

УДК 519.95:621.3+62.50

С.И. Кондрашов, Т.В. Дроздова

Национальный технический университет «Харьковский политехнический институт», Харьков

ИССЛЕДОВАНИЕ БАЙЕСОВСКОГО КЛАССИФИКАТОРА ПРИ ОЦЕНИВАНИИ КАЧЕСТВА ВУЗА В УСЛОВИЯХ АПРИОРНОЙ НЕОПРЕДЕЛЕННОСТИ

В статье рассматривается модель образовательного процесса как марковского в некотором приближении с дискретными состояниями и непрерывным временем. Все переходы системы из состояния в состояние, происходящие под действием некоторых потоков событий, представлены в виде размеченного графа состояний, который представляет собой «схему гибели и размножения». Рассматривается вопрос используемой терминологии теории распознавания, анализируются особенности проектирования байесовских классификаторов в условиях параметрической априорной неопределенности. Рассматриваются вопросы практического применения объекта исследования в задачах диагностики качества образовательного процесса при существенно ограниченных объемах обучающей выборки.

Ключевые слова: качество образовательного процесса, процесс «гибели и размножения», априорная неопределенность, ограниченная выборка, байесовский классификатор, объем обучающей выборки, вероятность ошибки классификатора.

Постановка задачи

На современном этапе важным направлением приложения усилий в сфере образования является борьба за повышение качества образовательных услуг. В этом контексте заслуживает внимания метод использования системы критериальных компетентностно-ориентированных задач для диагностики уровня сформированности компетенций, а также выработка управляющего воздействия для их коррекции [1].

Одними из особенностей функционирования вуза являются динамический характер и ограниченное количество людей, которые осуществляют образовательный процесс.

Образовательный процесс в вузе отличается большой протяженностью и ограниченным количеством предметов, которые определяют необходимый уровень знаний студента и являются критерием его обученности.

Следовательно, компетенция студента/выпускника во многом зависит от количества пройденных обучающих этапов (модулей, семестров, курсов), предметов (итогового балла студента по каждому из предметов) и др.

Таким образом, возникает задача оценивания качества обученности студентов с учетом динамики и ограниченности выборки критериев оценивания.

Анализ достижений и публикаций

В работах К.В.Тараканова, Л.А. Овчарова, А.Н. Тырышкина, Е. В. Вентцель систематически изложены методы математического описания динамических процессов функционирования систем, которым можно отнести систему образования, изложен веро-

ятностный подход к функционированию сложных динамических систем.

Работы Ш. Раудиса [2], П.Ф. Шапова [3] посвящены исследованию теории распознавания образов, а также проблем построения байесовского классификатора в условиях неопределенности с использованием теории математической статистики.

Байесовский классификатор многомерных наблюдений

Рассматривая систему образования как марковский процесс с дискретными состояниями и непрерывным временем, удобно представить все переходы системы S из состояния в состояние, происходящие под действием некоторых потоков событий, в виде размеченного графа состояний. Такой граф представляет собой так называемую «схему гибели и размножения».

При временном дрейфе динамику обученности можно представить точкой, перемещающейся по определенному закону в поле допуска и положение которой характеризуется интенсивностью усвоения знаний и интенсивностью забывания. Возможен контроль лишь вполне различимых положений этой точки при ее перемещении. Различимые точки в поле допуска представляются как дискретные состояния контролируемого параметра. Следовательно, можно предположить, что существует некоторый поток событий, вызывающий переход подсистемы из состояния в сторону границы поля допуска.

На рис. 1 изображена возможная реализация процесса гибели и размножения получения профессиональных знаний студентом на протяжении всего времени обучения, начальное состояние которого равно нулю [$S(0)=S_0=0$].

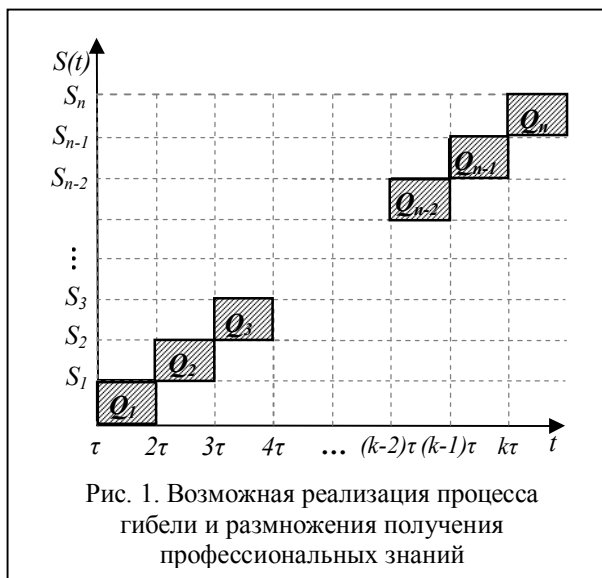


Рис. 1. Возможная реализация процесса гибели и размножения получения профессиональных знаний

