в каждом из классификаторов попадает в некоторый лист дерева, а для каждого листа дерева решений каждого классификатора значение вероятности правильной классификации известно по результатам тестирования. Поэтому поиск сводится просто к сравнению значений вероятностей правильной классификации, обеспечиваемой каждым из классификаторов для поиска наибольшего из них.

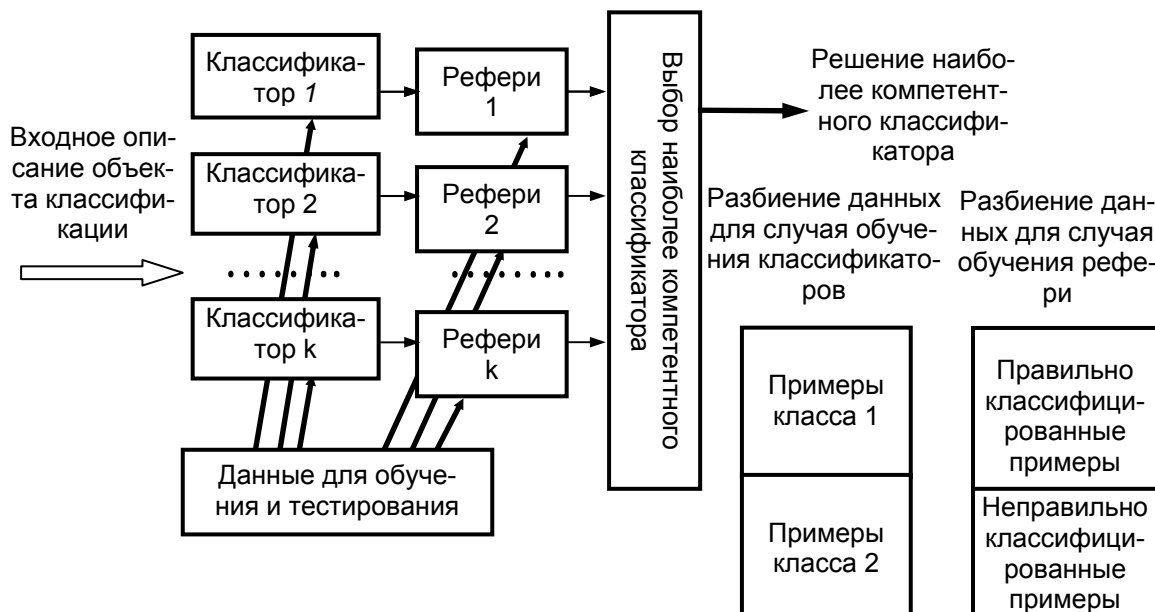


Рис. 5. Схема объединения решений на основе выбора наиболее компетентного классификатора и схема разбиения данных для обучения классификаторов и для обучения рефери.

Однако в реферируемой работе [55] предложена и общая схема обучения коллективному распознаванию на основе оценки компетентности. Эта схема состоит из двух шагов (рис. 5). На первом шаге производится обучение и тести-

¹По рассматриваемой группе подходов к задаче объединения решений данная работа является наиболее полной, а главное — конструктивной.

рование каждого конкретного базового классификатора. Этот шаг не отличается от обычных схем обучения. Любые методы, адекватные постановке задаче и типам данных, формирующих пространство признаков, могут быть здесь использованы. На следующем шаге, после тестирования каждого классификатора, обучающая выборка, которая использовалась на этапе тестирования для некоторого классификатора, разделяется на два подмножества, L^+ и L^- . При этом в первое подмножество включаются те экземпляры исходной тестовой выборки, которые при тестировании были классифицированы правильно. Во второе подмножество включаются остальные экземпляры тестовой выборки, т.е. те, которые были классифицированы ошибочно. Рассматривая эти множества данных в качестве областей компетентности и некомпетентности классификатора соответственно, их можно использовать в качестве обучающих данных для обучения алгоритма «рефери». Очевидно, что это стандартная задача обучения, которую можно решать различными методами. Нестандартным элементом здесь является только необходимость вычисления вероятности правильной классификации.

При классификации новых данных задача рефери состоит в том, чтобы для каждого входного примера определить, принадлежит ли он области компетентности алгоритма или нет, и если принадлежит, то какова вероятность правильной классификации этого примера. После этого рефери поручает выбранному (наиболее компетентному) классификатору самостоятельно решать задачу классификации.

Большинство известных методов классификации позволяют определить вероятность правильной классификации отдельного примера, хотя, возможно, в ряде случаев для этого потребуются некоторые доработки алгоритма. Как уже отмечалось, для метода C4.5 эта проблема решается тривиально [55]. То же самое можно сказать и о «наивном байесовском классификаторе» [62].

Общее свойство методов комбинирования решений на основе оценок компетентности алгоритмов состоит в том, что чем детальнее описаны области компетентности каждого из классификаторов, тем большей точностью обладает алгоритм в целом¹. В различных методах необходимая степень детальности описания областей компетентности алгоритмов достигается различным образом. В самом простом случае эти характеристики можно получить по результатам тестирования. Хотя такие характеристики компетентности алгоритма могут оказаться достаточно грубыми, но, тем не менее, их использование ведет к повышению точности алгоритма с выбором наиболее компетентного классификатора по сравнению с точностью, даваемой каждым из них в отдельности [55].

В несколько более поздней работе [60] рассматривается еще одна постановка рассматриваемой задачи, которая во многом сходна с постановкой задачи в работе [55], однако имеет и свои отличия. В частности, в ней задача метаобучения рассматривается в более общей форме. В этой постановке для каждого классификатора базового уровня после его обучения использованная обучающая выборка разбивается на две подобласти (как и в работе [55], которые рассматриваются как обучающие данные для алгоритма «рефери»). Отличительная особенность данного алгоритма состоит в том, что если несколько классификаторов базового уровня оцениваются для некоторого входа как одинаково компетентные, но при этом дают различные решения, то далее исполь-

¹ В пределе, когда достигается достаточно детальное дробление пространства значений атрибутов, метод вырождается в известную схему “*case-based reasoning*”.

зуется либо схема голосования, либо выбор решения делается на основе подсчета меры уверенности, которая вычисляется упрощенным образом по выборке данных на основе частотных оценок вероятностей классов. Эмпирические оценки точности такого подхода показали, как утверждают авторы, что на некоторых данных этот метод превосходит другие методы объединения решений.

