

ОСНОВНЫЕ ПРИНЦИПЫ ПОСТРОЕНИЯ БИОМЕТРИЧЕСКИХ СИСТЕМ ПРЕДУПРЕЖДЕНИЯ НЕСАНКЦИОНИРОВАННОГО ИСПОЛЬЗОВАНИЯ ВОЗДУШНЫХ СУДОВ

С. Б. Рудницкий, В. А. Дюк

Санкт-Петербургский институт информатики и автоматизации РАН

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

<rouditsky@spiiras.nw.ru>

УДК 681.3

Рудницкий С. Б., Дюк В. А. Основные принципы построения биометрических систем предупреждения несанкционированного использования воздушных судов // Труды СПИИРАН. Вып. 4. — СПб.: Наука, 2007.

Аннотация. *Рассматриваются различные аспекты построения биометрических систем для борьбы с терроризмом на воздушном транспорте. — Библ. 6 назв.*

UDC 681.3

Rouditsky S. B., Duke V. A. General principles of the biometric systems building for the warning the unauthorized use air court // SPIIRAS Proceedings. Issue 4. — SPb.: Nauka, 2007.

Abstract. *Different aspects of the biometric systems building for fight with terrorism on air transport are considered. — Bibl. 6 items.*

1. Введение

Под несанкционированным использованием воздушного судна (ВС) будем понимать функционирование ВС в условиях проведения террористических акций с захватом террористами ВС.

Теракт 11.09.2001 в США, августовские теракты 2004 г. в России вынудили пересмотреть и повысить требования к авиационной безопасности [1]. Однако, как отмечает представитель ФСБ России В. Шавлюга, около 70% технических средств обеспечения авиационной безопасности в аэропортах нашей страны морально и физически устарели, на их обновление требуется сумма порядка 100 млн долл. Таких средств у аэропортов нет, так что без помощи государства здесь не обойтись. Существует еще одна проблема — на законодательном уровне до сих пор нечетко прописаны требования к авиационной безопасности. Кроме того, даже при наличии финансирования должен пройти достаточно большой период времени до получения ожидаемых результатов. Проблема же требует безотлагательного решения.

В такой ситуации целесообразно применить новые технические решения, основанные на других физических принципах. В этой связи представляется перспективным широкое внедрение в систему безопасности на воздушном транспорте комплекса биометрических систем, как предотвращающих проникновение на борт террористов, так и автономно вырабатывающих управляющие сигналы на перехват ручного управления интеллектуальной автоматической системой управления (АСУ) ВС.

В последнее время Международной организацией гражданской авиации (ИКАО) наряду с традиционными средствами повышения безопасности полетов: системой предупреждения столкновений, системой сигнализации опасной близости земли, системой автоматического захода на посадку и категорирован-

ной посадки и ряда других — инициирована программа по разработке технических средств антитеррористической направленности. В частности, ведутся работы по разработке специального устройства в составе навигационно-пилотажного комплекса ВС, которое не позволит пилоту войти в зону атаки, например городского центра, атомной станции и т.п. [2].

В компьютер навигационно-пилотажного комплекса закладываются координаты запретных зон, при подлете к которым АСУ ВС перехватывает ручное управление и уводит самолет от опасных зон, вплоть до обеспечения вынужденной посадки. Однако при этом существует ряд проблем, связанных с анализом нештатной ситуации и подтверждением пилотом или диспетчером наземного пункта управления воздушным движением (УВД) случая захвата ВС террористами. Очевидно, что такое построение системы не обеспечивает гарантированного предотвращения теракта, так как пилот может быть или принуждаем террористами, или сам террористом, а канал связи с наземным пунктом УВД может быть заблокирован тем или иным способом. В этой связи представляется актуальным выработать рекомендации и определить облик автономной скрытой дистанционной системы оценки нештатной ситуации на основе анализа биометрических данных экипажа и пассажиров.

2. Методы биометрических измерений

Под биометрией в широком смысле понимают раздел науки, связанный с различными измерениями биологических объектов и анализом таких измерений. Биометрия в узком смысле ограничена решением задач идентификации индивидуума или его функционального состояния по его уникальному «биологическому портрету». На этом последнем приложении биометрии сконцентрировано внимание современных исследований, которые проводятся с целью разработки разнообразных систем безопасности. Актуальность подобного приложения сегодня не нуждается в комментариях.

