

УДК 021.8 + 025.1
ББК 78.34

АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ ПСИХИЧЕСКОГО СОСТОЯНИЯ ОБУЧАЕМЫХ ПО ИСТОРИИ ИХ РАБОТЫ В СЛЕДЯЩЕЙ ИНТЕЛЛЕКТУАЛЬНОЙ ОБУЧАЮЩЕЙ СИСТЕМЕ

Смирнова Н. В.

*(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)*

В статье представлены результаты по построению модели психического состояния обучаемого по истории его работы в интеллектуальной обучающей системе. Данная модель может задействоваться алгоритмом управления учебным процессом во время решения задач обучаемым. Для построения модели использовались известные методы машинного обучения. Полученные результаты могут быть полезными не только для моделирования обучаемых, но и для моделирования пользователей любого программного обеспечения, в котором логируются действия пользователя.

Ключевые слова: автоматизированная обучающая система, интеллектуальная обучающая система, модель обучаемого, машинное обучение, модель пользователя

1. Введение

В последние годы задачи автоматизированного выявления психического состояния пользователя вызывают все больший интерес исследователей в области человеко-машинного взаимодействия. Наименее инвазивным для пользователя¹ источником получения данных для подобного анализа является история

¹ т.е. не приводящим к возникновению у пользователя чувства слежения за его действиями и не требующим наличия дополнительного оборудования

работы пользователя в прикладном программном обеспечении. Перечислим несколько основных применений мониторинга состояния пользователей ПК.

Актуальными являются такие применения мониторинга психического состояния пользователей, как оценка новых программных продуктов с точки зрения их удобства использования и интеллектуализация пользовательских интерфейсов программ. При выявлении негативного эмоционального состояния пользователя можно перестроить интерфейс программы на более простой, предложить ему посмотреть подходящий справочный материал или отложить действия, которые могут вызвать раздражение у пользователя (например, такие действия, как предложение обновить программу). При выявлении ситуаций, когда неадекватное поведение пользователя может нанести существенный ущерб, можно наложить временные ограничения на использование некоторых функций программы или даже запретить доступ к программе. Мониторинг психического состояния пользователя применим и для реализации здоровьесберегающих информационных технологий, позволяющих пользователю контролировать уровень стресса (см. например, технологию для мобильных устройств Android Remote Sensing App (AIRS) [14]).

Наконец, мониторинг состояния пользователей необходим для интеллектуального управления учебным процессом в обучающих системах, в частности, для управления процессом выполнения заданий обучаемыми. В качестве управляющих воздействий можно выделить следующие: рекомендация другого задания или временного прекращения работы в системе (отдыха); предоставление необходимого справочного материала. Модель психического состояния обучаемого также может задействоваться инструментами генерации отчетности для преподавателя, позволяющими выявить плохо проработанные места учебного курса, выделить студентов, которым следует уделить более пристальное внимание на экзамене или зачете, и т.д.

Результаты статьи получены в рамках моделирования обучаемых в следящих интеллектуальных обучающих системах², но могут быть полезны и в других, перечисленных выше случаях.

2. Постановка задачи

Для построения моделей, позволяющих выявить психическое состояние обучаемого по истории его работы, требуется прибегать к методам машинного обучения, в частности, решать задачи классификации [17], постановка которых выглядит следующим образом. Психическое состояние студента трактуется как один или несколько процессов P_i , $i=1, \dots, n$, состояния которых можно описать с помощью конечных множеств M_i значений некоторых нечетких переменных (например, «злость», «усталость», «радость», «обычное состояние»). Для построения модели собираются экспериментальные данные, включающие истории работы обучаемых и экспертные оценки динамики их психического состояния в ходе работы с обучающей системой. На основе экспериментальных данных для каждого процесса формируется конечная совокупность пар <состояние обучаемого, метка>. Состояние обучаемого задается посредством вектора значений некоторых показателей (в литературе эти показатели называют признаками [17]). Метка соответствует одному из значений M_i . Предполагается, что существует некоторая зависимость между метками и значениями признаков, но она неизвестна. На основе экспериментальных данных требуется восстановить эту зависимость, т.е. построить алгоритм, способный любому новому описанию психического состояния обучаемого (т.е. вектору значений признаков) достаточно точно присвоить метку.

Несмотря на достаточное количество работ, посвященных диагностике психического состояния обучаемого по истории его работы, эта проблема еще до конца не решена. Критическим

² *следящими интеллектуальными обучающими системами называются системы, обеспечивающие интерактивную поддержку процесса решения задач обучаемым и автоматизированную оценку его решений*

местом моделей такого рода является трудоемкость процесса сбора экспертных оценок. К примеру, для построения модели на основе историй работы хотя бы 30 студентов, необходимо затратить около 30 часов работы времени экспертов. В качестве экспертов, как правило, задействуются преподаватели, осуществляющие оценивание без отрыва от основной работы. Исходя из того, что 1) один эксперт имеет возможность отсматривать не более 3 видеозаписей в неделю длиной не более 1.5 ч³, 2) необходимо дополнительное время на процедуру предварительного согласования оценок экспертов, при участии 2 экспертов на сбор экспертных оценок потребуется не менее месяца.

