

Моделі Параметричної Статистики в Системах Дистанційного Тестування Знань

Володимир Ляшик
кафедра Програмної Інженерії
Харківський національний університет
радіоелектроніки
Харків, Україна
volodymyr.liashyk@nure.ua

Ігор Шубін
професор
кафедра Програмної Інженерії
Харківський національний університет
радіоелектроніки
Харків, Україна
igor.shubin@nure.ua

Models of Parametric Statistics in Distance Testing Systems

Volodymyr Liashyk
Department of Software Engineering
Kharkiv National University
of Radio Electronics
Kharkiv, Ukraine
volodymyr.liashyk@nure.ua

Igor Shubin
PhD, professor
Department of Software Engineering Kharkiv National
University of Radio Electronics
Kharkiv, Ukraine
igor.shubin@nure.ua

Анотація—показано перевагу запропонованого методу побудови ПЗ адаптивного тестування на базі мереж Маркова з елементами мереж Петрі при використанні концепції основних і додаткових питань. Різниця в рівнях складності основних і додаткових питань і здійснюваний зв'язок між основними і гілками додаткових питань дозволяє в процесі тестування як мінімізувати кількість необхідних відповідей, так і суттєво поліпшити адаптаційні властивості тестування.

Abstract—The advantage of the proposed method of constructing adaptive testing software based on Markov networks with elements of Petri nets using the concept of basic and additional questions is shown. The difference in the levels of complexity of the main and additional questions and the connection between the main and branches of additional questions allows in the testing process to minimize the number of required answers of those taught to determine their level of knowledge and significantly improve the adaptive properties of testing.

Ключові слова—адаптивне тестування, мережі маркова, модель раша, параметрична статистика, апроксимація функцій, критеріальні оцінки

Keywords— adaptive testing, markov nets, rash model, parametric statistics, stochastic methods, approximation functions, criterial estimates

І. ВСТУП

Контроль знань або тестування – це процес, проведений з метою визначення рівня знань [1]. Це найбільш стандартизований і об'єктивний метод контролю й оцінювання знань, умінь і навичок випробуваного, який позбавлений таких традиційних недоліків інших методів контролю знань, як неоднорідність вимог, суб'єктивність екзаменаторів, невизначеність системи оцінок і т.п. Рівні знань зазвичай дискретизуються.

При такому підході, тестування може розглядатися як деякий діагностичний процес, а стани, що характеризують оцінки знань випробуваного, — як діагностичні стани. Тести є ефективним засобом перевірки якості знань, одержуваних студентами, і оперативного контролю ходу навчання [2]. Інформаційні освітні ресурси (ІОР) тестові матеріали, що містять, можна розбити на дві категорії: орієнтовані на проходження студентами тестів у письмовій формі з подальшою перевіркою вручну викладачем (як варіант – сканування результатів тестування з метою їх подальшої автоматизованої перевірки) та системи комп'ютерного тестування (СКС) з відповідним наповненням тестовими матеріалами. Переваги другої категорії тестових ІОР вони дозволяють звільнити викладача від рутинної роботи при проведенні іспитів і



проміжній оцінці знань у традиційному навчальному процесі, а при навчанні з використанням дистанційних технологій стають основним засобом контролю; надають можливість автоматизації обробки результатів, об'єктивність контролю й швидкість перевірки якості підготовки великого числа тестуємих по широкім колу питань. Це дозволяє визначити розділи, які представляють найбільшу складність у вивченні, і, можливо, коректувати процес навчання залежно від результатів тестування; надають можливість реалізації навчальної функції; дозволяють зробити індивідуалізацію процесу засвоєння знань учнями.

II. СИСТЕМИ АДАПТИВНОГО КЕРУВАННЯ

Функції контролю знань носять не тільки контролюючий характер, але також навчально-виховний і розвиваючий. Форми тестових завдань можна представити в наступних формах: закрита форма; відкрита форма; на відповідність; на встановлення правильної послідовності.

Проведений аналіз існуючих КСТ показав, що вони в більшості випадків орієнтовані на проведення тестів, а не на їхню розробку. При реалізації тестування жодна з розглянутих КСТ не підтримує адаптивні методи проведення тестів, слабо розвинена політонічна оцінка [3] виконання тестових завдань.

При наявності великої кількості систем комп'ютерного тестування, постійним створенні нових систем необхідно визначитися з основними критеріями, яким повинні відповідати такі системи. Попередній аналіз показує, що весь набір параметрів може бути зведений у наступні чотири основні групи: типи підтримуваних тестових завдань; сервіс по створенню тестових завдань; побудова траєкторії тестування; обробка результатів тестування; адміністративні функції.

