УДК 004.82

А.В. Олецкий, М.Ф. Махно

О ПОВЫШЕНИИ УРОВНЯ АДЕКВАТНОСТИ РЕЗУЛЬТАТОВ ОЦЕНИВАНИЯ УЧЕБНЫХ ПРОЕКТОВ НА ОСНОВЕ ПАРАМЕТРИЧЕСКОЙ РЕЛАКСАЦИИ МЕТОДА ПАРНЫХ СРАВНЕНИЙ

Ключевые слова: попарные сравнения, транзитивные шкалы, ранжирование альтернатив, нечеткое оценивание, учебные проекты.

Введение

Современное высшее образование, в частности, в сфере IT приобретает все более индустриально-ориентированный характер. Необходимо обеспечить гармоничное сочетание теоретической подготовки с практическими навыками, затребованными на рынке труда. Развивается также направление, связанное с дуальным образованием, при котором университетское обучение в той или иной форме сочетается с выполнением реальных заказов от IT-компаний.

В этом контексте особое значение в процессе обучения приобретает успешное выполнение студентами учебных заданий, связанных с проектированием и разработкой программных продуктов. Естественно, в рамках учебного процесса данные задания должны быть формально оценены по 100-балльной шкале; на учебный проект отводится определенное количество баллов (как правило, 30–40).

При оценивании возможны определенные проблемы. Выделим несколько типичных подходов к оцениванию, которые так или иначе могут комбинироваться. При этом каждый из этих подходов имеет свои преимущества и недостатки.

1. Оценивание экспертным путем на основе общего впечатления преподавателя с учетом его опыта и индивидуальных предпочтений. Предполагается, что оценки, выставленные опытным и квалифицированным преподавателем, достаточно естественны и адекватны. Но такой подход порождает органическую проблему — оценки неминуемо субъективны, и, как правило, преподаватель не может их достаточно полно обосновать. Студенты, особенно наиболее активные, могут предъявлять претензии, что обычно приводит к процессу «торговли» до достижения согласия. Недостатки такой практики очевидны: субъективное повышение балла одному из студентов неминуемо затронет интересы других, что очень нежелательно.

2. Оценивание с учетом определенного набора критериев. Практически повсеместно применяется линейная оценка в виде взвешенной суммы оценок критериев:

$$v_i = \sum_{j=1}^{n} a_{ij} r_j , \; i = 1,..., q, \qquad (1)$$

где q — количество оцениваемых объектов (в данном контексте — учебных проектов); n — количество критериев; r_j — степень важности (вес) j-го критерия; a_{ij} — количественная оценка j-го критерия для i-го объекта.

Во многих случаях такой принцип оценивания позволяет получить удовлетворительные результаты, однако для индустриально-ориентированных учебных проектов он может быть неприемлем. В ряде случаев можно выделить определенный набор формальных критериев. Однако практика показывает, что основная дифференциация связана с оценочными суждениями (внешний вид: нравится–не нравится; интерфейс: удобный–неудобный; программный код: грамотный–неграмотный; и т.п.), что трудно поддается количественной оценке. Кроме того, при экспертном оценивании иногда сложно отделить один критерий от другого.

3. Разработка более совершенных математических методов и алгоритмов автоматизированного оценивания. Их использование непосредственно связано с внедрением в процесс контроля и оценивания знаний специализированных экспертных систем. Среди способов реализации систем, которые хорошо зарекомендовали себя при решении многих задач оценивания, следует отметить базирующийся на применении методики парных сравнений. В основу данной методики положен принцип анализа иерархий Саати [1, 2], по которому проводится сравнение группы объектов путем последовательного сравнения отдельных пар между собой. При этом очевидно, что для экспертов проще сравнивать два объекта, чем получать количественные оценки объектов по заданной абсолютной шкале.

Еще одним эффективным подходом к реализации компьютерных систем контроля и оценивания уровня подготовки является использование схем адаптивного тестирования и интерпретации его результатов [3]. Под адаптивным способом контроля [4] понимают систему научно обоснованной проверки и оценки результатов обучения, отличающуюся высокой эффективностью за счет оптимизации процедур генерации, предъявления и оценивания результатов выполнения совокупности тестов. Адекватность результатов оценивания повышается при многошаговой стратегии отбора и назначения заданий следующего шага тестирования с учетом полного контекстного анализа и оценки результатов выполнения предыдущего шага.

Одним из наиболее известных примеров моделей адаптивного тестирования является методика оценивания знаний на основе тестовых заданий (itemresponsetheory, IRT) [4]. Данная методика, которая состоит в проведении многократных качественных педагогических измерений, включает в себя поиск наиболее адекватных прогностических моделей, а также определение уровня подготовки с учетом пригодности моделей для получаемых в процессе тестирования данных. Как правило, основные задачи IRT сводятся к разработке методов измерения на основе выборочных статистик и других эмпирических характеристик [4–7].

