

РАЗРАБОТКА РАНЖИРУЮЩИХ АЛГОРИТМОВ ДЛЯ РЕШЕНИЯ СЛАБОФОРМАЛИЗУЕМЫХ ЗАДАЧ

В.В. Ворошилов

(Самарский государственный архитектурно-строительный университет)

Научный руководитель – доктор технических наук, профессор С.А. Пиявский

(Самарский государственный архитектурно-строительный университет)

Предлагаются новые алгоритмы и методы для восстановления слабоформализуемых моделей в различных областях экономики и культурно-социальной сферы. Эти алгоритмы использовались для решения двух задач в достаточно слабо формализованных областях: ранжирование ведущих футболистов, ранжирование стран по степени готовности к внедрению электронного обучения.

Введение

Математические методы и моделирование сложных явлений на ЭВМ все более проникают в такие слабо формализованные области, как педагогика, психология, спорт, политика. Одним из важных направлений при этом является алгоритмизация знаний высококвалифицированных экспертов, необходимая для создания систем поддержки принятия решений и экспертных систем. В настоящей статье разрабатывается метод, позволяющий формализовать логику ранжирования некоторых объектов на основе анализа набора их количественных характеристик (критериев).

Пусть дана упорядоченная группа из n объектов, каждый из которых имеет количественные оценки по p критериям. Требуется на основании этих данных восстановить потенциально существующий, но неизвестный, алгоритм, с помощью которого упорядочены объекты.

Предлагаются два метода идентификации ранжирующих алгоритмов:

- интерактивный поиск ранжирующего алгоритма;
- оптимальный поиск ранжирующего алгоритма.

При этом структура ранжирующего алгоритма выбирается в одном из двух видов: линейной свертки или метода ПРИНН [1].

Основная часть

Интерактивный поиск ранжирующего алгоритма

В основе интерактивного метода ограниченного перебора используется метод ПРИНН-У. Соответственно структура ранжирующего алгоритма представляется в виде следующей функции, предложенной в [1]:

$$F = \frac{\sum_{j=1}^p 3^{\alpha_j-1} \cdot f_{i,j}}{\sum_{j=1}^p 3^{\alpha_j-1}},$$

где α_j – заранее неизвестный индекс, показывающий отношение к той или иной группе важности критерия j ($j=1, \dots, p$), $\alpha_j \in [1, k]$. Здесь k – коэффициент, указывающий высшую группу важности критерия; $f_{i,j}$ – оценка i -й альтернативы по j -му критерию ($i=1, \dots, n$). При заданных значениях коэффициентов можно вычислить значения ран-

жирующей функции для всех объектов, ранжировать объекты в соответствии с ними и рассчитать критерий качества формализации J по формуле

$$J = \sum_{i=1}^n abs(N_u(j) - N_p(j)).$$

Методом перебора целочисленных коэффициентов α_j при заданном значении k минимизируют этот критерий, получая наиболее близкое решение к исходному ранжированию. При этом количество переборов, конечно, при большом количестве критериев может оказаться весьма значительным. В данном случае специально разработанная программа предоставляет эксперту (ЛПР) возможность интуитивно указать для каждого критерия более узкие границы варьирования индекса α_j , получая созданный на основе его предпочтений ранжирующий алгоритм с определенной погрешностью.

Оптимальный поиск ранжирующего алгоритма

Несмотря на предоставленную возможность ограничивать перебор, может оказаться, что получить достаточно качественную формализацию интуитивным путем не удастся, а полный перебор невозможен. В этом случае необходимо использовать такой, более мощный метод оптимизации, как линейное программирование, для чего требуется разработать специальную математическую модель.

Пусть структура ранжирующего алгоритма, как и выше, представлена функцией

$$F = \frac{\sum_{j=1}^p 3^{\alpha_j-1} f_{i,j}}{\sum_{j=1}^p 3^{\alpha_j-1}}, \quad (1)$$

где α_j – индекс, показывающий отношение к той или иной группе важности критерия j ($j = 1, \dots, p$); $\alpha_j \in [1, k]$, где k – коэффициент, показывающий высшую группу важности критерия; $f_{i,j}$ – оценка i -й альтернативы по j -му критерию ($i = 1, \dots, n$).