Основная особенность биометрических методов идентификации состоит в их статистической природе. При наличии запоминаемого кода устройство выдает ответ типа «совпадает / не совпадает». В случае считывания копии того или иного биологического кода и его сравнения с аналогом речь идет о вероятности ошибки, которая является функцией измерительного прибора.

Таким образом, качество биометрической системы определяется двумя техническими характеристиками: вероятностью ошибочного допуска «чужого» (ошибка первого рода) и соответственно ошибочного задержания «своего» (ошибка второго рода).

В качестве уникального биологического кода человека может использоваться целый ряд параметров. Их можно разделить на физиологические, основанные на анатомической уникальности каждого человека, и на поведенческие, основанные на специфике действий человека. Каждый из параметров имеет свои достоинства и недостатки с точки зрения его использования в качестве критерия идентификации.

Эволюция биометрических систем безопасности началась от идентификации личности по какому-либо одному статическому образцу (например, по папиллярному узору на пальце человека) и происходит в направлении увеличения количества и интеграции идентификационных показателей. Так, современные комплексированные системы безопасности второго поколения анализируют различные информативные признаки голоса, лица, руки, глаза, объемного

пульса и пр. В этих системах повышается уровень сложности алгоритмов обработки исходной биометрической информации. Однако регистрация и анализ многомерного динамического образа человека позволяют повысить надежность его идентификации и снижает вероятность возможных фальсификаций.

3. Дистанционное измерение (классификация) проявлений психического

Развитие технических и алгоритмических средств многомерных динамических биоизмерений создают предпосылки для расширения круга задач, решаемых биометрическими системами. Одним из перспективных направлений развития биометрии является анализ функциональных состояний человека по дистанционному исследованию его физиологических и поведенческих характеристик. В частности, сюда относится задача дистанционного распознавания психического состояния человека и вероятностного прогнозирования его намерений.

Разрабатываемые сегодня биометрические системы безопасности третьего поколения, призванные решать указанную задачу, как и системы второго поколения, должны быть построены по принципам интегрированных комплексов. Вместе с тем уровень сложности таких комплексов повышается на порядок. Он определяется, с одной стороны, структурной сложностью и полиморфностью распознаваемых конструкторов, связанных с психической деятельностью биологических объектов. С другой стороны, сужается круг доступных для дистанционного измерения потенциальных корреляторов с внешними проявлениями психического. Этим обуславливается смещение акцента в задаче построения биометрических систем безопасности третьего поколения в сторону усложнения алгоритмов комплексной обработки большого числа информативных признаков комбинаций из нескольких, одновременно анализируемых биометрических параметров методами новейших информационных технологий, способных «выжимать» из регистрируемых данных скрытые закономерности, дающие максимум полезных сведений о психическом состоянии биологического объекта.

4. Задачи, решаемые биометрической системой

По своим функциям биометрическая система — это система распознавания шаблона, которая устанавливает аутентичность конкретных физиологических или поведенческих характеристик пользователя. Логически биометрическая система может быть разделена на два модуля: модуль регистрации и модуль идентификации.

Модуль регистрации отвечает за «обучение» системы идентифицировать конкретного человека или его функциональное состояние. На этапе регистрации биометрические датчики снимают информативные признаки человека для того, чтобы создать цифровое представление. Специальный модуль обрабатывает это представление, с тем чтобы выделить характерные особенности и сгенерировать более компактное и выразительное представление, называемое шаблоном. Шаблон для каждого пользователя хранится в базе данных биометрической системы. Эта база данных может быть централизованной или распределенной, когда шаблон каждого пользователя сохраняется на смарт-карте и передается пользователю.

Модуль идентификации отвечает за распознавание человека и/или его функционального состояния. На этапе идентификации биометрический датчик снимает характеристики человека, идентификация которого проводится, и преобразует эти характеристики в тот же цифровой формат, в котором хранится шаблон. Полученный шаблон сравнивается с хранимым шаблоном, с тем чтобы определить, соответствуют ли эти шаблоны друг другу.

Идентификация может выполняться в виде верификации, аутентификации (проверка утверждения типа «Я — Иванов») или распознавания, определяя личность человека из базы данных известных системе людей (определение того, кто я, не зная моего имени). В верификационной системе, когда полученные характеристики и хранимый шаблон пользователя, за которого себя выдает человек, совпадают, система подтверждает идентичность. В системе распознавания, когда полученные характеристики и один из хранимых шаблонов оказываются одинаковыми, система идентифицирует человека с соответствующим шаблоном.