Кроме того, сбор экспертных оценок осложняется тем, что в течение сеанса работы обучаемого происходит достаточно много событий, которые могут свидетельствовать об изменении его психического состояния. Возникает вопрос о том, в какие моменты времени эксперт должен выставлять оценки в ходе наблюдения за работой обучаемого, насколько точными должны быть эти оценки и т.д., т.е. необходимость создания правил, регулирующих экспертную деятельность. При разработке таких правил необходимо учитывать тот факт, что увеличение требований к эксперту ведет к возрастанию усталости эксперта и, как следствие, снижению надежности выставляемых им оценок, а понижение требований – к меньшей информативности и количеству экспертных оценок и, как следствие, к затруднениям при формировании совокупности пар <состояние студента, метка>. Поэтому правила, регулирующие экспертную деятельность, должны сопровождаться соответствующей методикой формирования совокупности пар <состояние обучаемого, метка> на основе экспериментальных данных.

Как будет показано далее, совершенствование таких правил и сопутствующих им методик формирования данных для построения модели является наиболее перспективным направлением улучшения существующих подходов к диагностике психического состояния обучаемых по истории их работы. В статье

³ оценка получена из опыта общения с экспертами в ходе исследований по данному вопросу

предлагаются правила сбора и обработки экспертных оценок, включающие:

- инструкцию для эксперта по выставлению оценок, определяющую моменты времени выставления оценок и другие предъявляемые к ним требования,
- методику формирования совокупности пар <состояние обучаемого, метка> на основе экспериментальных данных.

Далее в статье приводится более подробный обзор известных подходов и обосновывается выбор направления их улучшения (раздел 3), описываются применявшиеся в исследовании известные методы и технологии машинного обучения (п. 4.1), предлагаются правила сбора и обработки экспертных оценок (п. 4.2), описываются условия и результаты эксперимента, проведенного для апробации предлагаемых правил (п. 5.1-5.3). Также приводятся описания выделенных в результате проведения эксперимента признаков (п. 5.4), мало зависящих от интерфейса и возможностей обучающей системы, использование которых привело к повышению качества модели. Эти признаки могут быть полезны при построении подобных моделей в обучающих системах других типов, отличных от рассмотренного.

3. Известные результаты

Исследования различных аспектов поведения обучаемых по истории их работы в обучающей программе были начаты еще в 90-е годы прошлого века. Исследования проводились для обучающих систем различной направленности: следящих обучающих систем, содержащих курсы по предметам естественно-научного цикла(, [2]), для автоматизированных систем обучения программированию ([13],[19]), виртуальных лабораторий [7], систем для обучения иностранному языку [8], гипертекстовых обучающих сред с тестами по различным предметам [5] и т.д. (более подробный обзор см. в [21]).

К наиболее часто выделяемым исследователями компонентам психического состояния обучаемого относятся «Самостоятельность» ([1], [13], [15]) и «Усилия» ([19], [8], [15]). Ряд работ посвящен детектированию нерационального поведения студента

в обучающей среде (см. исследования в соавторстве с R.S.J.d Baker, например, [2]).

Исследователями были предложены различные списки формулировок «значимых» для оценки психического состояния обучаемых событий и способы вычисления значений признаков на их основе. Наиболее часто используются признаки, характеризующие взаимодействие обучаемого с подсистемой выдачи подсказок ([1], [2], [13], [11]) и подсистемой автоматической проверки фрагментов выполняемого задания ([1], [2], [13], [7], [8], [15], [11]). Методы построения моделей варьируются (например, в [13] компоненты психического состояния представляются в виде линейно изменяющихся величин с эмпирически заданными коэффициентами прироста, а в [7] и [11] – в виде байесовских сетей).

Как отмечалось выше, проблема диагностики психического состояния обучаемого по истории его работы еще до конца не решена. Направления улучшения существующих подходов включают:

- 1) совершенствование процедур подбора экспертов и согласования их оценок,
- 2) добавление новых признаков,
- 3) выбор наиболее подходящего алгоритма машинного обучения (например, логистическая регрессия, метод опорных векторов, байесовский классификатор и т.д.),
- 4) совершенствование правил сбора и обработки экспертных оценок (подробнее см. предыдущий раздел)

Обсудим перспективность работ по каждому из перечисленных направлений улучшения.

Процедуры отбора, проверки компетентности и формирования группы экспертов, а также генерации совместных экспертных оценок на основе индивидуальных достаточно хорошо исследованы в рамках области принятия решений (см., например, [20], [18]) и могут быть использованы и в рассматриваемом случае.

На формирование признаков влияют как возможности обучающей системы, так и удобство ее интерфейса. Действительно, при работе в простом гипертекстовом учебнике с тестами поведение студента менее вариативно, чем в следящей обучаю-

щей системе, в которой студент может вводить фрагменты своего решения, проверять их на правильность, брать подсказки и т.д. Недостатки интерфейса способствуют проявлению поведения, кажущегося хаотическим и нецеленаправленным, но таковым не являющимся. Например, студент, боясь потерять результаты своей работы, на всякий случай может держать открытыми одновременно несколько окон программы, и иногда, в поисках нужного окна, быстро осуществлять последовательность действий переключения между ними. Поскольку архитектура и возможности обучающих систем со временем изменяются, то выделение новых признаков, более точно характеризующих состояние студента в той или иной обучающей системе, имеет кратковременное скорее практическое, нежели научное значение. Некоторый интерес может представлять выделение признаков, применимых для ряда обучающих систем и слабо зависящих от особенностей интерфейса конкретных обучающих систем из этого ряда.