Условие задачи можно представить в виде

$$F_{i+1} \leq F_i, \quad i = 1, \dots, n-1; \quad (2)$$

Записав (2) через (1), получим:

$$\frac{\sum_{j=1}^p 3^{\alpha_j-1} f_{i+1,j}}{\sum_{j=1}^p 3^{\alpha_j-1}} \leq \frac{\sum_{j=1}^p 3^{\alpha_j-1} f_{i,j}}{\sum_{j=1}^p 3^{\alpha_j-1}},$$

или

$$\sum_{j=1}^p 3^{\alpha_j-1} f_{i+1,j} \leq \sum_{j=1}^p 3^{\alpha_j-1} f_{i,j}; \quad i = 1, \dots, n-1. \quad (3)$$

Здесь условие становится нелинейным. Преобразуем задачу из нелинейной в линейную. Для этого введем вместо α_j дополнительные целочисленные переменные Z_j^k , $Z_j^k = \{0,1\}$, $j = 1, \dots, p$; вместо 3^{α_j-1} запишем $(1 + 2Z_j^1 + 8Z_j^2 + \dots + (3^{k-1} - 1)Z_j^k)$ и введем еще одно ограничение:

$$Z_j^1 + Z_j^2 + \dots + Z_j^k \leq 1, \quad j = 1, \dots, p. \quad (4)$$

Тогда (3) принимает вид

$$\sum_{j=1}^p (1 + 2Z_j^1 + 8Z_j^2 + \dots + (3^{k-1} - 1)Z_j^k) f_{i+1,j} \leq \sum_{j=1}^p (1 + 2Z_j^1 + 8Z_j^2 + \dots + (3^{k-1} - 1)Z_j^k) f_{i,j},$$

или

$$\sum_{j=1}^p [f_{i+1,j} - f_{i,j}] (1 + 2Z_j^1 + 8Z_j^2 + \dots + (3^{k-1} - 1)Z_j^k) + KU_i \leq 0; \quad i = 1, \dots, n-1. \quad (5)$$

Целевую функцию можно записать так:

$$\Phi = \sum_{i=1}^{n-1} U_i \rightarrow \min,$$

а (4), (5) – условия. Таким образом, задача сводится к задаче целочисленного линейного программирования, которая решается уже известными методами.

Интересно провести сопоставление результатов формализации в случае, когда для структуризации ранжирующей функции используется менее естественный, чем метод ПРИНН, но более традиционный метод линейной свертки критериев.

Зададимся структурой ранжирующего алгоритма в виде функции

$$F = \sum_{j=1}^p \alpha_j f_{i,j}, \quad (6)$$

где α_j – неизвестный весовой коэффициент важности критерия j ($j = 1, \dots, p$); $f_{i,j}$ – оценка i -й альтернативы по j -му критерию ($i = 1, \dots, n$).

$$\alpha_j \leq 0, \quad j = 1, \dots, p; \quad (7)$$

$$\sum_{j=1}^p \alpha_j = 1. \quad (8)$$

Условие задачи можно представить в виде

$$F_{i+1} \leq F_i, \quad i = 1, \dots, n-1. \quad (9)$$

Запишем (9) через (6):

$$\sum_{j=1}^p \alpha_j f_{i+1,j} \leq \sum_{j=1}^p \alpha_j f_{i,j}; \quad i = 1, \dots, n-1. \quad (10)$$

Введем дополнительные целочисленные переменные U_i :

$$U_i = \{0, 1\}, \quad i = 1, \dots, n-1. \quad (11)$$

Используя (11), можем записать (10) в виде

$$\sum_{j=1}^p \alpha_j f_{i+1,j} + KU_i \leq \sum_{j=1}^p \alpha_j f_{i,j}; \quad i = 1, \dots, n-1,$$

или

$$\sum_{j=1}^p \alpha_j [f_{i+1,j} - f_{i,j}] + KU_i \leq 0; \quad i = 1, \dots, n-1. \quad (12)$$

где K – очень большое число. Тогда целевую функцию можно записать так:

$$\Phi = \sum_{i=1}^{n-1} U_i \rightarrow \min,$$

а (7), (8), (12) рассматривать как ограничения.