Биометрическая система, используемая для борьбы с терроризмом на воздушном транспорте, должна решать следующие задачи:

- биометрической идентификации членов экипажа, обслуживающего персонала и пассажиров в аэропорту при допуске на борт ВС;
- биометрической дистанционной (скрытой) идентификации психического состояния и определения намерений членов экипажа, обслуживающего персонала и пассажиров при посадке на борт и во время полета;
- оценки ситуации на борту по данным дистанционной (скрытой) биометрии и выработки управляющих сигналов для АСУ ВС, запрещающей полет над заданной территорией, вплоть до вынужденной посадки на определенный аэродром.

Для решения указанных задач необходимо:

- провести выбор и обоснование идентификационных параметров;
- определить структуру комплексной биометрической системы идентификации личности и ее намерений;
- разработать алгоритмы комплексной обработки биометрической информации и технологии обработки многомерных данных, соответствующие поставленным задачам;
- провести анализ возможных нештатных ситуаций и на его основе сформировать иерархическую структуру приоритетов для АСУ ВС, включая системы предупреждения столкновений, опасной близости Земли и близости зон, запретных для пролета над ними.

5. Анализ состояния дел в исследуемой предметной области

Работы в этом направлении исследований ведутся достаточно интенсивно, особенно в последнее время, в связи с реализацией широкомасштабной программы борьбы с международным терроризмом. Однако пока достижения практического плана касаются только биометрических систем идентификации самой личности, а не её намерений. Наблюдается переход от анализа одного из биометрических идентификационных параметров (отпечаток пальца, руки, фото / термография лица, включая его виброизображение, состояние радужной оболочки глаза и др.) к их совокупности, то есть переход к комплексированным биометрическим системам, правда, на сегодня без серьезных попыток ком-

плексной обработки первичных (так называемых «сырых») измерений таких многопараметрических систем. Тем не менее биометрические системы третьего поколения, которые ещё только предстоит разрабатывать, продекларированы именно как системы идентификации намерений. Здесь же, во всяком случае на первых этапах, без глубокого изучения проявлений электрической активности мозга и соответствующих им функциональных состояний сознания обойтись не удастся, по крайней мере на этапе поиска коррелятов психического состояния человека его физиологическим и поведенческим реакциям.

В природе биологические объекты используют различные дистантные способы для идентификации друг друга, оценки настроения, эмоций и прогнозирования намерений. Люди делают это по лицам и их выражениям, по различным вегетативным реакциям, по голосу, речевой и двигательной активности, по манерам поведения и т.п. Животные получают информацию друг о друге по запахам, позам, по издаваемым звукам, посредством эхолокации (летучие мыши, дельфины) и другим физическим агентам.

Технические системы в ряде случаев значительно уступают биообъектам в чувствительности к продуктам секреции — биологически активным веществам и в способности живых организмов содержательно интерпретировать трудноформализуемые динамические проявления психического. Тем не менее в отдельных случаях технические системы начинают приближаться по эффективности к решениям живой природы.

Так, в области химического аналитического приборостроения известна, например, разработка кафедры аналитической химии Воронежской технологической академии, где сконструирован стендовый образец «электронного носа», который по своей чувствительности близок к аналогичному человеческому органу [3]. Все более широкое применение для исследования взаимосвязи субфракционного состава в полидисперсных нативных биологических жидкостях с функциональными состояниями находит метод корреляционной лазерной спектроскопии (СПб. Институт ядерной физики, ООО «Интокс», СПбГМУ им. акад. И.П. Павлова). Широкий спектр высокоточных приборов для анализа проб воздуха и жидких сред предлагает, например, предприятие «ЛЮМЭКС» (Санкт-Петербург). Ведутся исследования по дистанционному контролю и оценке психоэмоционального состояния человека по специфическим показателям речевой активности (см., например, разработки Института медико-биологических проблем [4]) и др.

Наибольшее соперничество с природой в искусстве узнавать биообъекты, описывать и прогнозировать их состояние и поведение составляют, конечно, достижения психофизиологии, опирающиеся на количественные физические измерения. Психофизиологические исследования создали теоретический базис для построения систем оценки различных психических состояний по величине сопутствующих вегетативных реакций (изменение сердечного ритма, кожных потенциалов, кровяного давления, дыхания и т.п.). Отдельные современные разработки, основанные на обширном экспериментальном материале, обращают на себя внимание весьма высокой точностью и надежностью оценок. Так, на кафедре психиатрии Санкт-Петербургской медицинской академии последипломного образования построены шкалы свойств психики и степени отклонения в психическом статусе по показателям регуляции кардиоритма, дающие от 80 до 90% совпадений с клиническими феноменологическими оценками [5].