Другой, весьма интересный, вариант алгоритма данного класса предложен в работах [64] и [65]. В этих работах рассматривается задача объединения решений множества классификаторов базового уровня, обученных на одном и том же множестве данных, но построенных с помощью различных алгоритмов обучения. Полагается, что решение каждого такого классификатора, формируемое в процессе тестирования, включает в себя имя класса с некоторой мерой уверенности, например, с вероятностью принадлежности классифицируемого объекта указанному классу. Эти решения, построенные на тестовых данных и записанные в виде множества строк, каждая из которых содержит указанные пары для всех классификаторов базового уровня, образуют мета-данные. Каждой строке мета-данных, как обычно, приписывается имя истинного класса (правильная интерпретация). Мета-данные, построенные таким образом, используются для обучения мета-дерева решений. Мета-дерево решений имеет «почти» обычную структуру и строится обычным образом, например, с использованием алгоритма C4.5. Основное отличие мета-дерева от обычного дерева решений состоит в том, что в мета-дереве промежуточным узлам, как обычно, приписываются тестовые атрибуты из числа атрибутов мета-данных, однако в отличие от обычного дерева решений, *каждому листу мета-дерева приписывается имя алгоритма* базового уровня, который классифицирует тестируемый пример с наибольшей мерой уверенности. Напомним, что в «обычном» дереве решений листу приписывается имя класса принадлежности тестового примера. Последнее и является тем новшеством, которое придает данному алгоритму новые, весьма интересные, свойства. Иначе говоря, мета-дерево задает имя предпочтительного классификатора для подмножества данных, отвечающих листу мета-дерева. Очевидно, что для построения мета-дерева решений можно использовать не только алгоритм C4.5, но и любой другой метод построения дерева решений.

Результаты сравнения различных вариантов построения мета-дерева решений, а также результаты сравнения его свойств с другими методами объединения решений, которые получены на основании обширного эксперимента, тщательно проведенного авторами, могут быть найдены в работе [65]:

В целом, группа алгоритмов комбинирования решений, использующих рефери для определения наиболее компетентного классификатора в зависимости от входных данных, представляется весьма перспективной. Достоинство подходов этой группы, кроме повышения точности классификации, также и в том, что если в уже работающую схему ввести новый классификатор, то его введение не потребует переучивания ранее использовавшихся классификаторов. В этом случае путем обучения новому классификатору должна назначаться область компетентности, в которой он ответственен за принятие решений. Это заключение справедливо как для алгоритмов, описанных в работе [55], так и для алгоритмов, использующих мета-дерево решений [64, 65]. К сожалению, во всех работах оставлен в стороне вопрос о том, как контролировать покрытие всего пространства атрибутов областями компетентности. Косвенным образом влияние этого недостатка снижается требованием некоррелированности ошибок используемых классификаторов.

7. Нейросетевые методы в задачах объединения решений

В работах, посвященных использованию нейронных сетей для объединения решений классификаторов, преобладают два направления, суть которых представлена в работе [16]. В этой работе автор рассматривает проблему объединения решений в системах, состоящих из нейронных сетей, объединенных в определенную структуру. Все методы объединения решений делятся на методы, которые используют объединение в ансамблях сетей (ensembles of neural networks) и те, которые используют нейронные сети, построенные из модулей. По существу, это те же два направления, которые были ранее рассмотрены в данном обзоре, а именно, подход на основе многоярусного обобщения и подход на основе оценки компетентности классификаторов.

Первый из них, метод многоярусного обобщения (“stacked generalization”), рассматривает использование нейронной сети для объединения решений базовых классификаторов (также [34, 46]). Этот вариант представлен на рис. 6.

Выходом каждого базового классификатора в общем случае является вектор решений (вектор, содержащий в качестве значений «мягкие метки»), значения элементов которого принадлежат некоторому числовому интервалу $[a, b]$. Эти значения подаются на вход нейронной сети (она должна быть обучена объединению решений классификаторов базового уровня), выходом которой является решение в пользу того или иного класса. Выходом сети также может являться вектор, размерность которого равна количеству классов распознаваемых объектов, который на каждой позиции имеет значение некоторой меры доверия в пользу того или иного класса. В этом случае в качестве решения может быть выбран класс с максимальным значением такой меры.

Настройка весов нейронной сети происходит в процессе обучения, например, с помощью алгоритма обратного распространения ошибки [14]. На начальной фазе обучения, чтобы избежать попадания в локальный минимум, можно использовать один из алгоритмов глобальной оптимизации, например, генетический алгоритм [14].

Система объединения решений функционирует следующим образом:

- 1) выбирается и обучается множество базовых классификаторов;
- 2) подготавливаются мета-данные для обучения нейронной сети (раздел 4 данного обзора). Для этого базовые классификаторы тестируются с использованием интерпретированной выборки данных и формируется для каждого тестового примера вектор решений базовых классификаторов, к которому добавляется компонента, в которую вносится имя истинного класса принадлежности тестируемого примера;
- 3) выборка мета-данных используется для обучения нейронной сети, выполняющей объединение решений.

В работе [46] для объединения решений предлагается использовать двухслойную нейронную сеть, которую автор называет динамической нейросетью для объединения решений (Dynamic Combination Neural Network). В предложенной структуре сети в первом слое все весовые коэффициенты равняются единице и не модифицируются в процессе обучения, тем самым реализуют объединение векторов (результатов каждого классификатора) в один составной вектор. Далее результат работы сети (при условии что в качестве активационной функции нейронов используется сигмоидная смещенная функция) вычисляется в соответствии со стандартным алгоритмом:

$$D_j = \frac{1}{1 + \exp(-(\sum_{i=1}^M \sum_{k=1}^N w_{ijk} s_{ik} + \theta_j))}, \quad (16)$$

где D_j — j -ый выход сети, M — количество классификаторов, N — размерность выходного вектора каждого классификатора, θ_j — смещение j -го нейрона выходного слоя, s_{ik} — поддержка i -го классификатора в пользу k -го класса, w_{ijk} — весовой коэффициент сети от j -го входного нейрона i -го классификатора к k -му выходному нейрону.