Для порівняння були обрані відомі системи Прометей 4.0, АСТ, Ellekta 5.01, Test Commander 3.0, Webquiz XP 2.0.34. Як показує аналіз, основні параметри провідних систем багато в чому схожі. Загальним недоліком усіх систем (у напрямку його подолання багато хто з них, імовірно, будуть розвиватися) є відсутність або обмеженість можливостей по обміну інформацією з міжнародним стандартом IMS QTI, а також відсутність засобів адаптивного тестування [4].

Під адаптивним тестовим контролем розуміють комп'ютеризовану систему науково обгрунтованої перевірки й оцінки результатів навчання, що має високу ефективність за рахунок оптимізації процедур генерації, пред'явлення й оцінки результатів виконання адаптивних тестів.

Ефективність контрольно-оцінних процедур підвищується при використанні багатокрокової стратегії відбору й пред'явлення завдань, заснованої на алгоритмах з повною контекстною залежністю, у яких черговий крок відбувається тільки після оцінки результатів виконання попереднього кроку (або кроків). Після виконання випробуванням чергового завдання щораз виникає потреба в ухваленні рішення про вибір рівня складності

наступного завдання залежно від того, вірною або невірною була попередня відповідь [5].

Алгоритм підбору й пред'явлення завдань будується за принципом зворотного зв'язку, коли при правильній відповіді випробуваного чергове завдання вибирається більш важким, а невірна відповідь спричиняє пред'явлення наступного більш легкого завдання, ніж те, на яке випробуванним була дана невірна відповідь. Також є можливість завдання додаткових питань по темах, які той, якого навчають, знає не дуже добре для більш тонкого з'ясування рівня знань у даних областях. Таким чином, можна сказати, що адаптивна модель нагадує викладача на іспиті – якщо той, якого навчають, відповідає на, що задаються питання впевнено й правильно, викладач досить швидко ставить йому позитивну оцінку. Якщо той, якого навчають, починає «плавати», то викладач задає йому додаткові або навідні запитання того ж рівня складності або по тій же темі. І, нарешті, якщо той, якого навчають, із самого початку відповідає погано, оцінку викладач теж ставить досить швидко, але негативну [6].

До неодмінних умов реалізації таких алгоритмів слід віднести:

- наявність банку каліброваних завдань зі стійкими оцінками їх параметрів, що дозволяють прогнозувати успіх або неуспіх випробуваного при доборі чергового завдання адаптивного тесту;

- використання інструментальних-програмно-інструментальних засобів і комп'ютерних програм для індивідуалізації алгоритмів добору завдань, заснованих, як правило, на оцінці ймовірності правильного виконання навчальних завдань;

- використання параметричних моделей Item Response Theory [7] (психометрична теорія завдання і відповідь).

Багатокрокові стратегії адаптивного тестування можна підрозділити на фіксоване, варіюють від того, як конструюються багатокрокові адаптивні тести. Співставлення різних видів адаптивного тестування призводить до вибору комп'ютерного адаптивного тестування, заснованого на багатокрокових стратегіях, що варіюють, у якості основного підходу. Найбільш важлива перевага стратегій, що варіюють, пов'язане з можливістю оперативного реагування на результати виконання навчальних завдань шляхом переоцінки рівня підготовки того, якого навчають, після виконання кожного чергового завдання адаптивного тесту.

Тестування зазвичай починається із завдань середньої складності, але можна починати й з легких завдань, тобто йти за принципом підвищення складності. Тестування закінчується, коли параметри функції успіху (по моделі Раша [8]) є статистично певними й дозволяють оцінити по тій або іншій шкалі рівень підготовки випробуваного.

Особливості адаптивного тестування: дозволяють більш гнучко й точно вимірювати знання суб'єктів навчання; дозволяє вимірювати знання меншою кількістю завдань, ніж в класичній моделі; виявляє теми, які суб'єкт



навчання знає погано й дозволяє задати по них низку додаткових питань; заздалегідь невідомо, скільки питань необхідно задати суб'єкту навчання, щоб визначити його рівень знань. Якщо кількість питань, закладених у систему тестування, виявляється недостатньою, можна перервати тестування й оцінювати результат по тій кількості питань, на яку відповів суб'єкт навчання; можливе застосування тільки на ПК.

Надійність результатів тестування в цьому випадку досить висока, тому що здійснюється пристосування під рівень знань конкретного суб'єкта навчання, що забезпечує більш високу точність вимірів.

III. МОДЕЛЮВАННЯ АДАПТИВНОГО ТЕСТУВАННЯ

Для визначення базового алгоритму, необхідно привести сценарій роботи системи. У його основі лежить модель приймання іспиту викладачем у студента, як модель адаптивного тестування. Такий вибір сценарію роботи системи обумовлений тим, що, по-перше, дана процедура історично зложилася дуже давно й добре формалізована, по-друге, — при проектуванні тестів, їх розроблювачеві необхідно опиратися на загальноприйняті, відомі й використовувані їм методи з мінімальною модифікацією. Далі наведений сценарій приймання іспиту.