После каждого шага h студент переходит в одно из k возможных состояний ($0 \leq S_k \leq Q_n$), которое характеризует то количество знаний, умений и навыков, которым обладает обучающийся в данный момент времени. При состоянии обученности $S_k=Q_n$ процесс обучения можно завершить, следовательно, процесс гибели и размножения будем считать с конечным числом состояний [1].

Процесс обучения (совокупность процессов усвоения и забывания профессиональных знаний) состоит из совокупности точек, характеризующих уровень обученности в каждый момент времени t , последовательность которых подчиняется определенному закону в поле допуска на протяжении некоторого периода $k\tau < t < (k+1)\tau$.

Каждую реализацию положения уровня обученности студента можно представить как совокупность точек в n -мерном пространстве из заштрихованной области соответствующего этапа на рисунке 1. Совокупность таких точек образует определенные группы – классы. Совместное распределение множества точек, принадлежащих к одному классу, называется функцией правдоподобия выборки и соответствует условной плотности распределения $W(X/Q_k)$.

Разработка оптимальных устройств связана с поиском таких параметров, которые обеспечивают экстремум выбранного критерия качества. Следовательно, необходимо создание такого устройства, которое по измерениям ряда параметров исследуемой системы будет классифицировать их по группам, – классификатора (распознающего устройства) [4].

Задачу классификации, лежащую в основе технических и кваліметрических приложений, можно сформулировать следующим образом: при заданных вероятностях $P(Q_k)$, распределениях $W(X/Q_k)$, наборе решений $\{\gamma_k\}$, где $k=1..n$, матрице потерь, требу-

ется определить наилучшее, в смысле выбранного критерия качества, правило использования результатов наблюдения для выбора решения. Набор решений $\{\gamma_k\}$, где $k=1..n$, представляет ряд логических утверждений о том, какая из гипотез относительно классов $Q_1..Q_n$ истинна. Оптимальным правилом выбора решения будет такое, для некоторого при заданных априорных вероятностях классов, матрице потерь и условных функциях правдоподобия выборок, средний риск R_{cp} будет наименьшим среди величин рисков, относящихся к любым другим правилам выбора решения:

$$R_{cp} = \min R(\gamma). \quad (1)$$

Такой классификатор называется байесовским и соответствует оптимальному качеству классификации [3].

Создание оптимальных при минимальной вероятности ошибок классификаторов наталкивается на значительные трудности в условиях априорной неопределенности, когда объем обучающей выборки ограничен. В связи с этим требует развития направление, содержанием которого является разработка методов преодоления априорной неопределенности и получение классификаторов, близких к оптимальным. Оптимальность достигается применением критерия максимального правдоподобия. При построении байесовского классификатора обязательным является знание, по крайней мере, условных функций правдоподобия.

Особенности проектирования байесовского классификатора при существенно ограниченных выборках

Проектирование распознающего устройства в условиях априорной неопределенности, вызванной малым объемом обучающей выборки (ООВ), наталкивается на ряд трудностей. Во-первых, необходимо решить вопрос о представительности выборки, то есть определить, достаточна ли она для обучения классификатора. Во-вторых, необходимо построить систему признаков, отобрав наиболее информативные (обладающие степенью разделяющей способности признака, количественной характеристикой которой могут служить вероятность правильной классификации или вероятность ошибки классификатора, функционально связанная с последней).

Оценка достаточности объема обучающей выборки зависит от выбранного критерия качества. За последние годы предложено много критериев, из которых не все характеризуют качество полученного правила решения. Сравнительный анализ критериев определения качества обучения показывает, что наиболее предпочтительным является критерий вероятности отказа классификатора (ВОК) или критерий качества с ним связанные[4].

Как показывают исследования достаточности объема обучающей выборки, когда используется критерий ВОК, представительность объема обучающей выборки связана с выбором правила классификации, размерностью пространства признаков и зависит от выбранной модели распределения признаков. Чувствительность классификатора к ООВ возрастает с увеличением сложности правила классификации.