Необходимо отметить, что несмотря на определенные сложности методика оценивания на основе адаптивного тестирования активно внедряется в процессы дистанционного образования и имеет широкие перспективы в повышении эффективности учебного процесса.

В данной статье рассматривается и анализируется простой метод парных сравнений для процедуры оценивания проектов. Т. Саати предложил «стандартную» шкалу, по которой для сравнения вариантов задается набор разных градаций предпочтения с определенными количественными значениями (1 — варианты равнозначны; 3 — слабое превосходство и т.п.; допускаются промежуточные значения) [1, 2]. Однако опыт показывает, что это не всегда дает достаточно хорошие результаты. Становится все более очевидным, что количественные значения градаций предпочтения могут быть разными для различных предметных областей и даже для разных задач в пределах одной предметной области. Особенный интерес представляет ситуация, когда все варианты (проекты) достаточно близки по своему уровню с точки зрения основных требований (критериев). Таким образом, можно говорить, что метод парных сравнений может зависеть от определенного набора параметров, и для конкретного применения может потребоваться «настройка» метода на основе подбора этих параметров.

Подбор параметров для реализации метода парных сравнений приобретает особое значение при построении систем автоматического (или автоматизированного) оценивания вариантов. Понятно, что настройка (обучение) системы должна производиться на основе некоторой обучающей выборки; в рабочем режиме будут использоваться значения параметров, полученные на этапе настройки.

Статья посвящена экспериментальному исследованию и развитию подходов к автоматизации оценивания учебных проектов на основе процедур комбинирования параметризированного метода парных сравнений и нечеткой логики с учетом того, что вопрос о возможности и целесообразности применения методики парных сравнений для оценивания результатов работы в рамках учебного процесса в настоящее время изучен недостаточно. Предложены подходы для более глубокого анализа и сравнения вариантов (проектов) со слабым уровнем предпочтения.

Примеры оценивания при использовании стандартной шкалы

На основе метода парных сравнений с использованием стандартной шкалы приведем примеры оценивания учебных проектов с недостаточно «естественными» результатами.

Пример 1. Рассмотрим простой случай, когда имеется q групп (команд) студентов, каждая из которых выполняет учебный проект. Обычно число таких команд невелико; для определенности положим $q = 3$.

Уровни результатов выполнения проектов можно линейно упорядочить, все они достаточно качественны, и преимущество одного над другим незначительно. В соответствии со «стандартной» шкалой Т. Саати слабое превосходство оценивается числом [1].

При таких предположениях матрица попарных сравнений будет иметь вид

$$M = \begin{pmatrix} 1 & 3 & 3 \\ 1/3 & 1 & 3 \\ 1/3 & 1/3 & 1 \end{pmatrix}.$$

Здесь и далее оценки каждого объекта вычисляются как компоненты перроновского вектора (нормализованного собственного вектора, соответствующего максимальному собственному значению) матрицы M, если явным образом не будет оговорено что-либо иное. Для данного примера вектор оценок по Саати имеет вид (0,58416; 0,28083; 0,13501).

Если на учебный проект выделяется 30 баллов, а наилучший проект оценивается по максимально возможному количеству баллов, то с помощью соответствующей нормировки получаем такое распределение баллов (с округлением до целых): 30; 14; 7 (т.е. наилучший проект оценивается в 30 баллов, а остальные — соответственно в 14 и 7). Такой разброс между максимальной и минимальной оценками не соответствует общепринятой практике — он слишком большой, когда все проекты достаточно качественны.

Зачастую качество результатов, получаемых с помощью метода сравнений с использованием стандартной шкалы, зависит от понятия согласованности матрицы парных сравнений [1]. Но «естественность» результатов сравнения в контексте данной статьи представляет собой совершенно несвязанное с согласованностью понятие.

В примере 1 индекс согласованности равен 0,06781, т.е. отношение согласованности не попадает в диапазон, допустимый по Саати. Однако следующий пример показывает, что недостаточная «естественность» может иметь место даже при идеальной согласованности.

 ISSN 0572-2691

Пример 2. Пусть проекты, как и в примере 1, линейно упорядочены. Первый проект имеет незначительное превосходство над вторым и более значительное — над третьим, а второй проект незначительно превосходит третий. Идеально согласованная матрица парных предпочтений при использовании «стандартной» шкалы имеет вид

$$M = \begin{pmatrix} 1 & 3 & 9 \\ 1/3 & 1 & 3 \\ 1/9 & 1/3 & 1 \end{pmatrix}.$$

В данном случае приходится либо пожертвовать идеальной согласованностью и оценить превосходство первого проекта над третьим меньшим числом (например, 5 или даже 3, как в примере 1), либо все-таки добиваться идеальной согласованности. В последнем случае индекс согласованности равен нулю, а распределение баллов имеет вид: 30; 10; 3, т.е. полученное значение разброса значительно больше, чем в примере 1.