Как известно, в случаях, когда исследователю доступно большое количество данных, выбор того или иного алгоритма машинного обучения не играет большой роли [10]. В противном случае, рекомендуется сначала построить первоначальный вариант модели, используя простой алгоритм, а затем, в случае недостаточного качества модели, стараться улучшить качество модели с помощью различных способов, в том числе и путем подбора наиболее подходящего алгоритма машинного обучения. Поскольку на результаты работы любого алгоритма машинного обучения влияет не только количество данных, но и их степень зашумленности, которая может существенно варьироваться между различными экспериментами, представляется целесообразным рассматривать подбор подходящего алгоритма машинного обучения как один из этапов построения моделей для конкретных обучающих систем, а не как направление улучшения существующих подходов.

Как отмечалось ранее, проблема совершенствования правил сбора и обработки экспертных оценок для построения модели психического состояния студентов особенно актуальна: требуется уменьшить трудоемкость процесса сбора экспертных оценок, не снижая обоснованность и информативность результатов

экспертной деятельности. При этом под трудоемкостью понимается не только количество часов работы экспертов, но и степень когнитивной нагрузки на них.

Рассмотрим два известных варианта правил сбора и обработки экспертных оценок ([2], [3]). Первый вариант, согласно упоминаниям в статьях по рассматриваемому вопросу, соответствует наиболее часто используемым исследователями правилам. Второй вариант был предложен в качестве решения проблемы уменьшения трудоемкости сбора экспертных оценок.

В [2] эксперты наблюдали за работой студентов в режиме реального времени, и оценивали поведение студентов только в те моменты, когда, по их мнению, оно существенно изменялось. Для сопоставления векторов признаков, формируемых на основе истории работы студентов и оценок экспертов использовались «временные окна». Например, при 2-минутном «временном окне» для формирования пары <состояние обучаемого, метка> использовались действия, произошедшие не ранее чем за 1 минуту до t_0 и не позднее чем за 1 минуту после t_0 , где t_0 – время фиксации экспертной оценки. Данный подход имеет следующие недостатки. Затрудняется оценивание качества работы экспертов: возникают вопросы о том, достаточно ли оценок выставлено экспертами. От экспертов требуется внимательно следить не только за тем, в каком состоянии находится обучаемый, но и как можно более точно фиксировать точное время изменения этого состояния, что зачастую невозможно из-за недостаточной скорости работы эксперта и приводит к его быстрой утомляемости.

В [3] для повышения эффективности сбора экспертных оценок предлагается предоставлять экспертам для оценивания не видеозаписи работы студентов, а некоторое множество избранных фрагментов истории их работы в системе. Множество избранных фрагментов формируется почти случайным образом, при этом, для каждого фрагмента справедливо следующее: чем больше действий обучаемого в нем зафиксировано, тем больше вероятность того, что он будет выбран для экспертной оценки. Данный подход имеет следующие недостатки. Отдельные фрагменты истории работы обучаемого зачастую не содержат всю необходимую эксперту информацию: теряется информация о жестах, выражении лица обучаемого, его действиях вне обуча-

ющей системы, которые как будет показано далее, могут соответствовать как полезному, так и вредному с точки зрения учебного процесса психическому состоянию. Теряется контекст: эксперт не успевает получить общее впечатление об обучаемом. Между тем, при сравнении поведения обучаемого в текущий момент времени с поведением в предыдущие моменты, эксперт может рассматривать часть его действий как малозначительные, если они характерны для большинства периодов работы обучаемого или, наоборот, при появлении одного-двух, казалось бы, незначительных, но не характерных для обучаемого действий, сделать выводы об изменении психического состояния обучаемого.

Из вышесказанного следует, что проблема усовершенствования правил сбора и обработки экспертных оценок для построения модели психического состояния обучаемого является актуальным и наиболее перспективным направлением улучшения существующих подходов.

Далее в статье предлагаются правила, при использовании которых эксперты оценивают динамику состояний обучаемых на основе видеозаписей их работы через одинаковые промежутки времени. При этом в качестве оценки каждого компонента психического состояния обучаемых эксперты предоставляют только одно (максимум 3) наиболее частых, т.е. превалирующих значений этого компонента в течение рассматриваемого промежутка времени. Исследуется несколько возможных вариантов формирования совокупности пар <состояние обучаемого, метка> на основе экспериментальных данных, собранных с использованием предлагаемых правил.

На данном этапе исследований сравнение с другими подходами не представляется возможным, поскольку построение модели психического состояния обучаемого необходимо осуществлять на основе данных, собранных в разные дни с участием разных групп студентов, а процесс сбора экспертных оценок даже для одной группы обучаемых, работавших в системе в течение одного занятия, трудоемок. Насколько известно автору, единственными исследованиями по рассматриваемому вопросу, которые удовлетворяют этому требованию, является ряд исследований R.S.J.d Baker и др. ([2], [3]), в которых отмечается

неустойчивость полученной модели. Т.к. в данной статье описываются результаты разового эксперимента, проведенного с участием одной группы студентов, то сравнение с известными исследованиями не проводится. Вместо этого, исходя из оценки качества полученной модели, делаются выводы о применимости предлагаемых правил сбора и обработки экспертных оценок.

4. Предлагаемые правила сбора и обработки экспертных оценок

4.1. МОДЕЛЬ ПСИХИЧЕСКОГО СОСТОЯНИЯ СТУДЕНТА И ЕЕ ПОСТРОЕНИЕ

Прежде чем переходить к описанию предлагаемых правил сбора и обработки экспертных оценок, вкратце опишем выбранный алгоритм машинного обучения. Для построения модели был выбран известный метод логистической регрессии, т.к. этот метод сочетает простоту реализации и вычислительную эффективность.