Оценка психического состояния биологического объекта по показателям квазипериодических процессов, протекающих в живом организме, представляет

особый интерес для построения систем безопасности третьего поколения. Несмотря на то что амплитуды таких процессов, отражающихся на поверхности тела, незначительны и имеют порядок долей миллиметра, подобные колебания поддаются дистанционному мониторингу с помощью современных технических средств. В частности, для дистанционной регистрации столь малых по амплитуде перемещений частей тела биологического объекта в настоящее время применяют телевизионные камеры с высоким пространственным разрешением и специальной обработкой принимаемого сигнала, связанной с накоплением межкадровой разности.

Примером технического решения задачи дистанционной регистрации колебательных процессов биологического объекта может служить система AURA-VibrImage предприятия «ЭЛСИС» (Санкт-Петербург) [6]. Встроенная в систему программа Biolmage получает последовательность кадров изображения объекта, производит вычитание двух последовательных кадров и последующую обработку межкадровой разности. Обработка изображения включает накопление суммы разностей (от 2 до 30 последовательных кадров изображения).

Разработчики AURA-VibrImage и других аналогичных систем утверждают, что полученное при обработке изображение может быть использовано для получения информации о физиологическом или психофизическом состоянии живого объекта. Например, отмечается, что насыщенность изображения человека, находящегося в состоянии нервного возбуждения, под действием алкоголя или наркотика, значительно больше, чем у человека, находящегося в спокойном состоянии. Вместе с тем подобные утверждения, опираясь на отмеченный выше неоспоримый теоретический базис психофизиологии, не способны пока найти практическое воплощение в виде эффективной методики оценки психического состояния биологических объектов по следующим причинам:

- сильная зависимость от внешних условий (условия съемки, качество применяемой телевизионной техники, наличие контрастных перепадов между объектами и др.);
- высокая вариабельность результатов, получаемых для различных биологических объектов (людей), обусловленная системной сложностью таких объектов;
- использование разработчиками весьма ограниченного набора параметров получаемых изображений;
- отсутствие у разработчиков представлений о новейших возможностях информационных технологий в области анализа многомерных данных.

Таким образом, создание валидных и надежных шкал оценки психических состояний на основе дистанционного измерения параметров квазипериодических колебательных процессов у биологических объектов требует разработки принципиально нового подхода к манипулированию регистрируемой информацией.

По нашему мнению, этот новый подход должен включать в себя две основные составляющие:

- 1) переход от абсолютных значений измеряемых показателей к многомерным ипсативным (автонормирующимся) данным о биологическом объекте;
- 2) поиск многомерных ипсативных инвариантов, отражающих психическое состояние биологического объекта, на основе новейших технологий анализа многомерных данных.

Первая составляющая вряд ли нуждается в развернутом комментарии. Результаты накопления межкадровой разности с различными периодами накопления представляют собой многомерный вектор, соответствующий спектру

вибраций видеоизображения. Этот спектр нормируется в простейшем случае относительно среднего или максимального значения. Тем самым в последующей обработке мы отвлекаемся от абсолютных значений спектральных составляющих и концентрируем внимание на форме огибающей спектра.

Вторая составляющая, связанная с современными технологиями обработки многомерных данных, нуждается в более подробном рассмотрении.

6. Предлагаемые методы и подходы к решению поставленных задач

6.1. Автоматизированный анализ электрической активности мозга человека

Намерения могут коррелировать с функциональным состоянием испытуемого и выявления этих корреляций могут быть опосредованными, в том числе через проявления изменений в электроэнцефалограммах. Здесь мы опираемся на результаты разработки в СПИИРАН метода анализа фрактальной динамики (АФД) электроэнцефалограмм (ЭЭГ) [7].