Примеры такого рода можно продолжить, но общий вывод останется неизменным: «естественность» оценок в понимаемом здесь смысле не имеет прямой связи с согласованностью. Можно предложить способ достижения «естественного» разброса за счет параметризации метода парных сравнений.

Общие замечания о настройке параметров метода парных сравнений

На достаточно общем уровне задача настройки параметров метода парных сравнений может быть описана как некоторая задача подбора параметров следующего содержания.

Пусть дано множество сравниваемых объектов $X = \{x^1,..., x^q\}$, состоящее из q элементов. Предположим, что определен некоторый алгоритм оценивания $\Gamma(X, T)$, где $T = \{\tau_1,..., \tau_n\}$ — заданный набор параметров. Тогда при каждом заданном Т в результате выполнения алгоритма будет получено соответствующее распределение оценок, т.е. вектор оценок вариантов $v^{(T)} = (v_1^{(T)}, ..., v_q^{(T)})$, и требуется найти

$$T^* = \arg\max_T Q(v^{(T)}), \tag{2}$$

где Q — заданный критерий качества.

Существенная сложность задачи может заключаться в том, что критерий Q трудно формализовать, что в полной мере относится и к поставленной задаче оценивания учебных проектов. С другой стороны, если множество оценок, полученное путем явного перебора возможных предпочтений, невелико, эксперты могут достаточно просто сравнивать получаемые векторы оценок $v^{(\mathrm{T})}$ и, соответственно, выбирать набор параметров, который, по их мнению, является наиболее подходящим.

Как отмечалось ранее, для решения задачи оценивания учебных проектов существенное значение имеет «естественность» оценок. Понятие «естественности» трудно формализуемо; при этом следует ориентироваться лишь на то, что оценки, автоматически выставляемые обученным алгоритмом, не должны значительно отличаться от оценок, которые выставили бы квалифицированные преподаватели и/или эксперты.

Естественно, в большинстве случаев нет необходимости в количественной оценке экспертами каждого конкретного проекта. Вместо этого рассматривается некоторый интегрированный показатель, а именно разброс максимальной и минимальной оценок (т.е. оценок, выставленных за наилучший и наихудший проекты). Этот разброс количественно оценивается как $\dfrac{\max\limits_{i=\overline{1,q}} v_i^{(T)}}{\min\limits_{i=\overline{1,q}} v_i^{(T)}}$; предполагается, что $\min\limits_{i} v_i^{(T)} > 0.$

Отметим, что этот разброс не должен быть ни слишком большим, ни слишком малым — эксперименты показывают, что наибольшая естественность и, соответственно, наилучшие результаты достигаются при некоторых средних значениях данного разброса. В общем случае это может быть сформулировано в более формализованном виде. Кроме того, могут быть предложены соответствующие математические постановки, однако в рассматриваемых примерах качество решения оценивалось экспертным путем.

В [2] рассматривается простой подход к параметризации метода парных сравнений, в соответствии с которым вводится некоторый параметр $\tau > 1,$ и каждая следующая градация предпочтения получает количественное значение, в τ раз большее, чем предыдущая. Таким образом, слабое превосходство оценивается как τ, заметное — τ^2 и т.д. Именно этот подход был исследован в первую очередь; в этом простом случае задача (2) становится однопараметрической. Приведенные ниже эксперименты показывают, что при снижении значения параметра уменьшается и разброс максимальной и минимальной оценок, поэтому назовем параметр τ релаксационным. Таким образом, для рассматриваемой задачи первостепенное значение приобретает такой подбор релаксационного параметра, чтобы результаты оценивания были максимально естественными (далее в статье приведен пример такого подбора).

Оценка за наилучший проект не обязательно должна быть максимальной; это зависит от его качества. Поэтому в данном случае целесообразно ввести еще один показатель — оценка за наилучший проект H, которая определяется исходя из степени удовлетворенности эксперта (преподавателя) самим проектом и максимально возможной оценки M.

В простейшем случае можно положить $H = k \cdot M$, где k — степень удовлетворенности, оцениваемая экспертным путем. Однако величину этого параметра сложно оценить с достаточной степенью точности: оценка, как правило, субъективна. Кроме того, зависимость окончательной оценки от степени удовлетворенности не обязательно должна быть линейной. В некоторых сложных ситуациях целесообразно использовать аппарат теории нечетких множеств для фаззификации параметра и применения, в частности, способа принятия решений на основе метода центра тяжести max-min-композиции, который хорошо описан в [8].