Предполагается, что совокупность важных для управления учебным процессом в обучающей системе компонент психического состояния студента включает «Самостоятельность», «Усилия» и «Фрустрационное поведение». Психолого-педагогические основы построения модели психического состояния студентов разработаны при участии Шварц А.Ю.⁴

Текущее значение каждой компоненты психического состояния обучаемого определяется вектором вида

$$(1) \quad (f_{\theta'}(X), f_{\theta''}(X), f_{\theta'''}(X))$$

Координаты вектора вычисляются с помощью логистической функции с соответствующим набором регрессионных коэффициентов (θ' , θ'' или θ''') и выражают вероятности того, что уровень текущего значения компоненты подпадает под одну из категорий: «высокий», «средний», «низкий». Логистические функции в качестве аргумента используют переменную со следующей структурой:

⁴ н.с. психологического факультета МГУ

$$(2) \quad z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k.$$

В (2) $x_j, j = 1, \dots, k$ — признаки модели, которые могут принимать любые вещественные значения, а $\theta_j, j = 1, \dots, k$ — регрессионные коэффициенты.

Значения признаков формируются на основе сведений о произошедших «значимых» событий в истории работы обучаемого в течение соответствующего «временного окна». «Значимые» события могут быть одномоментными (например, проверка фрагмента решения на правильность) или длительными (например, чтение справочного материала). В качестве признаков в случае одномоментных событий используются сведения об их количестве, а для длительных — значения различных характеристик, например, средняя длительность, совокупная длительность и т.д. Также могут использоваться и другие признаки, не связанные со «значимыми» событиями, например, объективная сложность решаемой задачи.

Регрессионные коэффициенты подбираются следующим образом. Выборка (т.е. совокупность пар <состояние обучаемого, метка>) делится на две части — обучающую и контрольную. На обучающей выборке происходит настройка алгоритма (с использованием численных методов оптимизации решается задача минимизации функции ошибок, отражающей величину расхождений между метками алгоритма и метками эксперта), а на тестовой выборке — оценка качества алгоритма. В данном случае, т.к. каждая компонента психического состояния обучаемого измеряется посредством нечеткой переменной, имеющей 3 возможных значения, при построении модели для каждой компоненты психического состояния студента необходимо найти 3 набора регрессионных коэффициентов, и, соответственно, решить 3 задачи минимизации функции ошибок.

Модель считается хорошей, если значения F-мер⁵ на тестовой выборке достаточно высоки и существенно не отличаются от соответствующих значений на обучающей выборке (в про-

⁵ F-мера - гармоническое среднее точности и полноты классификации, принимает значения из $[0,1]$

тивном случае имеет место недообучение или переобучение, см. [16]).

В данном исследовании для подбора коэффициентов логистических функций используется программный пакет *scikit-learn* [6]. Для минимизации функции ошибок в программном пакете используется «Trust region newton method» – метод доверительных областей [4]. Основная идея метода заключается в том, что для нахождения минимума функции $f(x)$ можно осуществлять итеративно, при этом на каждой итерации f аппроксимируется посредством более простой функции q , которая достаточно правдоподобно отображает поведение f в окрестности N точки x . Достоинством метода является возможность использования в оптимизационных задачах большой размерности.

4.2. ПРЕДЛАГАЕМЫЕ ПРАВИЛА СБОРА И ОБРАБОТКИ ЭКСПЕРТНЫХ ОЦЕНОК

Для оценки поведения обучаемого эксперту предоставляется текстовое описание поведения и успехов студента во время обучения интерфейсу программы и видеозапись, на которой зафиксирован процесс решения задач. Видеозаписи предварительно обрабатываются таким образом, чтобы эксперту были одновременно доступны как запись с экрана обучаемого, так и запись его лица с веб-камеры. Каждая видеозапись заранее делится на 5-минутные фрагменты. После окончания каждого 5-минутного фрагмента видеозапись останавливается для того, чтобы эксперт внес оценки. Ниже приведена часть инструкции эксперта, регламентирующая выставление оценок после просмотра очередного 5-минутного фрагмента.

«Если в течение 5-минутного фрагмента

- значение компоненты психического состояния существенно не изменялось, укажите ее превалирующее значение. *Например: Н (т.е. значение преимущественно было низким).*
- существенно изменялось один раз, укажите ее превалирующее значение ближе к началу, середине и концу 5-минутного фрагмента. *Например: НСН (в*

начале значение параметра было низким, ближе к середине – средним, к концу – опять низким).

- существенно изменялось более одного раза, напишите МН. Этот фрагмент будет в дальнейшем рассмотрен с другим экспертом или исключен из рассмотрения при подборе коэффициентов модели.»

Перед тем, как переходить к оцениванию основного набора видеозаписей, эксперты согласовывают свои оценки на небольшом наборе видеозаписей. Видеозаписи для согласования оценок должны давать представление о том, как решают задачи в обучающей системе студенты из разных групп. Деление студентов на группы осуществляется с помощью показателя «% незасчитанных шагов», равного отношению количества незасчитанных системой шагов (т.е. фрагментов решения) к общему количеству проверенных шагов решения студента.