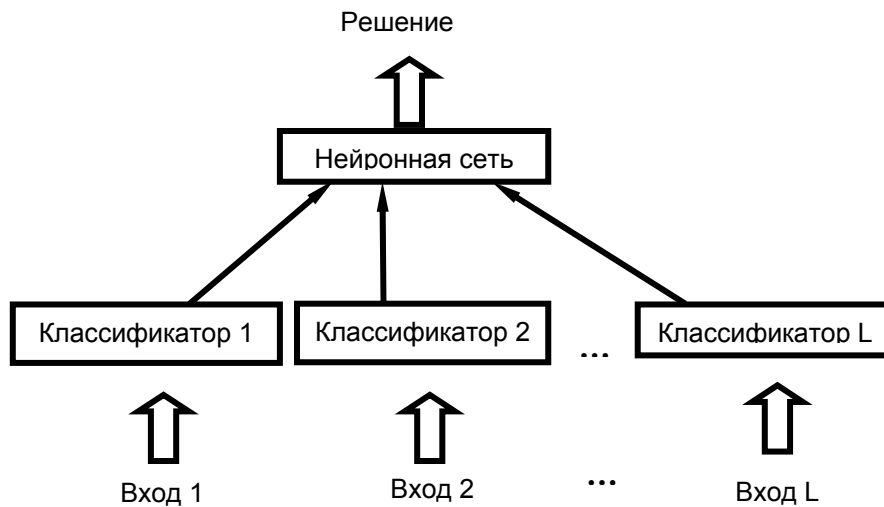


Рис. 6. Схема объединения решений множества классификаторов используя нейронную сеть.

Как видно использование такой структуры сети состоит в том, что каждый весовой коэффициент сети играет понятную роль в принятии решения, а именно вес w_{ijk} определяет поддержку j -му классу когда i -ый классификатор определил поддержку s_{ik} k -му классу. В общем же случае можно выбирать структуру сети на основании каких либо имеющихся условий, ограничений или рекомендаций.

Для модульных нейронных сетей в работе [16] (также [46, 47, 48] предлагается использовать так называемую «шлюзовую сеть» (“gating network”), нейронную сеть для оценки компетентности классификаторов для конкретного входного вектора данных предъявляемых классификаторам. По существу этот вариант рассматривает нейросетевую парадигму для объединения решений на основе оценок компетентности классификаторов (раздел 5 данного обзора). Соответствующая теория здесь называется “mixture of experts” — «смесь экспертов».

В статье [47] авторы предлагают адаптивный подход к объединению решений, в котором используется фактически нейросетевой аналог подхода, предложенного в работе [55], каждому классификатору ставится в соответствие программа «рефери», которая предсказывает степень его компетентности по отношению к конкретному входу, подаваемому на вход множества классифика-

торов базового уровня. Рис. 7 представляет структуру объединения решений в таком случае.

В зависимости от входного вектора X решения различных классификаторов могут быть выбраны и использованы для принятия объединенного решения. Количество входов предсказывающей сети равно размерности входного вектора пространства признаков. Количество выходов сети равняется количеству классификаторов, т.е. L . Предсказывающая нейросеть обучается предсказывать меру компетентности каждого классификатора при предъявлении ей

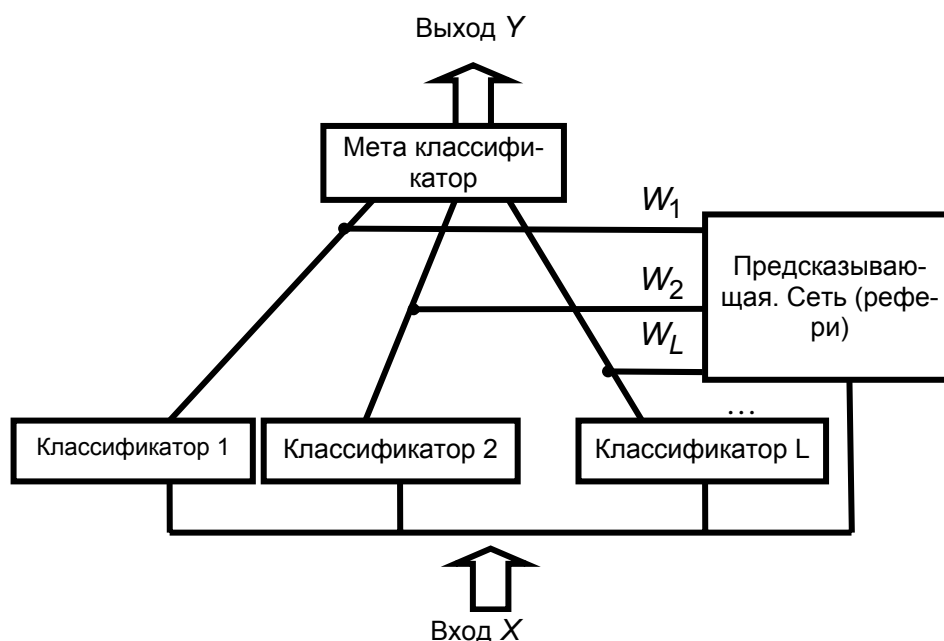


Рис.7. Схема принятия решений основанная на выборе классификаторов предсказывающей сетью.

конкретного входного вектора, т.е. оценку того факта, что классификатор выдает правильное решение. Обычно степень компетентности оценивается числом из интервала $[0,1]$.

Процедура построения такой системы объединения решений выглядит аналогично тому, как это было описано в разделе 5:

- 1) разделить имеющееся множество интерпретированных данных на обучающее и тестирующее подмножества;
- 2) выбрать и обучить все классификаторы (их общее число обозначено символом L);
- 3) сгенерировать мета-данные для обучения предсказывающей нейросети, т.е. протестировать все классификаторы базового уровня на тестирующем подмножестве данных. При этом для каждого вектора x_i из тестирующего множества сформировать вектор $t_i = [t_{i1}, \dots, t_{iL}]^T$, элементы которого равны единице, если соответствующий классификатор выдал правильное решение, и нуль — в противном случае, т.е. $t_{ij}=1$ если входной вектор x_i был правильно классифицирован j -м классификатором;
- 4) на основании данных полученных на предыдущем шаге обучить предсказывающую сеть;

- 5) определить пороговое значение β для включения соответствующих классификаторов в общее решение, т.е. решение j -го классификатор учитывается если $p_j > \beta$, где p_j — значение j -го выхода предсказывающей сети.

Процесс классификации происходит следующим образом:

- 1) подать входной вектор X на вход предсказывающей сети и определить множество тех классификаторов, для которых $p_j > \beta$;
- 2) подать входной вектор X на входы классификаторов из множества, полученного на шаге 1;
- 3) объединить решения классификаторов в соответствии с выбранным методом объединения решение множества классификаторов.