Основная идея такого подхода состоит в оценке проекта (ограничимся рассмотрением только наилучшего), которая может быть вычислена как центр тяжести нечеткого множества $\tilde{B}'$ по формуле

$$\tilde{B}' = \tilde{A}' \bullet (\tilde{A} \to \tilde{B}),$$

где $\tilde{A}$ — нечеткое множество, характеризующее нечеткое понятие «отличный проект»; $\tilde{B}$ — нечеткое множество, характеризующее понятие «отличная оценка»; $\tilde{A}'$ — нечеткое множество, характеризующее фактический уровень проекта; $\to$ — отношение нечеткой импликации, соответствующее логическому правилу «если проект отличный, то оценка отличная»; $\bullet$ — операция max-min-композиции.

Для практического использования нужно постулировать функции принадлежности нечетких множеств $\tilde{A},$ $\tilde{B}$ и $\tilde{A}',$ а также лингвистическую переменную, описывающую такие понятия, как «отличный проект», «очень хороший проект», «хороший проект», «средний проект», «плохой проект», «очень плохой проект» и т.п.

Сформулируем общую эвристическую методику оценивания проектов следующим образом.

1. Вычислить оценку H наилучшего проекта.
2. Получить матрицу попарных сравнений A.
3. Получить перронов вектор v для матрицы A.
4. В качестве рекомендуемых оценок по проектам взять округленные до целочисленных значений координаты полученного вектора, масштабированные так, чтобы максимальное значение было равно H.

Дальнейшее развитие подхода может быть связано с анализом соотношения (1). Первостепенное значение здесь приобретает вопрос надежности оценок и доверия к ним. В частности, если в соотношении (1) коэффициенты a_{ij}, $i = \overline{1, q}$, $j = \overline{1, n}$, и веса r_j, $j = \overline{1, n}$, известны, обоснованы и надежны, то можно непосредственно вычислить оценки v_i, $i = \overline{1, q}$. Если известны a_{ij} и v_i, $i = \overline{1, q}$, $j = \overline{1, n}$, то веса r_j, $j = \overline{1, n}$, можно получить на основе решения соответствующей системы линейных алгебраических уравнений (СЛАУ). Но возможны и ситуации, когда об абсолютной надежности и достоверности говорить сложно и имеются существенные расхождения между оценками, полученными разными способами.

В более общем случае можно говорить о решении нечетких СЛАУ типа $\tilde{A}\tilde{x} = \tilde{v}$. Эти соотношения аналогичны (1), но их компоненты заданы неточно (в частности, для формализации неточности можно использовать трапециевидные, треугольные нечеткие числа и величины [9, 10]). Это приводит, в зависимости от природы неточности, к возникновению различных новых типов задач.

— $\tilde{A}$ известна, правую часть $\tilde{v}$ можно оценить на основе парных сравнений, требуется найти $\tilde{x}$;

— исходя из оценок $\tilde{v}$ и $\tilde{x}$ и некоторой априорной информации об $\tilde{A}$ нужно восстановить $\tilde{A}$.

Если значения параметров метода задаются с некоторой степенью точности, то при определенных условиях можно рассматривать оптимизационные задачи вида (2). Следует отметить, что в последнее время предложены новые подходы к решению задач такого рода и представлению фигурирующих в них нечетких и неопределенных величин [8–12]; использование этих подходов в контексте обсуждаемых вопросов представляется весьма перспективным.

Примеры использования методики оценивания

Для исследования описанной методики оценивания были проведены вычислительные эксперименты с использованием программных средств.

Эксперимент 1. Этот эксперимент непосредственно связан с учебным процессом. В процессе изучения одного из университетских учебных курсов студенты были разделены на три команды. Командам было предложено разработать технологически сложный проект в виде реально функционирующего приложения; у всех было одинаковое задание. Таким образом, в этом случае $q = 3$.

Оценка проектов осуществлялась группой экспертов, сформированной из университетских преподавателей и представителей IT-компании.

После защиты проектов эксперты пришли к однозначному выводу, что первый проект был лучше второго, а второй — лучше третьего, т.е. имела место транзитивность и линейная упорядоченность. Таким образом, параметризированная матрица попарных сравнений имела вид

$$A=\begin{pmatrix} 1 & \tau & \tau^2 \\ \frac{1}{\tau} & 1 & \tau \\ \frac{1}{\tau^2} & \frac{1}{\tau} & 1 \end{pmatrix}.$$

Индекс согласованности равен нулю, т.е. матрица идеально согласована.

Максимально возможная оценка за проект — 30 баллов. В этом эксперименте существенных замечаний к наилучшему проекту не было. Поэтому в применении описанной выше методики оценки наилучшего проекта на основе нечетких множеств не было необходимости — его оценка равнялась максимально возможной, т.е. 30.

Это простой случай, для которого задача подбора параметра τ решалась простым перебором, так как область перебора невелика, а особо высокая точность не требуется. Для разных значений рассчитаны предлагаемые оценки, основные из которых представлены в табл. 1. Эксперты имели возможность сравнить результаты работы алгоритма оценивания и выбрать наиболее подходящее распределение оценок (или несколько таких распределений) исходя из их представлений об «естественности».