Сам алгоритм адаптивного тестування повинен мати на кожному кроці (при переході від одного завдання до іншого) максимальною інформативністю (тобто надавати максимум інформації про здатності студента на одне задане питання). У той же час, не можна повністю відмовлятися від обліку суб'єктивних властивостей студента, які можуть виражатися в нерозумінні очевидно поставленого питання або завдання.

Додаткові питання є невід'ємною частиною тестів за певних умов, однак потрібно врахувати: не всі додаткові питання є тотожні як з погляду складності, так і з погляду повноти відповідності основному питанню; ланцюжка додаткових питань є логічно зв'язаними послідовностями; питання (як основні, так і додаткові) задаються послідовно, тобто неможливо задати два або більш питання одночасно; частота завдання додаткових питань різна.

За базову, але не єдину модель системи адаптивного тестування, обрана вистава у вигляді Марківських мереж, яке має наступні властивості: основною одиницею є питання певної складності, у якого можуть бути ланцюга додаткових питань; вибір додаткового питання визначається виходячи з імовірності появи кожного додаткового питання як потоку найпростіших подій з експонентним розподілом або шляхом логічного висновку на вибір невірної відповіді; система тестів є замкненою в розумінні Марківських мереж, тобто якщо питання (основний або додатковий) являє собою стан s_i , що має ймовірність перебування в ній системи $P(s_i)$, то сумарна ймовірність перебування системи в S тотожно визначається як:

$$P_{\Sigma} = \sum P(s_i) = 1;$$

зазначений стан — це тестове завдання як процес оцінки знань і вмінь випробуваного, повинне виконуватися у випадку вдалої відповіді на одне з питань та/або додаткового питання (або їх ланцюгів), тобто «вирахування» одного з питань не повинне «руйнувати» зазначений стан.

Останнє дозволяє обходити питання нескінченно, тобто завершення тесту можливо в наступних випадках: вичерпані всі питання в банку тестових завдань; досягнуто кінець тесту; рівень знань оцінений з достатньою точністю; рівень знань розташований на досить далекій відстані від значення критерію проходження тесту; суб'єкт тестування показує свою неспроможність при відповідях на запитання тесту.

Подання алгоритму тестування у вигляді Марківських мереж не є вичерпним. Як при правильній, так і при невірній відповіді на запитання повинне ухвалюватися одне з наступних рішень: перехід до наступного основного питання з вибором рівня його складності; перехід до додаткових питань (або до їхніх дерев), при цьому потрібно відкинути вже задані додаткові питання (як окремі, так і їх дерева); повернення до основного питання, якщо отримана відповідь на додатковий (або додаткові) питання; закінчення тестування. «Питання» (тестове завдання) як суб'єкт у розглянутому процесі тестування припускає виконання правил, що встановлені екзаменатором. Це відповідає й традиційному процесу приймання іспиту, тобто залежно від відповіді на поставлені питання екзаменатор ухвалює зазначені вище рішення. При цьому він бере до уваги як фрагментарно розділену відповідь (оцінка за кожне питання й висновок середнього), так і за певними їм правилами весь ланцюжок відповідей суб'єкта тестування.

Загальним підходом у прийнятих методах, крім загальних правил для всіх завдань, є компонування завдань із різних частин (рисуноків, таблиць, мультимедійного змісту), обумовлених як стимули. Це дозволяє заощаджувати ресурси по розміщенню стимулів і виділяти програму обробки та загальні правила в різних частинах ПЗ.

Даний підхід не дозволяє в достатній ступені індивідуалізувати «питання». Насамперед, це пов'язане з тим, що кожне питання поєднує як безпосереднє завдання, так і рішення, що пов'язане з виконанням цього завдання з відповідями на додаткові питання. Враховуючи викладене, для забезпечення гнучкості в прийнятті рішень, простоти створення питань і правил, що визначають прийняття рішень для конкретного питання, доцільним є об'єднання питання й процедури ухвалення рішення, пов'язаного з ним. Такий підхід спрощує як процедуру тестування, так і саму систему тестування з погляду задоволення вимогам мінімальної складності застосовуваних алгоритмів.

Прийняття загальних для всього процесу тестування рішень, вимагає загальних підходів в одному сеансі тестування. Це вимагає використання протоколу тестування й застосування: алгоритмів операційного й



статистичного аналізу результатів тестування з погляду надмірності або недостатності інформації; алгоритмів, що визначають, рівень підготовки суб'єкта тестування; алгоритмів, що забезпечують стохастичні переходи по мережах тестових завдань.