Соотнесение оценок эксперта вида $X_1X_2X_3$ с действиями обучаемого во время 5-минутного фрагмента для формирования совокупности пар <состояние студента, метка> может быть реализовано несколькими способами. Рассмотрим два наиболее очевидных способа.

Первый способ заключается в огрублении экспертных оценок путем преобразования оценок вида $X_1X_2X_3$ в оценки вида X_{1-3} с использованием некоторого набора эвристических правил (например, оценка «НСС» преобразуется в «С», а оценка «ВСВ» в «В» и т.д.). В этом случае на основе данных по каждому 5-минутному фрагменту формируется только одна пара <состояние обучаемого, метка>, причем в качестве метки используется оценка X_{1-3} . Для краткости будем называть этот способ формирования точек обучающей выборки «Огрублением».

Второй способ заключается в том, что 5-минутный фрагмент делится на 3 «временных окна» длительностью 100 с. В этом случае формируются 3 пары вида <состояние обучаемого, метка>, причем в качестве меток выступают оценки X_1 , X_2 , X_3 соответственно (этот способ будем называть «Равномерным разбиением»).

5. Апробация предлагаемых правил сбора и обработки экспертных оценок

5.1. УСЛОВИЯ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТА

В качестве участников эксперимента были задействовано 23 студента, изучавших курс линейной алгебры. Им было предложено пройти обучение интерфейсу программы, а затем попытаться решить задачи в системе. Студентам предлагалось решить одну задачу высокой сложности или одну задачу средней сложности и одну простую задачу. Во время работы студентов, помимо фиксации их деятельности в журнале действий системы, осуществлялась запись видео с экрана студента, а также видео с веб-камеры ноутбука студента. Для интерактивного оценивания видеозаписей использовался программный продукт ELAN v.4.4.0 ([9], [12]).

Перед экспериментом, на основе оценок студентов за домашние работы в течение семестра, множество участников было разбито на 3 группы: «сильные», «средние» и «слабые» студенты. Условия работы с участниками, входящими в каждую группу, варьировались: так, некоторым участникам предлагались предположительно слишком трудные или слишком легкие задачи, а также иногда предоставлялась ложная информация о трудности их задач. Получение инструкции того или иного типа зависело от того, к какой группе относится студент и каким по очереди он подошел к столу с инструкциями. Варьирование условий эксперимента осуществлялось с целью повышения разнообразия поведения и динамики проявления и смен психических состояний участников эксперимента.

Во время эксперимента в ИОС «Волга» был включен простой алгоритм управления учебным процессом, запускавшийся раз в 5 мин и предлагавший студентам различные подсказки в зависимости от их прогресса в задаче (подробнее см. [22]). Поскольку в случае успешного решения задач в системе студенты имели право получить предварительную оценку за экзамен, а оценка за решение задач снижалась за использование большого количества подсказок, то большинство студентов старались использовать подсказки только в самом крайнем случае. Поэтому показатель «Самостоятельность» был исключен из списка

показателей, для которых будет осуществлен подбор коэффициентов.

К оценке видеозаписей были привлечены 2 эксперта (преподаватели психологического факультета). Эксперты участвовали в процедуре согласования оценок два раза: перед началом оценивания видеозаписей и после того, как оценили половину видеозаписей (в последнем случае потребовалось повторное согласование оценок, т.к. в силу жизненных обстоятельств пришлось изменить первоначальную политику распределения видеозаписей по экспертам). С целью минимизации денежных и временных затрат большинство видеозаписей были оценены только одним из экспертов. Экспертами было оценено 215 5-минутных фрагментов. Как видно из рис. 4, в обеих задачах классификации имеет место проблема смещенных классов (англ. *skewed classes*). Например, для показателя «Усилия» преобладающими являются оценки В («высокий уровень»), а для показателя «Фрустрационное поведение» – оценки Н («низкий уровень»).

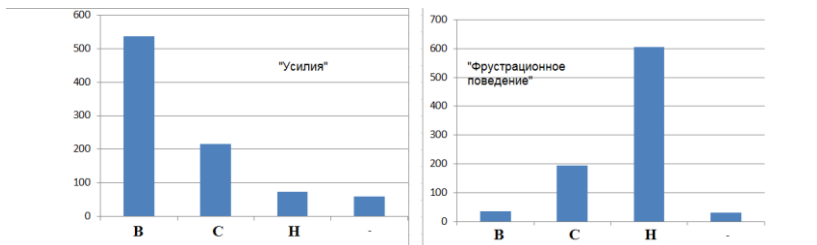


Рис. 1. Распределение количеств экспертных оценок, соответствующих высокому/среднему/низкому/неопределенному уровню компоненты «Усилия» и компоненты «Фрустрационное поведение».

5.2. ПОДБОР КОЭФФИЦИЕНТОВ ДЛЯ КОМПОНЕНТЫ «УСИЛИЯ»

Сначала было проведено сравнение двух упомянутых выше способов формирования совокупностей пар <состояние обучаемого, метка> на основе экспериментальных данных. Оценки качества модели, полученной каждым способом, вычислялись с

помощью процедуры стратифицированного 10-кратного скользящего контроля. Вкратце, суть данной процедуры заключается в следующем. Исходная выборка данных для обучения делится на обучающую и контрольную несколькими возможными способами, а оценка качества модели вычисляется как средняя оценка качества модели по всем разбиениям. Требование стратифицированности означает, что в обеих частях каждого разбиения исходной выборки количество элементов каждого класса (т.е. с одинаковой меткой) должно быть пропорционально количеству элементов этого класса во всей выборке данных для обучения.