Таблица 1

№ п/п	Релаксационный параметр	Рекомендуемые оценки
1	1,02	30 29 29
2	1,05	30 29 27
3	1,06	30 28 27
4	1,07	30 28 26
5	1,08	30 28 26
6	1,09	30 28 25
7	1,1	30 27 25
8	1,15	30 26 23
9	1,2	30 25 21
10	1,3	30 23 18
11	1,4	30 21 15
12	1,5	30 20 13

Экспертным путем определено, что наиболее «естественные» результаты представлены строками 2–5 (табл. 1), что соответствует значениям параметра $\tau \in [1{,}05, 1{,}08]$.

Эксперимент 2. Предположим также, что при наличии той же матрицы парных сравнений, что и в эксперименте 1, наилучший проект не заслуживает максимально возможной оценки и оценен экспертами как «хороший, но не отличный».

Найдем оценку для наилучшего проекта. В качестве носителя (базовой шкалы) для $\tilde{B}$ возьмем значения оценок w от 0 до 30, в качестве базовой шкалы для $\tilde{A}$ и $\tilde{A}'$ — оценок u от 0 до 10.

Зададим в табличном виде функции принадлежности для соответствующих нечетких множеств:

$\mu_{\tilde{A}}(u)$ = {(0; 0), (1; 0), (2; 0), (3; 0,1), (4; 0,15), (5; 0,2), (6; 0,4), (7; 0,5), (8; 0,7), (9; 0,9), (10; 1,0)};

$\mu_{\tilde{A}'}(u)$ = {(0; 0), (1; 0), (2; 0,05), (3; 0,15), (4; 0,25), (5; 0,35), (6; 0,5), (7; 0,8), (8; 1), (9; 0,95), (10; 0,9)};

$\mu_{\tilde{B}}(w)$ = {(0; 0), (10; 0), (15; 0,1), (20; 0,3), (22; 0,5), (25; 0,7), (27; 0,8), (29; 0,9), (30; 1)}.

Найдем нечеткое множество $\tilde{B}'$. Результат зависит от того, какую формулу использовать для отношения нечеткой импликации. В случае min-импликации Мамдани — $\mu_{\tilde{B}'}(w)$ = {(0; 0), (10; 0), (15; 0,1), (20; 0,3), (22; 0,5), (25; 0,7), (27; 0,8), (29; 0,9) (30; 0,9)}. Центр тяжести этого множества — $H \approx 26{,}36$.

Дальнейшие рассуждения в эксперименте 2 (применение метода парных сравнений) ничем не отличаются от выводов эксперимента 1, можно использовать полученные в нем результаты. Возьмем одно из допустимых распределений в виде 30; 28; 26, которое получается при τ = 1,07. Масштабирование по H дает окончательное распределение оценок вида 26; 24; 22, которое выглядит довольно «естественным» и неплохо согласуется с практикой оценивания студенческих проектов.

Разумеется, эти рассуждения носят эвристический характер, как и выбор вида функций принадлежности. При изменении этих функций результат может несколько измениться, но при разумном выборе функций принадлежности изменения не будут существенными, что подтверждено экспериментальным путем. Приведем один из таких результатов.

Эксперимент 3. Изменим вид функций принадлежности. Зададим функцию принадлежности для $\tilde{A}$ в виде

c = {(0; 0), (1; 0,1), (2; 0,2), (3; 0,3), (4; 0,4), (5; 0,5), (6; 0,6), (7; 0,7), (8; 0,8), (9; 0,9), (10; 1)}.

Для $\tilde{A}'$ изменим значение меры принадлежности, на котором достигается максимум, уменьшив при этом величину носителя (функция принадлежности становится более крутой):

$\mu_{\tilde{A}}(u)$ = {(0; 0), (1; 0), (2; 0), (3; 0), (4; 0), (5; 0), (6; 0,1), (7; 0,5), (8; 0,9), (9; 1), (10; 0,5)}.

Для нечеткого числа $\tilde{B}$ зададим функцию принадлежности вида

$\mu_{\tilde{B}}(w)$ = {(0; 0), (10; 0,1), (15; 0,2), (20; 0,5), (22; 0,9), (25; 1), (27; 1), (29; 1), (30; 1)}.

Теперь $H \approx 25{,}23$, и получаем следующее распределение оценок: 25; 23; 21.

Отметим, что зависимость получаемых оценок от выбора функций принадлежности имеет место. С другой стороны, можно надеяться, что при рациональном выборе функций принадлежности этот фактор не будет играть решающей роли.

Более сложный случай процесса оценивания

Эксперимент 4. Как и ранее, максимально возможная оценка H равна 30, наилучший проект хороший, но не отличный (аналогично экспериментам 2 и 3).