При малых ООВ статистические характеристики классов, влияющих на точность оценки достаточности ООВ, однозначно определить невозможно.

В работах [2, 3] указано, что качество обучения распознающего устройства при заданном ООВ линейно зависит как от количества признаков, так и от информативности последних. Если в условиях полной априорной информации добавление любого малоинформативного признака улучшает качество распознавания, то в условиях априорной неопределенности все значительно сложнее.

Выводы

Рассмотрение понятия качества образования с точки зрения элемента теории массового обслуживания позволяет применить некоторые элементы данной теории для оптимизации методов оценивания качества. Основные идеи и методы математической теории распознавания и классификации приобретают все большее признание в качестве фактора, составляющего основу построения современных информационно-измерительных систем, в том числе систем оценивания качества вуза. Основными особенностями проектирования байесовского классификатора в условиях неопределенности являются

достаточность объема выборки критериев и информативность последних.

В качестве такового подходит критерий вероятности отказа классификатора, который лучше всего описывается с помощью нечетко-вероятностного подхода.

Список литературы

1. Кондрашов С.И. Нечетко-вероятностная модель восстановления квалитетических характеристик образовательного процесса / С.И. Кондрашов, Т.В. Дроздова // «Метрологія та прилади» науково-виробничий журнал. Тематичний випуск. №1 II (45) 2014. – Харків, 2014. – С. 120-123.
2. Раудис Ш. Ограниченность выборки в задачах классификации / Ш. Раудис // Материали к семинару при Институте физики и математики АН Литовской ССР «Статистические проблемы управления». Выпуск 18. – Вильнюс, 1976. – 185 с.
3. Щапов П.Ф. Исследование и разработка байесовского классификатора в условиях априорной неопределенности: диссертация на соискание ученой степени кандидата технических наук : 05.13.01 / Щапов Павел Федорович. – Х.: 1980. – 183 с.
4. Кондрашов С.И. Особенности построения системы оценки качества в терминах нечеткой логики при ограниченных выборках / С.И. Кондрашов, Т.В. Дроздова // XXIII Национален научен симпозиум с международно участие «Метрология и метрологично осигуряване 2013», 9-13 септември 2013 г. – Созопол, България, 2013. – С. 379-383.

Поступила в редколлегию 17.03.2014

Рецензент: канд. техн. наук, проф. В.К. Гусельников, Национальный технический университет «Харьковский политехнический институт», Харьков.

ПРОБЛЕМИ ПОБУДОВИ БАЙЕСОВСЬКОГО КЛАСИФІКАТОРА В УМОВАХ АПРІОРНОЇ НЕВИЗНАЧЕНОСТІ

С.І. Кондрашов, Т.В. Дроздова

У статті розглядається модель освітнього процесу як марковського деяким наближенні з дискретними станами і безперервним часом. Усі переходи системи зі стану в стан, що відбуваються під дією деяких потоків подій, представлені у вигляді розміченого графа станів, який являє собою "схему загибелі і розмноження". Розглядається питання використання термінології теорії розпізнавання, аналізуються особливості проектування байесовських класифікаторів в умовах параметричної априорної невизначеності. Розглядаються питання практичного застосування об'єкта дослідження в завданнях діагностики якості освітнього процесу при істотно обмежених обсягах навчальної вибірки.

Ключевые слова: *якість освітнього процесу, процес "загибелі і розмноження", априорна невизначеність, обмежена вибірка, байесовський класифікатор, обсяг навчальної вибірки, імовірність помилки класифікатора.*

THE PROBLEM OF CONSTRUCTING BAYESIAN CLASSIFIER UNDER APRIORI UNCERTAINTY

S.I. Kondrashov, T.V. Drozdova

In the article the model of the educational process as a Markov some approximation with discrete states and continuous time is considered. All transitions from state to state, occurring under the influence of some streams of events, presented in the form of a mark-up of the state graph. Such graph is a "scheme of birth and death". The question of the terminology used, analyzes the characteristics of the design of Bayesian classifiers under parametric a priori uncertainty is considered. The practical application of the research object in the diagnostics of the quality of the educational process at substantially limited amounts of training sample is considered.

Keywords: *quality of the educational process, process of «birth and death», priori uncertainty, limited sample, Bayesian classifier, volume of training sample, probability of error of classifier.*