Метод АФД электрической активности мозга человека, предложенный профессором Р. И. Полонниковым, неплохо зарекомендовал себя для решения задач обучаемой классификации таких функциональных состояний пациента, как норма и определенные церебральные патологии. Подход к представлению структуры процессов электрической активности мозга человека в виде регулярного фрактала, регулярной биоритмической составляющей и фрактальных флуктуаций оказался в достаточной мере универсальным: он позволил, в частности, выявить некоторую общность, связанную с развитием как здорового, так и поврежденного мозга. Это дает основания полагать, что синтезированные информационные характеристики для трёх вышеуказанных составляющих окажутся полезными и для решения поставленной задачи, необходимо лишь модифицировать метод АФД применительно к новым требованиям.

Кроме того, есть основания полагать, что часть информативных признаков будет синтезирована с использованием результатов сингулярного разложения специальным образом преобразованной матрицы измерительной информации. Всего мы ожидаем получить 20–30 информативных признаков.

6.2. Метод сканирования геометрии лица

Система допуска по распознаванию лица — наиболее древний и распространенный способ. Именно такая процедура сопровождает каждого из нас, когда мы, например, предъявляем свой паспорт на пропускном пункте в международном аэропорту. Пограничник сверяет фото на паспорте с лицом владельца паспорта и принимает решение — его этот паспорт или нет. Примерно такую же процедуру выполняет компьютер, с той только разницей, что фото уже находится в его памяти. Хотя лицо человека и уникальный параметр, но достаточно динамичный; человек может улыбаться, отпускать бороду и усы, надевать очки — все это добавляет трудностей в процедуре идентификации и требует достаточно мощной и дорогой аппаратуры, что соответственно сокращает степень распространения данного метода.

В качестве примера приведем действующую систему контроля доступа на базе распознавания посетителей киосков для обналичивания чеков, установленных компанией Mr. Payroll в нескольких штатах США. По свидетельству компании, клиенты считают такую процедуру весьма удобной. При первом посещении киоска производится цифровой снимок лица клиента, который передается в сервисный центр, и впоследствии система сверяет соответствующее изображение с лицом клиента и только после этого выполняет транзакцию.

О качестве современных систем распознавания лица можно судить по высказыванию вице-президента компании Polaroid, который в качестве доступа к своему ноутбуку использует систему распознавания по лицу Facelt, разработанную компанией Visionics. «Раньше я все время забывал пароль, — говорит он, — но Facelt мое лицо не забывает. Она узнает меня в очках и без очков, с бородой и без нее».

Недавно появилось сообщение об устройствах Technology Recognition Systems (США), в которых происходит распознавание лица в инфракрасном свете. Данная технология основана на том, что термограмма лица человека (тепловая картинка, созданная излучением тепла кровеносными сосудами лица) является уникальной для каждого человека и, следовательно, может быть использована в качестве биокода для систем контроля допуска. Данная термограмма является более стабильным кодом, чем геометрия лица, поскольку не зависит от времени и изменений внешности человека.

Для вероятностного распознавания намерений, по нашему мнению [8], целесообразно использовать сочетание систем, использующих термограммы, и системы, анализирующей параметры виброизображений лица и его отдельных частей (область глаз, губ), подобной описанной выше системе Elsis.

6.3. Распознавание по голосу

Метод узнавания по голосу, так же как и метод распознавания по лицу, был известен до появления биометрии. Поэтому достоинства и недостатки метода знакомы каждому. Как не всегда по ответу на вопрос «кто там?» можно понять, что за дверью находится сосед, и тогда необходимо дополнительно заглянуть в глазок, так и техническая система может ошибаться в силу изменений голоса отдельного человека. А голос, как известно, зависит от настроения, состояния здоровья, меняется с возрастом и даже может «ломаться».

В связи с этим устройства идентификации человека по голосу широкого применения не получили.

Однако проведенные нами предварительные исследования (на 33 испытуемых) по идентификации эмоционального состояния, сопутствующего лжи, на портативном приборе True Tester (Республика Корея), принцип действия которого основан на анализе речевой активности, позволяют предположить возможность использования подобных методов для классификации функциональных состояний человека в комплексных биометрических системах.

6.4. Подходы к комплексной обработке большого массива биометрических данных

Наиболее перспективный подход к обработке данных биомедицинских исследований в настоящее время связывают с использованием технологий Data

Mining, предназначенных для обнаружения закономерностей в информации об объектах со сложной системной организацией.

В основу современных технологий Data Mining положена концепция шаблонов (паттернов) и зависимостей, отражающих многоаспектные взаимоотношения в данных. Поиск паттернов производится автоматическими методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей.