Существует множество работ, в которых предлагаются аналогичные подходы, «различающиеся», в основном, тем, как называется предсказывающая сеть. Например, в работе [46] авторы называют такую сеть *сетью динамического выбора* (Dynamic Selection Network), которая вычисляет коэффициенты, используемые для масштабирования выходов (результатов работы) классификаторов, тем самым увеличивая или уменьшая вклад каждого классификатора в общее решение. Этот подход аналогичен рассмотренному выше за исключением того, что здесь не используется отсечение некомпетентных классификаторов. Вместо этого учитывается вклад, который вносит каждый классификатор, хотя в данном случае нейронная сеть также вычисляет значения, которые могут быть интерпретированы как оценки компетентности соответствующих классификаторов. Масштабирование производится путем простого перемножения результатов работы классификатора (вектор значений в общем случае) на соответствующий коэффициент, выданный сетью для данного классификатора:

$$g_i(x) CL_i(x) = CL_i(x) \cdot g_i(x), \quad (17)$$

где x — входной вектор, $CL_i(x)$ — выход i -ого классификатора, $g_i(x)$ — i -ый выход сети.

В работе [48] рассматривается аналогичный подход. В отличие от предыдущей работы, автор вводит понятие «шлюз» (gating module), который определяет степень участия каждого классификатора в принятии общего решения, т.е. фактически шлюз представляет собой реализацию функций $g_i(x)$ (формула 16). Значениями этих функций могут быть как бинарные значения $\{0,1\}$ (отсечение некомпетентных классификаторов с точки зрения шлюза), так и вещественные числа, например из диапазона $[0,1]$, представляющие собой подкрепления в пользу соответствующих классификаторов. Несмотря на то, что автор явно не указывает на реализацию этого модуля в виде нейронной сети, он отмечает, что часто именно так обычно и происходит реализация алгоритма. Далее взвешенные результаты работы классификаторов или результаты усеченного множества классификаторов могут быть объединены с помощью любой техники объединения решений (например, с помощью одной из техник, рассмотренной в данном обзоре).

В некоторых других работах, например [35, 53, 67] данный подход рассматривается в контексте объединения мнений экспертов (“mixture of experts”).

Например, в работе [53] предлагается архитектура системы объединения решений, которая состоит из нескольких экспертов (нейросетей) и шлюзовой

сети (gating network). В отличие от рассмотренных ранее подходов, здесь выходы сети нормализуются следующим образом:

$$g_i = \frac{e^{g_i}}{\sum_{j=1}^L e^{g_j}}, \quad (18)$$

где g_i — выходы сети, g_i — нормализованные значения выходов. Использование данной функции приводит к тому, что сумма всех нормализованных выходов сети равняется единице, реализуя слабое соревнование¹ между экспертами. На верхнем уровне, как в ранее рассмотренных случаях, может быть использован любой алгоритм объединения решений.

Отличительной особенностью работ [67] и [35] является то, что авторы предлагают использовать несколько шлюзовых сетей — иерархическое объединение экспертов (Hierarchical Mixtures of Experts, HME). В основе подхода лежит принцип «разделяй и властвуй», т.е. принцип разделения исходной задачи, решение которой трудно получить, на множество подзадач, решения которых может быть получено уже достаточно просто.

8. Варианты структур взаимодействия классификаторов при коллективном принятии решений

В общем случае структура взаимодействия классификаторов, принимающих участие в коллективном принятии решений может быть представлена в виде, изображенном на рис. 5. Обычно в этой структуре присутствует уровень индивидуального принятия решений классификаторами базового уровня, которые могут использовать как множество источников данных, так и только один источник. Существенно, что классификаторы базового уровня являются различными, причем это различие может быть различного свойства. Структура, представленная на рис. 8, демонстрирует, в чем именно классификаторы могут различаться. Как видно из рисунка, это различие может проявляться в таких вариантах:

- признаковые пространства, используемые классификаторами, различны;
- классификаторы используют различные источники данных для принятия решений; эти источники могут фиксировать как одни и те же признаки, так и различные. Это может обуславливать различие в механизмах обучения и в алгоритмах принятия решений;
- классификаторы, использующие одни и те же признаковые пространства, могут использовать различные выборки данных для обучения и/или тестирования;
- используя одни и те же данные, классификаторы могут использовать различные алгоритмы принятия решений, а это приведет и к различию в механизмах обучения.

Здесь следует заметить, что различие классификаторов, обусловленное, например, перечисленными причинами, не означает их разнообразия, хотя эти понятия, различие и разнообразие, являются взаимосвязанными. Формально

¹ В оригинале “Soft competition”, название обусловлено тем, что функция (17) в работе [54] называется soft max function.

проблема разнообразия классификаторов понимается иначе, о чем пойдет речь в следующем разделе.

Рассмотрим другие отличительные черты задач коллективного распознавания, которые следует принимать во внимание при конструировании структуры

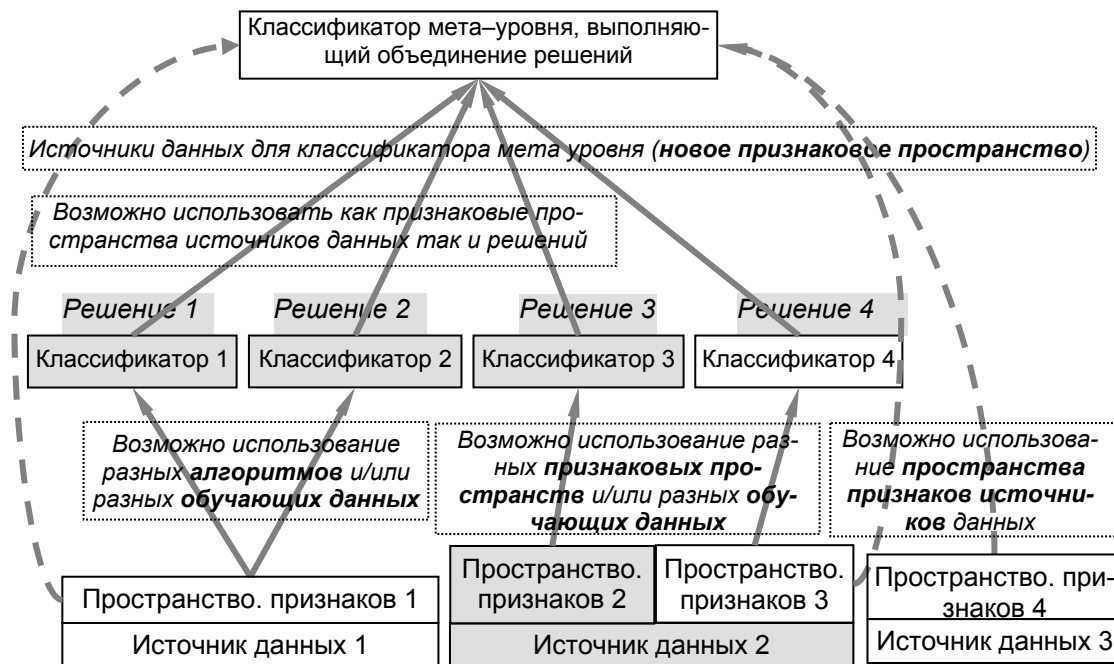


Рис. 8. Источники разнообразия классификаторов и общая схема принятия решений.