Пусть имеется шесть проектов, т.е. q = 6. Основные предпочтения имеют вид

$$1 \succ 2;\ 2 \succ 3;\ 3 \succ 4;\ 4 \succ 5;\ 5 \succ 2;\ 5 \succ 6$$

(здесь символом $\succ$ обозначено отношение строгого предпочтения «лучше»).

Как видим, в этом случае свойство транзитивности отсутствует. Известно, что метод Саати хорошо зарекомендовал себя и в случае отсутствия транзитивности, но нужно рационально определить матрицу парных сравнений.

Построим матрицу парных сравнений. Граф предпочтений можно разбить на непересекающиеся сильно связные компоненты, в данном случае имеются три сильно связные компоненты: G_1 = {1}; G_2 = {2, 3, 4, 5}; G_3 = {6}. Интуитивно понятно, что оценки узлов, относящихся к одной сильно связной компоненте, не должны сильно отличаться друг от друга. Поэтому разумно определить степени превосходства между компонентами G_2 как слабые. Исходя из этого построим такую матрицу парных сравнений:

$$\begin{pmatrix} 1 & \tau & \tau & \tau & \tau & \tau^2 \\ \frac{1}{\tau} & 1 & \tau & \tau & \frac{1}{\tau} & \tau \\ \frac{1}{\tau} & \frac{1}{\tau} & 1 & \tau & \tau & \tau \\ \frac{1}{\tau} & \frac{1}{\tau} & \frac{1}{\tau} & 1 & \tau & \tau \\ \frac{1}{\tau} & \tau & \frac{1}{\tau} & \frac{1}{\tau} & 1 & \tau \\ \frac{1}{\tau^2} & \frac{1}{\tau} & \frac{1}{\tau} & \frac{1}{\tau} & \frac{1}{\tau} & 1 \end{pmatrix}.$$

В данном случае можно использовать выводы относительно рекомендованных значений параметра τ, полученные в результате эксперимента 1. Однако здесь иная ситуация, в которой желательно провести обучение в целях подбора параметра τ, дающего более «естественные» оценки для этой ситуации. Методика подбора полностью аналогична эксперименту 1. Некоторые результаты представлены в табл. 2.

Таблица 2

№ п/п	Релаксационный параметр	Индекс согласованности	Рекомендуемые оценки
1	1,02	≈ 0	25 25 25 25 25 24
2	1,05	≈ 0	25 24 24 24 24 23
3	1,07	≈ 0	25 24 24 23 23 22
4	1,09	0,0013	25 24 24 23 23 21
5	1,10	0,0016	25 23 23 23 23 21
6	1,11	0,019	25 23 23 22 22 20
7	1,12	0,023	25 23 23 22 22 20
8	1,14	0,0031	25 23 23 22 22 19

Как и ожидалось, на основе предварительного анализа графа предпочтений первый проект имеет преимущество над остальными. Второй, третий, четвертый и пятый проекты получают приблизительно равные оценки, шестой несколько от них отстает. Результаты, получаемые при $\tau \in [1{,}09{,}1{,}14]$, представляются довольно «естественными».

Отметим также, что на основе нечеткой min-max-композиции можно было бы рассчитывать оценки для всех объектов, а не только для наилучшего, но использование сравнений и привязка оценок к оценкам победителя делает их более обоснованными.

Поскольку преимущество наилучшего проекта перед другими может быть довольно незначительным, вполне вероятно, что эти проекты попали бы в одну и ту же нечетко формализованную категорию качественных проектов, описанную соответствующей функцией принадлежности. Тогда все оценки проектов совпадали бы между собой. Кроме того, учет привязки других учебных проектов к оценкам наилучшего усиливает соревновательный характер обучения.

Усложнение редуцированной схемы оценивания при отсутствии транзитивности

Увеличение количества оцениваемых объектов приводит к росту размеров матрицы парных сравнений. Явное выделение сильно связных компонент позволяет разбить одну матрицу парных предпочтений на некоторое количество подматриц, каждая из которых соответствует отдельной сильно связной компоненте.

 ISSN 0572-2691

При этом предпочтения между разными сильно связными компонентами, очевидно, должны быть выражены более отчетливо. Поэтому разумно рассматривать, по крайней мере, несколько релаксационных параметров: для сравнений узлов, относящихся к одной сильно связной компоненте, и сравнения между собой сильно связных компонент. Следует также отметить, что отношение предпочтения между сильно связными компонентами уже, очевидно, будет транзитивным.

Таким образом, можем рассчитать оценки для каждой сильно связной компоненты отдельно, а потом скомбинировать эти оценки. При этом вместо общей матрицы попарных сравнений имеем дело с несколькими матрицами парных сравнений меньшего размера.

В данном случае редуцированный алгоритм оценивания может быть записан в следующем виде.

1. Выделить в графе предпочтений сильно связные компоненты G_k, $k=\overline{1,\, m}$.