Таким образом, эта задача сводится к решению задачи целочисленного линейного программирования, которая решается уже известными методами.

Примеры применения

Предложенные методы идентификации ранжирующих алгоритмов были реализованы в универсальной информационной системе (ИС), разработанной на языке Borland Delphi 5.5. Информационная система использовалась для решения двух задач в достаточно слабо формализованных областях:

- ранжирование ведущих футболистов,
- ранжирование стран по степени готовности к внедрению электронного обучения.

Ранжирование ведущих футболистов

Рассмотрим задачу восстановления алгоритма комплексной оценки ценности различных футболистов по отдельным характеристикам их футбольной квалификации. Для примера были выбраны 9 лучших футболистов по версии FIFA 2003 года. Их исходное ранжирование по этой версии приведено в табл. 1. В качестве критериев использовались 31 характеристика и оценки футболистов, предоставленные компанией Sports Interactive [4]. На рис. 1–4 показаны некоторые экраны ИС на примере обсуждаемой задачи.

Футболист	Рейтинг
Павел Недвед	1
Тьерри Анри	2
Паоло Мальдини	3
Андрей Шевченко	4
Зинедин Зидан	5
Руд ван Нистелрой	6
Рауль Гонсалес	7
Роберто Карлос	8
Девид Бекхем	9

Таблица 1. Рейтинг лучших футболистов по версии FIFA 2003 года

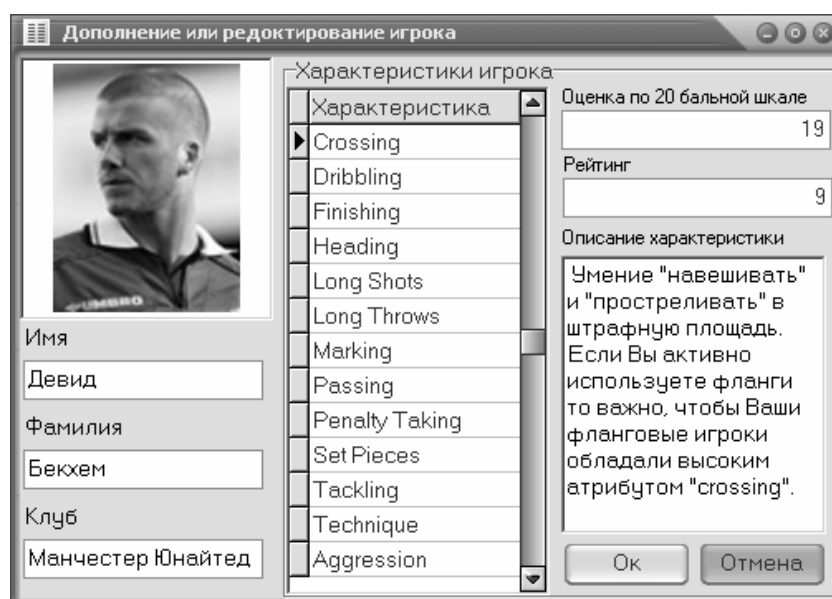


Рис. 1. Окно ввода информации о ранжируемых объектах

Посредством интерактивного поиска ограниченного перебора был определен ранжирующий алгоритм.

1. Для трех групп важности ранжирующий алгоритм был найден с погрешностью, равной 6 (максимальная погрешность 40), и получены следующие результаты:

- к третьей группе важности было отнесено 6 критериев: Marking, Tackling, Technique, Teamwork, Balance, Strength;
- ко второй группе – 13 критериев: Crossing, Dribbling, Finishing, Penalty Taking, Set Pieces, Bravery, Creativity, Decisions, Influence, Acceleration, Jumping, Pace, Stamina;

- к первой – 12: Heading, Long Shots, Long Throws, Passing, Aggression, Anticipation, Determination, Flair, Off the Ball, Positioning, Workrate, Agility.