Построение модели осуществлялось с помощью класса `sklearn.linear_model.LogisticRegression` пакета `scikit-learn` со следующими значениями параметров: `fit_intercept=True`, `penalty = 'l2'`, `C=100`, `class_weight = 'auto'`. Поясним роль этих параметров. Параметр `fit_intercept` указывает на то, должна ли присутствовать в величине z константа, т.е. слагаемое θ_0 . `Penalty` и `C` влияют на степень регуляризации (регуляризация приводит к уменьшению всех регрессионных коэффициентов и позволяет избежать переобучения модели). Параметр `class_weight` изменяет функцию ошибок, повышая относительную значимость правильного определения тех или иных классов по сравнению с другими классами. Использование процедуры *стратифицированного* k-кратного скользящего контроля и настройка параметра `class_weight='auto'` позволяют до некоторой степени решить проблему смещенных классов.

Результаты сравнения двух различных способов формирования совокупности пар <состояние обучаемого, метка> представлены в табл. 1.

Как видно из табл. 1, оба способа формирования данных для обучения показывают примерно одинаковые результаты. Для построения модели был выбран способ «Огрубление».

Далее был осуществлен анализ ошибок, проведена беседа с экспертами. Была сформирована вторая версия набора признаков логистических функций для компоненты «Усилия».

Табл. 1. Обобщающие способности различных алгоритмов формирования совокупности пар <состояние обучаемого, метка> на основе экспериментальных данных.

Способ формирования данных для обучения	F-мера на обучающей выборке			F-мера на тестирующей выборке		
	выс.	ср.	низ.	выс.	ср.	низ.
Уровни компоненты «Усилия»						
«Равномерное разбиение»	0.74	0.43	0.42	0.69	0.27	0.29
«Огрубление»	0.81	0.71	0.64	0.67	0.29	0.34

Далее было апробировано несколько эвристических правил преобразования оценок $X_1X_2X_3$ в оценку X_{1-3} . Наилучшие результаты дали следующие правила:

- 1) X_{1-3} равно той оценке, которой соответствует наибольшая частота в $X_1X_2X_3$. Если же такой оценки нет, то $X_{1-3} = H$ (F-мера на тестирующей выборке [0.75, 0.42, 0.42])
- 2) X_{1-3} равно В, если частота В в $X_1X_2X_3 \geq 2$, иначе X_{1-3} равно С (т.е. «не-В», низкий или средний уровень).

Для дальнейшей работы было выбрано правило 2). После отбрасывания незначимых признаков и поиска комбинации параметров алгоритма машинного обучения, обеспечивающих наиболее высокое качество модели, была выбрана комбинация параметров `fit_intercept=False`, `penalty='l1'`, `C=100`, `class_weight={0: 0.2, 1: 0.8}` со следующими F-мерами на обучающей и контрольной выборках: [0.75, 0.66], [0.75, 0.65].

5.3. ПОДБОР КОЭФФИЦИЕНТОВ ДЛЯ КОМПОНЕНТЫ «ФРУСТРАЦИОННОЕ ПОВЕДЕНИЕ»

Перед формированием набора признаков логистических функций была проведена беседа с экспертами, которые предположили, что в истории действий студента фрустрационное поведение может быть мало выражено. При оценке уровня

фрустрационного поведения эксперты в большой степени ориентировались на изменения выражения лица и невербальных жестов студентов. При этом «хаотическая деятельность» в системе⁶, которая предположительно сопутствует высокому уровню фрустрационного поведения, встречалась редко. Чаще всего фрустрационное поведение возникало после неуспешной проверки фрагмента решения и выражалось в том, что студент прекращал работу в системе и пытался получить помощь у соседей, или экспериментаторов (средний уровень фрустрационного поведения), или впадал в ступор (высокий уровень фрустрационного поведения). Также эксперты отметили, что фрустрационное поведение длилось недолго (не более чем 100 с) и сразу же прекращалось после того, как студенту помогали соседи или преподаватели.

С учетом результатов п. 4.2 формирование совокупности пар <состояние обучаемого, метка> осуществлялось методом «Огрубление». Наилучшие результаты дало следующее правило преобразования оценок $X_1X_2X_3$ в оценку X_{1-3} : если в оценке $X_1X_2X_3$ нет оценок В и/или С, то $X_{1-3} = Н$, иначе $X_{1-3} = С$. Таким образом, 5-минутные фрагменты разделились на 2 группы: те, в которых было зафиксировано повышение уровня фрустрационного поведения и те, в которых повышения не было. Процедура отбрасывания незначимых признаков не привела к повышению качества модели.

Наилучшей оказалась следующая комбинация параметров алгоритма `fit_intercept=False`, `penalty='l1'`, `C=200`, `class_weight={}` со следующими F-мерами на обучающей и контрольной выборках: [0.79, 0.69], [0.72, 0.61].

5.4. ВЫДЕЛЕННЫЕ ПРИЗНАКИ ДЛЯ ДИАГНОСТИКИ ПСИХИЧЕСКОГО СОСТОЯНИЯ СТУДЕНТА ПО ИСТОРИИ ЕГО РАБОТЫ

⁶ Такой деятельности могут соответствовать совокупности возникающих через короткие промежутки времени актов нецеленаправленного выбора и просмотра различных элементов интерфейса программы

Опишем способы измерения признаков, которые привели к повышению качества модели хотя бы одной компоненты психического состояния студента и удовлетворяют следующему условию: малая зависимость от особенностей интерфейса обучающих систем.