2. Для каждой G_k рассчитать по методу Саати веса входящих в нее узлов u_{ik}, $i=\overline{1,\, p_k}$, $k=\overline{1,\, m}$.

3. Для каждой G_k масштабировать u_{ik} так, чтоб выполнялось соотношение $\max\limits_i u_{ik}=1$, $i=\overline{1,\, p_k}$, $k=\overline{1,\, m}$.

4. Построить матрицу парных сравнений для сильно связных компонент A.

5. Вычислить вектор весов w как главный собственный вектор матрицы A, масштабированный так, чтобы его максимальная компонента была равна единице.

6. Получить комбинированные оценки для всех объектов:

для каждой G_k {для всех объектов из G_k положить $v_i=u_{ik}w_k$ }, $i=\overline{1,\, p_k}$, $k=\overline{1,\, m}$.

7. Масштабировать полученные на шаге 6 оценки так, чтобы максимум равнялся H, где H — рассчитанное экспертным путем количество баллов за наилучший объект.

Проиллюстрируем работу алгоритма на данных из эксперимента 4. При этом очевидно, что вычислять веса u_{ik} имеет смысл только для компонент с двумя и более узлами. В нашем примере это — G_2; вычислим веса для входящих в нее узлов. Матрица предпочтений имеет вид

$$\begin{pmatrix} 1 & \tau & \tau & \frac{1}{\tau} \\ \frac{1}{\tau} & 1 & \tau & \tau \\ \frac{1}{\tau} & \frac{1}{\tau} & 1 & \tau \\ \tau & \frac{1}{\tau} & \frac{1}{\tau} & 1 \end{pmatrix}.$$

Положим $\tau=1{,}07$ (параметр из допустимого диапазона, полученного в эксперименте 1). Тогда индекс согласованности равен 0,0019. Перронов вектор равен (0,2544; 0,2541; 0,2456; 0,2459).

При масштабировании получаем (1,00000; 0,99890; 0,96571; 0,96685).

Для первого и шестого объектов положим компонентные оценки u_{11}, u_{31} равными единице.

Теперь получим оценки для сильно связных компонент. Для соответствующей матрицы парных сравнений возьмем другой релаксационный параметр $\sigma>\tau$, пусть $\sigma=1{,}14$ (параметр подобран экспериментально). Матрица парных сравнений для сильно связных компонент будет иметь вид

$$A = \begin{pmatrix} 1 & \sigma & \sigma^2 \\ \frac{1}{\sigma} & 1 & \sigma \\ \frac{1}{\sigma^2} & \frac{1}{\sigma} & 1 \end{pmatrix}.$$

В данном случае ее индекс согласованности равен нулю, перронов вектор — (0,3778; 0,3314; 0,2907), а масштабированный по единице — (1,0000; 0,8772; 0,7695).

Получим немасштабированные комбинированные оценки всех объектов (шаг 6 алгоритма):1,0000; 0,8772; 0,8762;0,8471; 0,8481; 0,7695.

Масштабируя по H = 25,227 (использованы функции принадлежности из эксперимента 3), получаем окончательные рекомендуемые оценки: 25; 22; 22; 21; 21; 19.

Таким образом, имеем оценки, соответствующие «естественному» соотношению проектов.

Заключение

В статье показана возможность построения настраиваемых алгоритмов автоматизированного оценивания результатов работы студентов во время учебного процесса, в частности построенного на основе дуального обучения. Предлагается эвристическая методика расчета, основанная на нечетком оценивании объектов и парных сравнениях между ними. Методика зависит от некоторого набора параметров, в частности:

— релаксационные параметры, используемые при построении матрицы парных сравнений; один из них может применяться при сравнениях в рамках одной сильно связной компоненты графа предпочтений, а другой — между разными сильно связными компонентами, что имеет значение в случае отсутствия транзитивности;

— нечеткие множества, характеризующие оценку качества проектов. В частности, может быть выставлена оценка за наилучший проект, на основе которой оцениваются все остальные.

Кроме того, на них могут строиться автоматизированные системы оценивания, которые позволили бы задавать правила для соответствующих алгоритмов, а также непосредственно осуществлять парные сравнения.

Описанные подходы можно применять и для решения других подобных задач. По сути, речь идет о ситуациях, когда получение или применение обоснованных критериев для оценивания объектов невозможно или затруднительно и на первый план выходит автоматизированное экспертное оценивание, на базе которого происходит построение алгоритма на основе метода парных сравнений и подбор его параметров.

Эти направления должны стать предметом дальнейших исследований.