объединения решений, а при программной реализации соответствующей системы — ее архитектуры. В некоторых случаях различные алгоритмы коллектива, решения которых так или иначе необходимо объединять, могут на выходах иметь данные различных типов: наряду с категориальными значениями, которые указывают метку класса, они могут быть числовыми, когда выходом алгоритма является та или иная мера уверенности о принадлежности классифицируемого объекта тому или иному классу из числа возможных, бинарные, когда, например, классификаторы настроены на распознавание объектов только одного, «своего» класса на фоне объектов всех остальных классов (именно такой случай рассматривается в различных моделях распознавания, построенных на основе парадигмы искусственных иммунных сетей) и др. Пространство признаков может включать признаки различных уровней абстракции, что сильно влияет выбор структуры объединения решений. Типичным примером может служить ситуация, когда наряду с сигналами, на вход системы распознавания подаются изображения или текстовые сообщения. Существенное влияние на выбор структуры могут оказывать характеристики точности и достоверности принятия решений разными классификаторами, характеристики вероятностей ложной тревоги и пропуска сигналов.

Независимо от используемых алгоритмов объединения решений, основные типы компонент, на основании которых выбираются структуры объединения решений и архитектуры программных систем рассматриваемого типа, могут иметь вид, представленный на рис. 9а–9д). Рассмотрим их кратко.

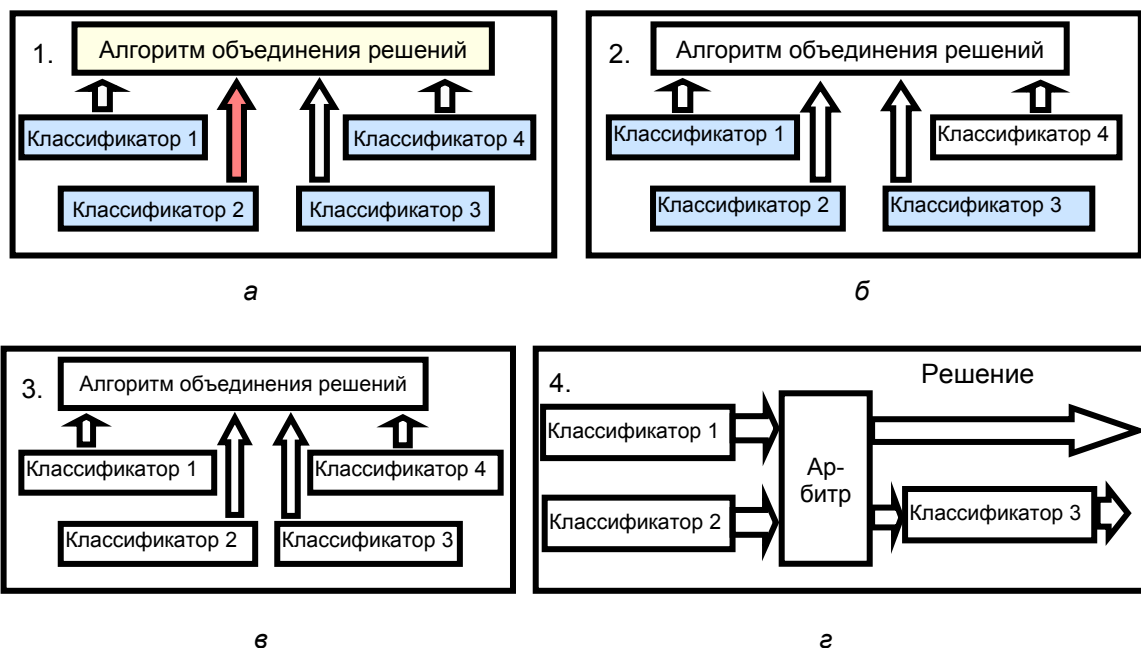


Рис.9. Варианты простейших структур объединения решений, которые могут комбинироваться в различные более сложные структуры.
 а — выбор классификатора; б — объединение решений классификаторов; в — комбинирование стратегий а и б; г — смешанная стратегия (параллельная/последовательная)

- 1) *Выбор классификатора (параллельная схема, рис. 9а).* В ней каждый классификатор рассматривается как «экспертом», компетентный в некоторой локальной области признакового пространства. В этом случае основная задача (а иногда и основная проблема) состоит в определении этого эксперта для конкретного варианта входных данных;
- 2) *Объединение решений классификаторов (параллельная схема, рис. 9б).* В этой схеме решения всех классификаторов рассматриваются как одинаково компетентные и объединение решений выполняется на основании правил, формируемых на основе обучения по мета-данным (раздел 5);
- 3) *Объединение в единой структуре двух вышеуказанных стратегий (параллельная схема, рис. 9в):* В этой структуре, для заданного входа, производится выбор нескольких наиболее компетентных классификаторов, решения которых далее объединяются;
- 4) *Параллельно-последовательная схема (рис. 9д).* В ней сначала из множества классификаторов выбирается наиболее компетентный (для заданного входа), а затем программа арбитр определяет, является ли принятое решение окончательным, либо нужно привлечь еще дополнительные процедуры классификации.

Использование древовидной структуры классификаторов (классификационное дерево) представляет собой еще один пример последовательной стратегии объединения решений с динамическим выбором классификатора. Рис. 10 демонстрирует классификационное дерево.

Схемы объединения решений также могут отличаться тем, как происходит выбор наиболее компетентного классификатора:

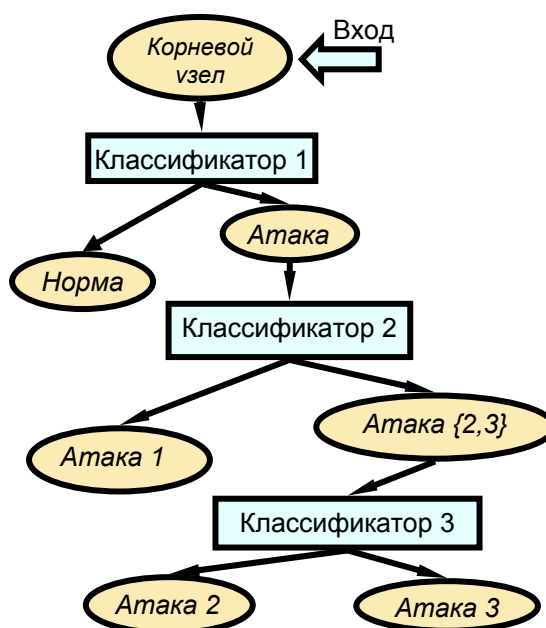


Рис.10. Древоподобная структура классификаторов.