Важное положение Data Mining — нетривиальность разыскиваемых паттернов. Это означает, что они должны отражать неочевидные, неожиданные (unexpected) регулярности в данных, составляющие так называемые скрытые знания (hidden knowledge). Многие специалисты осознали, что для выявления закономерностей в реальных жизненных явлениях нужен особенный аналитический инструментарий, соответствующий их системной сложности. В свою очередь, к обществу пришло понимание, что «сырые» данные содержат глубокий пласт знаний, при грамотной раскопке которого могут быть обнаружены настоящие самородки.

Термин Data Mining, появившийся в 1978 г., оказался удачным и приобрел высокую популярность в современной трактовке примерно с первой половины 90-х годов. Поэтому вполне понятным оказалось стремление разработчиков аналитических приложений, реализующих самые различные методы и подходы, отнести себя к данной категории. Вместе с тем это не всегда обоснованно.

Например, методы традиционной математической статистики, составляющие основу статистических пакетов, полезны главным образом для проверки заранее сформулированных гипотез (verification-driven data mining) и для «грубого» разведочного анализа, составляющего основу оперативной аналитической обработки данных (online analytical processing, OLAP). Главная причина ограниченной эффективности большинства процедур для выявления взаимосвязей в данных, входящих в состав статистических пакетов, — концепция усреднения по выборке, приводящая к операциям над несуществующими величинами (например, средняя температура пациентов по больнице, средняя высота дома на улице, состоящей из дворцов и лачуг и т.п.). Так называемые «многомерные методы» типа дискриминантного, факторного и других подобных видов анализа приходят к конечному результату через операции над фиктивными векторами средних значений, а также ковариационными и корреляционными матрицами. Поэтому их результаты нередко неточны, грешат подгонкой и отсутствием смысла.

Программные продукты, реализующие нейросетевой подход, также нередко относят к категории Data Mining. Основной недостаток классической нейросетевой парадигмы заключается в том, что нейронная сеть представляет собой «серый» ящик. Во-первых, топология нейросетей здесь задается исходя из эвристических соображений. И, во-вторых, в натренированных нейросетях со сложной топологией веса сотен и тысяч межнейронных связей не поддаются анализу и интерпретации человеком.

Подход, связанный с разработкой так называемых самоорганизующихся (растущих или эволюционирующих) булевых нейросетей, структура которых поддается расшифровке в виде логических высказываний, соответствует целям и задачам Data Mining, но страдает серьезными недостатками, в целом присущими эволюционным алгоритмам.

Идея систем рассуждений на основе аналогичных случаев (case based reasoning — CBR) на первый взгляд крайне проста. Для того чтобы сделать

прогноз на будущее или выбрать правильное решение, эти системы находят в прошлом близкие аналоги наличной ситуации и выбирают тот же ответ, который был для них правильным. Поэтому этот метод еще называют методом «ближайшего соседа» (nearest neighbor). В последнее время распространение получил также термин «memory based reasoning», который акцентирует внимание на том, что решение принимается на основании всей информации, накопленной в памяти.

Системы CBR показывают неплохие формальные результаты в самых разнообразных задачах. Главным их минусом считают то, что они вообще не создают каких-либо моделей или правил, обобщающих предыдущий опыт; в выборе решения они основываются на всем массиве доступных исторических данных, поэтому невозможно сказать, на основе каких конкретно факторов CBR системы строят свои ответы. Другой, более серьезный минус заключается в произволе, который допускают системы CBR при выборе меры «близости». От этой меры самым решительным образом зависит объем множества прецедентов, которые нужно хранить в памяти для достижения удовлетворительной классификации или прогноза. Кроме того, бесосновательным выглядит распространение общей меры близости на выборку данных в целом.

В наибольшей мере требованиям Data Mining удовлетворяют методы поиска логических закономерностей в данных. Их результаты чаще всего выражаются в виде IF-THEN и WHEN-ALSO правил. С помощью таких правил решаются задачи прогнозирования, классификации, распознавания образов, сегментации баз данных (БД), извлечения из данных «скрытых» знаний, интерпретации данных, установления ассоциаций в БД и др. Логические методы работают в условиях разнородной информации. Их результаты прозрачны для восприятия.