О.В. Олецький, М.Ф. Махно

ПРО ПІДВИЩЕННЯ РІВНЯ АДЕКВАТНОСТІ РЕЗУЛЬТАТІВ ОЦІНЮВАННЯ УЧБОВИХ ПРОЕКТІВ НА ОСНОВІ ПАРАМЕТРИЧНОЇ РЕЛАКСАЦІЇ МЕТОДУ ПАРНИХ ПОРІВНЯНЬ

Розглянуто проблему оцінювання учбових проектів студентів під час навчального процесу. Пропонується евристична методика побудови автоматизованих систем алгоритмічного оцінювання, що базується на нечіткому оцінюванні об'єктів та на попарних порівняннях між ними. Для підвищення адекватності та природності оцінок пропонується підхід, що ґрунтується на введенні релаксаційного параметра. Це дозволяє зменшити розкид між максимальними та мінімальними оцінками в порівнянні зі стандартною шкалою Сааті. Для оцінки найкращого варіанта застосовано один із методів нечіткого прийняття рішень, а саме метод центра тяжіння композиції «максимум–мінімум»; оцінки інших варіантів отримуються відповідним нормуванням. Крім того, пропонується алгоритм оцінювання для нетранзитивних відношень переваг, що базується на виділені сильно зв'язних компонент та попарних порівняннях між ними; при цьому для кожної підзадачі релаксаційні параметри слід підбирати окремо. Таким чином, пропонується комбінована методика оцінки альтернатив, яка залежить від таких параметрів, як релаксаційні параметри для матриць попарних порівнянь в межах окремих сильно зв'язних компонент; релаксаційний параметр для матриці попарних порівнянь між сильно зв'язними компонентами; нечітка функція належності, що задає якість найкращого варіанта.

Ключові слова: попарні порівняння, транзитивні шкали, ранжування альтернатив, нечітке оцінювання, навчальні проекти.

O.V. Oletsky, M.F. Mahno

ENHANCING ADEQUACY OF GRADING STUDY PROJECTS ON THE BASE OF PARAMETRIC RELAXATION OF PAIRWISE COMPARISONS

A problem of automated assessing of students' study projects is regarded. A heuristic algorithm based on fuzzy estimating of projects and on pairwise comparisons among them is proposed. For improving adequacy and naturalness of grades, an approach based on introducing a parameter named relaxation parameter was suggested in the paper. This enables to reduce the spread between maximum and minimum values of projects in comparison with the one in the standard scale suggested by T. Saati. Reasonable values of this parameter were selected experimentally. For estimating the best alternative, a center of mass of a fuzzy max-min composition should be calculated. An estimation algorithm for a case of non-transitive preferences based on getting strongly connected components and on pairwise comparisons between them is also suggested. In this case, relaxation parameters should be chosen separately for each subtask. So the combined technique of evaluating alternatives proposed in the paper depends of the following parameters: relaxation parameters for pairwise comparisons matrices within each strongly connected components; relaxation parameter for pairwise comparisons matrices among strongly connected components; membership function for describing the best alternative.

Keywords: pairwise comparisons, transitive scales, ranking of alternatives, fuzzy evaluation, study projects.

REFERENCES

1. Саати Т. Принятие решений. Метод анализа иерархий. М. : Радио и связь, 1993. 278 с.
2. Черноруцкий И.Г. Методы принятия решений. СПб : БХВ-Петербург, 2005. 416 с.
3. Івохін Є.В., Махно М.Ф. Розробка засобів адаптивного тестування та автоматичного оцінювання знань. *Вісник Київського національного університету імені Тараса Шевченка. Серія фіз.-мат.науки*. 2015. № 1. С. 130–133.
4. Hambleton R.K., Jones R.W. Comparison of classical test theory and item response theory and their applications to test development. Educational Measurement: Issues and Practice. 1993. **12** (3). P. 38–47.
5. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. М. : Прометей, 2000. 168 с.
6. Lord F.M., Novick M.R. Statistical theories of mental test scores. Book; Reading, MA : Addison-Wesley, 1968. 593 p.
7. Croker L., Algina J. Introduction to classical and modern test theory. SF: Cengage Learning, 2006. 527 p.
8. Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей. Рига : Зинатне, 1990. 184 с.
9. Лю Б. Теория и практика неопределенного программирования. М. : БИНОМ. Лаборатория знаний, 2005. 416 с.
10. Ivokhin E.V., Apanasenko D.V. Clustering of composite fuzzy numbers aggregate based on sets of scalar and vector levels. *Journal of Automation and Information Sciences*. 2018. **50**, N 10. P. 47–59. DOI: 10.1615/JAutomatInfScien.v50. i10.40.
11. Глибовець М.М., Олецький О.В. Штучний інтелект. К. : Вид. дім «Академія», 2002. 366 с.
12. Семенова Н.В., Колечкина Л.Н., Нагорная А.Н. Векторные задачи оптимизации с линейными критериями на нечетко заданном комбинаторном множестве альтернатив. *Кибернетика и системный анализ*. 2011. № 2. С. 88–99.

Получено 01.03.2019
После доработки 25.11.2020

ISSN 0572-2691