- 1) *статический выбор областей компетентности классификаторов.* Области компетентности классификаторов определяются до фазы их использования, и в дальнейшем области компетентности уже не изменяются. Назначение классификаторам областей компетентности происходит после их обучения;
- 2) *динамический выбор областей компетентности классификаторов.* В этом случае выбор классификатора для конкретного входа выполняется во время его работы. Как правило, этот выбор делается на основании специальных механизмов определения компетентности классификаторов (например, [8, 45, 55], и т.д.). Такие примеры даются алгоритмами, в которых правила определения компетентного классификатора получаются на основе обучения (рис. 4), например, как это описано в работе [55]. Другой пример динамического выбора классификатора дает классификационное дерево (рис. 10).

9. Разнообразие классификаторов

Разнообразие множества классификаторов является его важным свойством, которое определяет возможность получения положительного эффекта от совместного использования этих классификаторов, а именно, повышение точности классификации. Понятие разнообразия классификаторов содержательно определяется как некоторая мера, которая характеризует некоррелированность ошибок различных классификаторов множества. Иначе говоря, если классификаторы выдают ошибочные решения на различных подмножествах примеров пространства признаков, то это дает некоторую гарантию того, что коллективное решение будет обладать большей точностью, чем точность наилучшего классификатора используемого множества. Это достигается за счет того, что на множестве ошибочных решений некоторого классификатора другие классификаторы не ошибаются. С другой стороны, свойство разнообразия классификаторов позволяет также обнаруживать те из них, которые бесполезны в рассмат-

риваемом коллективе решателей, поскольку их использование не ведет к повышению точности работы системы в целом. Некоторый классификатор будет бесполезным в коллективе решателей, если множество, на котором он не ошибается, является подмножеством аналогичного множества другого классификатора. Такие классификаторы могут быть удалены без ущерба для точности работы системы в целом, однако такой шаг будет способствовать снижению требований к ресурсам процессора.

Наилучшим вариантом является использование классификаторов с независимыми ошибками классификации, когда все классификаторы ошибаются на различных экземплярах объектов (в различных областях пространства признаков). Рис. 1 демонстрирует именно такой случай. Но, к сожалению, не так просто построить такие классификаторы, особенно на «плохих» наборах обучающих и тестовых данных¹.

В настоящее время ведутся активные исследования, имеющие целью формализацию понятия разнообразия классификаторов в терминах «мер разнообразия» а также по ее использованию в задачах предсказания точности работы того или иного множества классификаторов. Основные результаты по этой проблеме, полученные к настоящему времени, наиболее полно сформулированы в работе [45]. Дадим их краткую характеристику.

- 1) В случае, если выходы классификаторов представляются в форме «мягких меток», и объединение выходов производится путем усреднения, или правил упорядочения наподобие *Min rule*, *Max rule*, и т.д., ошибка классификации зависит от корреляции между оценками мягких меток. Основным выводом в связи с этим является утверждение о том, что чем меньше корреляции на множестве мягких меток, тем лучшей точностью будет обладать множество классификаторов, т.е. лучшие результаты могут быть достигнуты в случае, когда имеется отрицательная корреляция между метками, выдаваемыми классификаторами;
- 2) Для случая когда выходы классификаторов базового уровня представляют собой метки не являются мягкими, т.е. классификаторы выдают в качестве решения имя класса, можно отметить 4 уровня разнообразия: (1) уровень 1, когда не более чем один классификатор выдает неверное решение; (2) уровень 2, в котором для каждого входного вектора данных могут быть неверными решения не более, чем $\lfloor L/2 \rfloor^2$ классификаторов (голосование большинством всегда будет верным); (3) уровень 3, когда по крайней мере один классификатор выдает верное решение для каждого входного вектора данных и (4) уровень 4, в котором существуют такие входы, для которых не существует классификаторов, выдающих правильные решения.

В работе [45], рассматриваются восемь мер разнообразия (зависимости) классификаторов (можно утверждать, что свойство зависимости множества классификаторов является противоположным свойству разнообразия). Эти меры разделены на две группы:

- 1) меры попарного разнообразия (Q-statistics, корреляция, рассогласование и двойная ошибка);

¹Необходимым условием для обучающих данных быть «хорошими» является условие более или менее равномерного распределения обучающих данных в признаковом пространстве.

² $\lfloor a \rfloor$ обозначает наибольшее целое меньшее или равное a .

- 2) не парные меры (энтропия голосования, индекс «сложности», рассеивание Ковачи–Вольперта (Kohavi–Wolpert variance) и др.

Точное описание указанных зависимостей могут быть найдены в цитируемой работе, где представлены результаты эмпирического исследования указанных мер зависимостей применительно к объединению решений на основе голосования.

Вообще говоря, к настоящему моменту данная проблема изучена достаточно слабо и нуждается в дальнейшем глубоком исследовании. Основное достижение в этой части состоит в том, что осознана ее важность в рамках проблемы объединения решений.

10. Заключение

«К настоящему моменту, область задач исследования, касающаяся методов объединения классификаторов является притягательной для исследователей и в ней имеется много убедительных результатов» [34]. В последнее время появилось много различных стратегий и подходов к объединению решений. Большинство из них рассмотрены в этом обзоре.

С того времени, как была предложена сама идея объединения решений множества классификаторов, было получено несколько фундаментальных результатов. Главным из них является то, что «вместо поиска лучшего признакового пространства и лучшего классификатора, в качестве задачи рассматривается поиск наилучшего множества классификаторов и затем поиск наилучшего метода объединения. Можно предположить, что в ближайшее время актуальной станет задача поиска наилучшего множества методов объединения и наилучшего способа их совместного использования» [33]. Действительно, в то время как для традиционной классификационной задачи наиболее важным является поиск лучшего (индивидуального) классификатора, т.е. того, который обладает лучшей точностью и удовлетворяет требованиям к ошибкам первого и второго рода (т.е. требованиям к вероятностям ложной тревоги и пропуска сигнала), в рамках задачи объединения решений свойство точности индивидуальных классификаторов не так важно. Наиболее важными являются мета-свойства множества классификаторов. Одним из примеров таких мета-свойств является фактор точности от их совместного использования (combined coverage factor), косвенно представленный в терминах требований к свойству разнообразия выбранного множества классификаторов. Результатом этого требования является то, что в задаче объединения решений, индивидуальные классификаторы могут быть «не так хороши»¹, но, работая в команде как единое целое, они обладают хорошей точностью и низкими требованиями к вычислительным ресурсам. Такая же идея строго доказана в работах [41] и [42].