Вместе с тем специальное тестирование известных коммерческих систем, реализующих самые распространенные подходы (деревья решений, ограниченный перебор), дало основание для следующих выводов [9]:

1. Наиболее популярные аналитические инструменты Data Mining в ряде случаев оказываются не способны решать даже простейшие задачи.
2. Применяющиеся подходы к обнаружению знаний в базах данных выявляют лишь усеченные фрагменты полных и точных логических закономерностей, скрытых в многомерных данных.
3. Инструменты для поиска логических правил в данных высокой размерности не способны отличать «ложные закономерности» от устойчивых регулярностей.
4. Известные системы для поиска логических правил не поддерживают функцию обобщения найденных правил и функцию поиска оптимальной композиции таких правил. Вместе с тем указанные функции являются весьма существенными для построения баз знаний, требующих умения вводить понятия, метапонятия и семантические отношения на основе множества фрагментов знаний о предметной области.

Все протестированные системы независимо от используемого подхода делают принципиальную ошибку уже на первом шаге своей работы.

Алгоритмы построения деревьев решений пытаются найти наилучший признак (корень дерева), который «оптимальным» образом разделяет выборку на части. Никакие математические ухищрения не способны исправить основной дефект подобного подхода, связанный с тем, что признак вырывается из целостной системы описания многомерного объекта (записи базы данных). В систе-

мах, реализующих переборный подход, принципиальная ошибка первого шага связана с поиском «оптимальной» сегментации количественных признаков.

Отмеченная проблема «первого шага» является производной от главной ключевой проблемы поиска if-then правил в данных. Эта ключевая проблема связана с тем, что здесь мы имеем дело с задачей полного перебора комбинаций элементарных логических условий. Считается, что данная задача при высокой размерности пространства признаков не может быть решена за приемлемое время в обозримом будущем даже с помощью суперкомпьютеров. Поэтому известные подходы к поиску if-then правил вынуждены использовать те или иные эвристические ограничения и их результаты нередко представляют собой «обрывки» истинных регулярностей в данных — «осколки знаний».

Таким образом, несмотря на продуктивность в целом технологий Data Mining для анализа сложно структурированной информации, данные технологии нуждаются в определенной доработке.

7. Заключение

Современное состояние биометрических и информационных технологий и проведенные нами исследования и эксперименты дают основание ожидать, что комплексные дистанционные биометрические системы позволят проводить обучаемую классификацию таких функциональных состояний пациента, как всплески эмоциональной напряженности, в том числе агрессивного характера и возникающие при даче заведомо ложных ответов на задаваемые вопросы.

Методы, алгоритмы и программное обеспечение комплексной обработки биометрической информации позволяют получить работоспособный инструмент, существенно дополняющий действующие биометрические системы и преобразующий их из систем с качественной оценкой в систему с автоматической обработкой результатов измерений и автоматической выработкой решения об отнесении объекта (человека) или нештатной ситуации к тому или иному классу.

Литература

1. Проблемы безопасности полетов №1 // Всероссийский институт научной и технической информации (ВИНИТИ). М.: ВИНТИ, 2004, С. 38.
2. Воздушный шлагбаум // Известия. 2003, 2 августа. С. 4.
3. Сайт «Медицинская справочно-информационная система для пациентов», <<http://www.medinfo.ru/news/rusnews/1409.phtm>>.
4. Разработка метода для дистанционного контроля и оценки психического состояния персонала, работающего в стрессогенных условиях, основанного на контент-анализе их коммуникаций с внешними абонентами. Эл. адрес <http://www.congress.copris.com/congress2001/rus/doc8_r.htm>
5. Фролов Б. С., Овечкин В. В., Овечкина И. В. Экспресс-оценка и мониторинг функционального состояния организма и психического статуса человека по сердечному ритму // Вестник аритмологии. 2000. № 16.
6. Сайт «Элсис», <<http://www.elsys.ru>>.
7. Вассерман Е. Л., Карташев Н. К., Полонников Р.И. Фрактальная динамика электрической активности мозга. СПб.: Наука, 2004. 208 с.
8. Рудницкий С. Б. О проблеме автоматизированного дистанционного распознавания (классификации) психического состояния человека // X Санкт-Петербургская международная конференция «Региональная информатика - 2006» (РИ-2006), Санкт-Петербург, 24–26 октября 2006 г.: Материалы конференции. СПб.: СПОИСУ, 2006. С. 247–248.
9. Дюк В. А., Эммануэль В. Л. Информационные технологии в медико-биологических исследованиях. СПб.: Питер, 2003. 528 с.