Фактично, модель суб'єкта тестування в окремій сесії визначається протоколом опитування та результатами оцінювання знань.

Таким чином, даний підхід формує те, що є природнім для викладача, має аналог у класичному розумінні іспиту й визначається як модель суб'єкта навчання.

IV. ВИСНОВКИ

Процедуру додаткових питань пропонується виконувати двома способами. Перший спосіб полягає в тому, що до закінчення множини додаткових питань одного основного (або єдиного питання) не здійснюється ніяких дій, крім фіксування відповідей на запитання з наступної передачі протоколу основному питанню, де ухвалюється одне з наступних рішень: зарахувати відповідь на основне питання як правильний з можливим коректуванням складності; вважати питання незадовільним і перейти до групи питань більш низького рівня; повторно поставити додаткове запитання із залученням, що залишилися додаткових питань; повторно поставити основне запитання без залучення додаткових питань.

Останнє є загальною формою, тобто псевдо-інтерактивна процедура може бути зведена до маршової шляхом видалення логічного аналізу й прийняття рішень (тільки перехід у класичному розумінні Марківських мереж).

У випадку, коли додаткові питання, що належать основному питанню, утворюють фрагмент деревоподібної мережі, то при використанні якого-небудь додаткового питання із усього безлічі цих питань певного для тесту взагалі, гілка деревоподібної мережі (графа) знищується повністю. Такий же результат має місце тоді, коли це додаткове питання задане безпосередньо в контексті конкретного основного питання тесту, але в цьому випадку цей результат має місце після закінчення відповіді на додаткові питання, що належать його гілці.

У результаті проведеного аналізу моделей і методів адаптивного тестування можна зробити висновки, що кращими є підходи, засновані на класі багатоступінчастих методів адаптивного тестування, які ефективно підтримують комплексну модель адаптивного тестування.

Дана модель у свою чергу є оптимальною для реалізації контролю знань у дистанційному навчанні.

Однак, виявлене протиріччя між адекватною статистичною оцінкою знань і вмінь випробуваного й обраного методу, заснованого на принципі максимуму інформативності.

Обрана модель представлення тестових завдань із використанням додаткових питань на основі мереж Маркова, яка може забезпечити вирішення зазначеного вище протиріччя.

Для вирішення завдань тестування обраний метод адаптивного тестування Г. Раша. Визначені методи статистичної обробки результатів тестування й побудови функції розподілу ймовірностей правильних відповідей. Наявність логічного аналізу й ухвалення рішення в тестовім завданні моделює правило поведінки екзаменатора. Тобто прийнятий ситуативний підхід на підставі правил для конкретного питання про поведінку (передача керування іншому питанню) у рамках окремо взятого завдання. Це дозволяє не тільки враховувати елементарну безліч результатів відповіді (ложно або істинно), але й те, яка відповідь із безлічі неправильних відповідей обраний з урахуванням стану протоколу тестування до звертання до конкретного питання.

Запропонована модель лінійного програмування для конструювання систем багатоступінчастого тестування, засновану на критерії максимуму інформації, що дозволяє брати до уваги різноманітні обмеження для різних тестових структур. Можливість включення таких обмежень у системи східчастого адаптивного тестування привернула увагу до цього методу.

ЛІТЕРАТУРА REFERENCES

- [1] Learning management system // URL: https://en.wikipedia.org/wiki/Learning_management_system.
- [2] Бейкер, Р. Educational data mining and learning analytics/Р. Бейкер, Г. Сіменс – The Cambridge handbook of the learning sciences, 2019. – 274 с.
- [3] Ржеуцька С. Опыт применения методов кластеризации для анализа результатов дистанционного обучения // Информатизация инженерного образования: материалы международной науч.-практ. конф. 2016. № 56. С. 617 – 620.
- [4] Expectation-maximization algorithm // URL: https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm.
- [5] Кукієр, К. Big Data: A Revolution That Will Transform How We Live, Work, and Think/К. Кукієр, В. Штойнберг, 2017. – 236 с.
- [6] Дашкевич О. Аналіз можливостей Apache Kafka в рамках забезпечення стрімінгу Big Data // Інформаційні системи і технології: матеріали 7-ї Міжнарод. науч.-техн. конф. 2018. № 12. С. 34-35.
- [7] Data Mining with WEKA MOOC – Material // Machine Learning at Waikato University. URL: <https://www.cs.waikato.ac.nz/ml/WEKA/mooc/dataminingwithWEKA>.
- [8] A Simple Step by Step Guide to WEKA // Analyticscosm. URL: <https://analyticscosm.com/a-simple-step-by-step-guide-to-WEKA>