Таким образом, объединение решений представляет собой специальную проблему в области распознавания и классификации, которая не сводится к обычной задаче классификации (с одним классификатором). Она подлежит дальнейшему углубленному исследованию, а ее использование на практике может привести к качественно лучшим свойствам систем классификации, которые используют идею совместного использования множества классификаторов, формирующих коллективное решение.

¹Salvatore Stolfo [62] использует для таких классификаторов термин "lightweight classifiers".

Благодарности

Данный обзор был выполнен в рамках проекта №236 "Многоагентная технология распределенного обучения систем поддержки принятия решений" программы фундаментальных исследований РАН № 14 «Фундаментальные проблемы информатики и информационных технологий» по Направлению 1: «Интеллектуальные технологии и математическое моделирование», а также при поддержке Российского фонда фундаментальных исследований, проект № 04–01–00494а "Многоагентные модели и распределенные алгоритмы оценки и прогнозирования ситуаций".

Литература

1. *Бешелев С. Д., Гурвич Ф. Г.* Экспертные оценки. М.: Наука, 1973. 159 с.
2. *Бонгард М. М.* Проблема узнавания. М.: Наука, 1967. 320 с.
3. *Воробьев Н. Н.* Вопросы математизации принятия решений на основе экспертных оценок // Материалы IV симпозиума по кибернетике. 1972. Ч. 3. С. 47–51.
4. *Глушков В. М.* О прогнозировании на основе экспертных оценок // Кибернетика. 1969. № 2. С. 2–4.
5. Математические методы в социальных науках / Под ред. П. Лазарфельда. и Н. М. Генри: Прогресс, 1973. 352 с.
6. *Льюс Р., Райфа Х.* Игры и решения. М.: Изд-во иностранной литературы, 1961. 642 с.
7. *Миркин Б. Г.* Проблема группового выбора. М.: Наука, 1974. 342 с.
8. *Растрюгин Л. А., Эренштейн Р. Х.* Метод коллективного распознавания. Библиотека по автоматике. Вып. 615. М.: Энергоиздат, 1981. 78 с.
9. *Растрюгин Л. А., Эренштейн Р. Х.* Коллектив алгоритмов для обобщения алгоритмов решения задач // Известия АН СССР. Техническая кибернетика. 1978. № 2. С. 116–126.
10. *Растрюгин Л. А., Эренштейн Р. Х.* Принятие решений коллективом решающих правил в задачах распознавания образов // Изв. АН СССР. Автоматика и телемеханика. 1975. № 9. С. 134–144.
11. *Растрюгин Л. А., Эренштейн Р. Х.* Коллектив алгоритмов // Материалы Международной объединенной конференции по искусственному интеллекту. Том 3. М., 1975. С. 138–144.
12. *Растрюгин Л. А., Эренштейн Р. Х.* Обучение коллектива решающих правил // Адаптивные системы. 1974. Вып. 4. 1974. С. 8–20.
13. *Розенблат Ф.* Принципы нейродинамики. М.: Мир, 1965. 196 с.
14. *Комарцова Л. Г., Максимов А. В.* Нейрокомпьютеры. М.: Изд-во МГТУ имени Баумана, 2002. С. 115–117.
15. *Ali K., Pazzani M.* Error Reduction through Learning Multiple Descriptions // Machine Learning. 1996. No. 24(3). P. 173–202.
16. *Sharkey A. J. C.* Combining Artificial Neural Nets: Ensembles and Modular Multi-net Systems. Berlin: Springer-Verlag, 1999. P. 1–30.
17. *Bay S. D., Pazzani M. J.* Characterizing Model Error and Differences // Proceedings of 17th International Conference on machine learning (ICML-2000). San Francisco: Morgan Kaufmann, 2000. P. 49–56.
18. *Blin J., Fu K., Whinston A.* Application of Pattern Recognition to some Problems in Economics // Techniques of Optimization / Ed. A. Balakrishnan. 1972. No. 416. P. 1–18.
19. *Breiman L.* Bagging Predictors // Machine Learning. 1996. No. 24 (2). P. 123–140.
20. *Breiman L.* Stacked Regression // Machine Learning. 1996. No. 24(1). P. 49–64.
21. *Buntine W. L.* A Theory of Learning Classification Rules: Ph.D thesis, 1990. P. 172. (University of Technology, School of Computing Science, Sydney, Australia)
22. *Clark P., Niblett T.* The CN2 Induction Algorithm // Machine Learning Journal. 1989. No. 3. P. 261–283.
23. *Clemen R.* Combining Forecasts: A Review and Annotated Bibliography // International Journal of Forecasting. 1989. No. 5. P. 559–583.
24. *Cost S., Salzberg S.* A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features // Machine Learning. 1993. No. 10(1). P. 57–78.
25. *Condorcet N. C.* Essai sur l'application de l'analyse à la Probabilité des Décisions rendues a la Pluralité des voix. Paris : L'Imprimerie Royale, 1785.