Выделим несколько групп возможных действий обучаемого в системе:

- *поисковые* (просмотр полученных подсказок, обозначений, введенного решения, поиск нужного параграфа в оглавлении и т.д., обозначается буквой «s»),
- *диалоговые* (проверка введенного фрагмента решения, запрос подсказки и т.д., обозначается буквой «d»),
- *редактирование фрагментов решения* (обозначается буквой «e»),
- *чтение теоретического материала* (обозначается буквой «t»),
- *бездействие* (т.е. отсутствие действий из вышеупомянутых групп, обозначается знаком «-»).

Предлагается разделить 5-минутный фрагмент на m интервалов (в данном случае $m=10$) и поставить в соответствие фрагменту вектор с обозначениями групп действий, превалирующих на соответствующих интервалах. Поскольку события «редактирование фрагментов решения» и «чтение теоретического материала» являются длительными событиями, то для сравнения с одномоментными событиями были получены их количества путем деления длительности события на некоторое число (10 и 30 с соответственно). Примером такого вектора является вектор «eestttt---», который расшифровывается следующим образом: в течение первой минуты студент преимущественно редактировал фрагмент решения, в течение следующих 30 с он преимущественно осуществлял поисковые действия, в течение следующих 2 мин студент преимущественно был занят чтением теоретического материала и т.д.

Для описания вектора превалирующих групп действий обучаемого p используются следующие признаки:

- количество встречающихся в нем обозначений «s»,
- количество обозначений «d»,
- количество обозначений «t» и «e» на позициях 0-2, 3-6, 7-9 (т.е. в начале, середине и конце вектора),
- количество обозначений «-» на позициях 0-2, 3-6, 7-9.

Также на основе данного вектора вычисляется значение признака «рваность (непоследовательность) действий обучаемого» по следующей формуле:

$$\frac{\sum_{i=1}^{m-1} \varphi(i)}{m-1}, \text{ где } \varphi(i) = \begin{cases} 1, & \text{если } p(i) \neq p(i-1) \\ 0 & \text{в противном случае} \end{cases}$$

Например, для вектора «eeeeeeeeee» значение этого признака будет равно 0, а для вектора «etetetetet» – 1.

Также оказались важными события «Бездействие дольше 7 сек». Эксперты отметили, что появление таких событий может свидетельствовать как о понижении «Усилий» и/или повышении «Фрустрационного поведения» (студент слишком часто отвлекается на посторонние звуки, на диалоги с соседями или впадает в ступор после неуспешной проверки шага решения), так и о повышении «Усилий» и/или понижении «Фрустрационного поведения» (студент что-то вычисляет с помощью листа бумаги и ручки перед вводом следующего шага).

Поэтому было решено отдельно учитывать «хорошие» и «плохие» периоды бездействия: хорошим предшествует успешная проверка фрагмента решения, а «плохим» – неуспешная. Для обеих компонент психического состояния обучаемого оказались полезными признаки, отражающие количество «хороших» и «плохих» периодов бездействия длительностью [7,25) с, [25, 50) с, > 50 с.

6. Выводы

В статье изложен опыт решения задач классификации, возникающих при построении модели психического состояния обучаемого по истории его работы. Апробирован оригинальный протокол сбора и обработки экспертных оценок, при использовании которого эксперты каждой компоненте психического

состояния обучаемого должны давать оценки вида $X_1X_2X_3$ каждые n секунд (в статье $n=300$). Показано, что в этом случае перспективным способом формирования совокупности пар <состояние обучаемого, метка> является способ «Огрубление», когда на основе каждого 5-минутного фрагмента формируется одна пара с оценкой X_{1-3} , полученной в результате преобразования оценок $X_1X_2X_3$ с помощью некоторого набора эвристических правил.

В ходе решения вышеописанных задач классификации предложено несколько признаков, которые могут быть полезными для любых автоматизированных обучающих сред.

Предложенный протокол сбора и обработки экспертных оценок может быть использован для построения модели психического состояния обучаемого по истории его работы. При его использовании будут классифицированы 2 градации уровня показателей «Усилия» («высокий уровень» и «средний или низкий уровень») и 2 градации уровня показателя «Фрустрационное поведение» («зафиксировано повышение уровня» и «повышения уровня не было»).

Для повышения качества построенной модели требуется больше данных. Поскольку существуют временные и денежные ограничения на затрачиваемое экспертами время, рекомендуется включать в выборку больше данных от тех студентов, у которых показатель «% незасчитанных шагов» (см. п. 4.1) имеет более высокое значение. Такие студенты демонстрируют более вариативное поведение. Также, поскольку в большинстве случаев при оценке каждой компоненты психического состояния студента в течение 5-минутного фрагмента эксперты давали более одной оценки, то может оказаться полезной следующая модификация протокола: вместо 5-минутных фрагментов просить экспертов давать оценки показателям МВК каждые 100 с.

Литература

1. Aleven V. Toward Meta-cognitive Tutoring: A Model of Help-Seeking with a Cognitive Tutor // International Journal of Artificial Intelligence in Education. 2006. – № 16. – С. 101–130.

2. Baker R.S.J. d. (et al.) Developing a Generalizable Detector of When Students Game the System // User Modeling and User-Adapted Interaction. 2008. – Vol. 18. № 3. – P. 287–314.
3. Baker R.S.J.d, Carvalho A.M.J.B. de. Labeling Student Behavior Faster and More Precisely with Text Replays // Educational Data Mining. 2008. – P. 38–47.
4. Chih-Jen L., Weng R.C., Keerthi S.S. Trust Region Newton Method for Large-Scale Logistic Regression // Journal of Machine Learning Research. 2008. – № 9. – P. 627–650.
5. Cocea M., Weibelzahl S. Disengagement detection in online learning: validation studies and perspectives // IEEE Transactions on Learning Technologies. 2011. – Vol. 4. – № 2. – P. 114–124.
6. Pedregosa F. (et al.) Scikit-learn: Machine learning in Python / , F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot E.D. // Journal of Machine Learning Research. 2011. – Vol. 12. – P. 2825–2830.
7. Hernandez Y. (et al.) A Probabilistic Model of Affective Behavior for Intelligent Tutoring Systems / Y. Hernandez, J. Noguez, E. Sucar, G. Arroyo-Figueroa // MICAI 2005. LNAI 3789 / eds. H.T. A. Gelbukh, A. de Albornoz. Berlin: Springer-Verlag Berlin Heidelberg, 2005. – P. 1175–1184.
8. HersHKovitz A., Nachmias R. Developing a Log-based Motivation Measuring Tool // Interdisciplinary Journal of E-Learning and Learning objects. – Vol. 5. – P. 197–214.
9. Lausberg H., Sloetjes H. Coding gestural behavior with the NEUROGES-ELAN system // Behavior Research Methods, Instruments & Computers. – Vol. 41. № 3. – P. 841–849.

10. Manning C.D. (et al.) Introduction to Information Retrieval / C.D. Manning, P. Raghavan, P. Schütze // Cambridge University Press, 2008.
11. Mavrikis M. Data-driven Modelling of Students' Interactions in an ILE. // Proceedings of the 1st International Conference on Educational Data Mining. Montreal, Canada, 2008.
12. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen T.N. ELAN. A professional tool for the creation of complex annotations on video and audio resources [Электронный ресурс]. URL: <http://tla.mpi.nl/tools/tla-tools/elan/> (дата обращения: 31.05.2013).
13. Soldato T., Boulay B. de. Formalization and Implementation of Motivational Tactics in Tutoring Systems // Journal of Artificial Intelligence in Education. 1995. – Т. 6. № 4. – С. 337–378.
14. Trossen D., Pavel D. AIRS: A Mobile Sensing Platform for Lifestyle Management Research and Applications // Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. – Vol. 65, 2013. – P. 1–15.
15. Yusoff M., Boulay B. de. Integrating Domain-Independent Strategies into an Emotionally Sound Affective Framework for an Intelligent Learning Environment [Электронный ресурс]. URL: <http://www.cogs.susx.ac.uk/users/bend/papers/aisb2005.pdf> (дата обращения: 09.07.2011).
16. Воронцов К.В. Комбинаторный подход к оценке качества обучающих алгоритмов // Математические вопросы кибернетики. Т. 13 / под ред. О.Б. Лупанов. Москва: Физматлит, 2004. – С. 5–36.
17. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). Курс лекций ВМК МГУ и МФТИ [Электронный ресурс]. URL:

<http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf>
(дата обращения: 18.12.2013).

18. Литвак Б.Г. Экспертные оценки и принятие решений. Москва: Патент, 1996. – 271 С.

19. Никифоров А.Ю., Русаков В.А. Модель студента в программном комплексе поддержки лабораторных занятий // Научная сессия МИФИ-2005. Сборник научных трудов. – Т. 2. – С. 79–80.

20. Орлов А.И. Экспертные оценки. Учебное пособие. Москва, 2002.

21. Смирнова Н.В., Шварц А.Ю. Мотивационно-волевой компонент модели обучаемого в следящих интеллектуальных системах. Часть 1 // Искусственный интеллект и принятие решений. 2012. – Т. 1. – № 1. – С. 65–80.

22. Смирнова Н.В., Шварц А.Ю. К вопросу об измерении и параметрической идентификации мотивационно-волевого компонента модели обучаемого // Информационно-технологическое обеспечение образовательного процесса государств–участников СНГ [Электронный ресурс] : сборник докладов Международной интернет-конференции. Минск: БГУ, 2012. URL: <http://elib.bsu.by/handle/123456789/27772> (дата обращения: 18.12.2013)

AUTOMATED ANALYSIS OF STUDENT'S PSYCHOLOGICAL STATE BASED ON LOGS OF HIS/HER ACTIVITY IN A MODEL-TRACING INTELLIGENT TUTORING SYSTEM

Natalia Smirnova, Institute of Control Sciences of RAS, Moscow, researcher (smirnovanatalia2008@gmail.com).

Abstract: This paper presents results of a research on student's psychological state model which uses logs of student's activity in the tutoring system. The suggested model is for use by the tutoring process control algorithms during students' problem solving sessions. The suggested model was built using machine learning techniques. The results of the paper can be used not only for student modelling but also for user modelling in any program that is able to log the user activity.

Keywords: intelligent tutoring system, student model, machine learning, user model

Статья представлена к публикации
членом редакционной коллегии ...заполняется редактором...

Поступила в редакцию ...заполняется редактором...
Опубликована ...заполняется редактором...