26. *Dietterich T.* Machine Learning Research: Four Current Directions // *AI Magazine*. 1997. No. 18(4). P. 97–136.
27. *Fix E., Hodges J.* Nonparametric Discrimination // Technical Report No. 11. USAF School of Aviation Medicine, Rendolph Field, Texas, 1952.
28. *Freund Y., Shapire R.* Experiments with a New Boosting Algorithm // *Machine Learning. Proceedings of the 13th International Conference*. San Francisco: Morgan Kaufmann, 1996. P. 148-156.
29. *Gama J., Brazdil P.* Cascade Generalization // *Machine Learning*. 2000. No. 41(3). P. 315–342.
30. *Gorodetski V., Karsayev O.* Algorithm of Rule Extraction from Learning Data // *Proceedings of the 8th International Conference Expert Systems & Artificial Intelligence (EXPERTSYS-96)*. 1996. P. 133–138.
31. *Hashem S.* Optimal linear combination of neural networks: Ph.D. thesis, 1997. (Purdue University, School of Industrial Engineering, Lafayette)
32. *Ho T. K., Hull J. J., Shirari S. N.* Decision Combination in Multiple Classifier Systems // *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 1994. No. 16(1). P. 66–75.
33. *Ho T. K.* Multiple Classifier Combination: Lessons and Next Step // *Hybrid Methods in Pattern Recognition*, World Scientific, 2002. P. 17.
34. *Huang T., Hess C., Pan H., Liang Zhi-Pei.* A Neuronet Approach to Information Fusion [Электронный ресурс] // <http://viola.usc.edu/paper/mmsp97/contents/papers/056/index.htm> (по состоянию на 21.04.06).
35. *Jacobs R., Jordan M., Nowlan S., Hinton G.* Adaptive Mixtures of Local Experts [Электронный ресурс] // <http://cs.toronto.edu/~hinton/absps/jjnh91.html> (по состоянию на 21.04.06).
36. *Jordan M., Jacobs R.* Hierarchical mixtures of experts and the EM algorithm // *Neural Computations*. 1994. No. 6(2). P. 181–214.
37. *Kanal L.* Interactive Pattern Analysis and Classification. Survey and Commentary // *Proceedings of IEEE*. 1972. Vol. 60, no.10. P. 1200–1215.
38. *Kittler J., Hatef M., Duin R. P. W., Matas J.* On combining classifiers // *IEEE Transactions on pattern Analysis and Machine Intelligence*. 1998. No. 20(3). P. 226–239.
39. *Kittler J., Roli F. (Eds.)*. Multiple Classifier Systems // *Lecture Notes in Computer Science*. 2000. No. 1857.
40. *Kittler J., Roli F. (Eds.)*. Multiple Classifier Systems // *Lecture Notes in Computer Science*. 2001. No. 2096.
41. *Kleinberg E. M.* Stochastic Discrimination // *Annals of Mathematics and Artificial Intelligence*. 1990. No. 1. P 207–239.
42. *Kleinberg E. M.* A Mathematically Rigorous Foundation for Supervised Learning // *Lecture Notes in Computer Science*. 2000. No. 1857. P. 67–78.
43. *Kuncheva L., Bezdec J., Duin R. P. W.* Decision Templates for Multiple Classifier Fusion // *Pattern Recognition*. 2001. No. 34(2). P 299–314.
44. *Kuncheva L.* Switching between Selection and Fusion in Combining Classifiers: An Experiment // *IEEE Transactions On Systems Man And Cybernetics, Part B-cybernetics*. 2002. Vol. 32, no. 2. P. 146–156.
45. *Kuncheva L., Whitaker C.* Measures of diversity in Classifier Ensembles // *Machine Learning*. 2003. No. 51. P 181–207.
46. *Lee D. S.*, A Theory of classifier combination: the neural network approach: Ph.D. thesis, 1995. (Dept. of Computer Science, State University of New York at Buffalo)
47. *Lipnickas A., Korbicz J.* Adaptive selection of neural networks for a committee decision // *Proceedings of the IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application*. 2003. P. 109–114.
48. *McKay C.* Classifier Ensembles: A Practical Overview [Электронный ресурс] // http://music.mcgill.ca/~cmckay/software/computer_science/ClassifierEnsembles/greeting_page.html (по состоянию на 21.04.06).
49. *Merz C.* Combining classifiers using correspondence analysis // *Advances in Neural Information Processing*. 1997. No. 10.
50. *Merz C., Murphy P.* UCI Repository on Machine Learning Databases. [Электронный ресурс] // http://ics.uci.edu/mllearn/MLR_repository.html (по состоянию на 21.04.06).
51. *Michalski R. A.* Theory and Methodology of Inductive Learning // *Machine Learning*, 1983. № 1. P. 83–134.
52. *Murthy S. K., Kasif S., Salzberg S., Beigel R.* OC1: Randomized Induction of oblique decision trees // *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*. 1993. P. 322–327.

53. *Moerland P.* Mixtures of experts estimate a posteriori probabilities // Proceedings of The International Conference on Artificial Neural Networks (ICANN'97). 1997. P 499–504.
54. *Niyogi P., Pierrot J-B., Siohan O.*, Multiple Classifiers by constrained minimization // Proceedings of International Conference on Acoustics, Speech, and Signal Processing. 2000.
55. *Ortega J., Coppel M., Argamon S.* Arbitrating Among Competing Classifiers Using Learned Referees // Knowledge and Information Systems. 2001. No. 3. P 470–490.
56. *Perrone M., Cooper L.* When networks disagree: Ensemble methods for hybrid neural networks // Artificial Neural Networks for Speech and Vision. New York: Chapman & Hall, 1993. P. 126-142.
57. *Prodromidis A., Chan P., Stolfo S.* Meta-learning in distributed data mining systems: Issues and approaches // Advances in Distributed Data Mining. AAAI Press, 1999.
58. *Quinlan R.* C4.5 Programs for Machine Learning. San Francisco: Morgan Kaufmann, 1993. P. 302.
59. *Rumelhart D. E., Hinton G. E., Williams R. J.* Learning internal representation by error propagation // Parallel Distributed Processing: Exploration of the microstructure of cognitions. MIT Press. 1986. No. 1. P. 318–362.
60. *Seewald A., Fuernkranz J.* An evaluation of grading classifiers // Proceedings of 4th International Conference Intelligent data Analysis, LNCS 2189. 2001. P 115–124.
61. *Stolfo S., Chan P.* A comparative evaluation of voting and meta-learning on partitioned data // Proceedings of Twelfth 4th International Conference on Machine Learning. 1995. P 90–98.
62. *Ting K.* The characterization of predictive accuracy and decision combination // Proceedings of 13th International Conference on Machine Learning. 1996. P 498–506.
63. *Ting K., Witten I.* Issues in stacked generalization // Journal of Artificial Intelligence Research. 1999. No. 10. P 271–289.
64. *Todorovski L., Dzeroski S.* Combining classifiers with meta decision trees // Proceedings of 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD-2000). 2000. P 54–64.
65. *Todorovski L., Dzeroski S.* Combining multiple models classifiers with meta decision trees // Machine Learning Journal. 2003. Vol. 50(3). P. 223–249.
66. *Vilalta R., Drissi Y.* A perspective view and survey of meta-learning // Journal of Artificial Intelligence Review [Электронный ресурс] / <<http://research.ibm.com/people/v/vilalta/papers/jaireview01.ps>> (по состоянию на 21.04.06).
67. *Waterhouse S., Robinson A.* Classification using hierarchical mixtures of experts // Proceedings of IEEE Workshop on Neural Networks for Signal Processing. 1994. P 177–36.
68. *Woods K., Kegelmeyer W. P., Bowyer K.* Combination of multiple classifier using local accuracy estimates // IEEE Transactions on Pattern Analysis and Machine Intelligence 1997. No. 19. P 405–410.
69. *Dietterich T.* Ensemble Methods in Machine Learning // Multiple Classifier Systems. 2001 Vol. 1857. P 1–15.
70. *Wolpert D.* Stacked generalization // Neural Network. 1992. No. 5(2). P 241–260.