Рис. 2. Окно ввода информации о частных критериях ранжирования



Рис. 3. Окно просмотра базы данных

2. Для двух групп важности ранжирующий алгоритм был найден с погрешностью, равной 16 (максимальная погрешность 40), и получены следующие результаты:

- ко второй группе был отнесен 21 критерий: Crossing, Dribbling, Long Shots, Long Throws, Marking, Passing, Set Pieces, Technique, Aggression, Bravery, Decisions, Influence, Positioning, Teamwork, Workrate, Acceleration, Agility, Balance, Jumping, Pace, Strength;
- к первой группе было отнесено 10 критериев: Finishing, Heading, Penalty Taking, Tackling, Anticipation, Creativity, Determination, Flair, Off the Ball, Stamina.

Отсюда видно, что при увеличении групп важности погрешность определения ранжирующего алгоритма заметно уменьшается.

На основе идентификации ранжирующих алгоритмов в форме метода ПРИНН-У ИС был определен ранжирующий алгоритм с нулевой погрешностью (идеальный ранжирующий алгоритм) и получены следующие результаты:

- к третьей группе важности было отнесено 6 критериев: Heading, Marking, Decisions, Acceleration, Balance, Strength;
- ко второй группе – 8 критериев: Crossing, Dribbling, Long Shots, Long Throws, Tackling, Bravery, Creativity, Pace;
- к первой – 17: Finishing, Passing, Penalty Taking, Set Pieces, Technique, Aggression, Anticipation, Determination, Flair, Influence, Off the Ball, Positioning, Teamwork, Workrate, Agility, Jumping, Stamina.

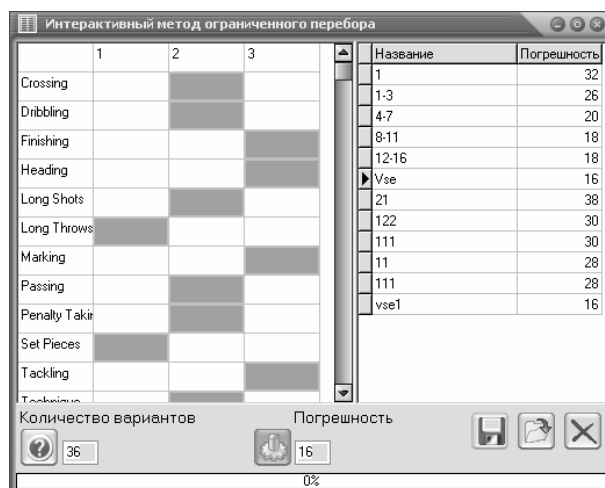


Рис. 4. Окно установки границ перебора и выдачи результатов

Футболист	Суммарный балл
Павел Недвед	1418
Тьерри Анри	1375
Паоло Мальдини	1375
Андрей Шевченко	1371
Зинедин Зидан	1371
Руд ван Нистелрой	1369
Рауль Гонсалес	1367
Роберто Карлос	1361
Девид Бекхем	1290

Таблица 2. Итоговый рейтинг и суммарный балл футболистов

На основе метода идентификации ранжирующих алгоритмов в форме линейной свертки программа определила ранжирующий алгоритм с нулевой погрешностью (был найден идеальный ранжирующий алгоритм). В табл. 3 приведены результаты.

Итак, мы получили два ранжирующих алгоритма:

- методом идентификации ранжирующих алгоритмов в форме линейной свертки;
- методом идентификации ранжирующих алгоритмов в форме метода ПРИНН-У.

Для анализа результатов были выбраны наиболее важные критерии, полученные первым методом (Aggression, Flair, Balance, Jumping), и критерии, отнесенные к последней группе важности (наиболее важной) вторым методом (Decisions, Acceleration, Balance, Strength). На основе этих методов и показателей были найдены ранжирующие алгоритмы для тех же 9 футболистов и получены результаты, приведенные в табл. 6.

Критерий	Коэффициент	Критерий	Коэффициент	Критерий	Коэффициент
----------	-------------	----------	-------------	----------	-------------

	значимости		значимости		значимости
Crossing	0,081	Stamina	0	Flair	0,178
Dribbling	0,005	Strength	0,051	Influence	0,037
Finishing	0	Set Pieces	0	Off the Ball	0
Heading	0	Tackling	0,031	Positioning	0,024
Long Shots	0	Technique	0,008	Teamwork	0,007
Long Throws	0,091	Aggression	0,132	Workrate	0
Marking	0,01	Anticipation	0,003	Acceleration	0
Passing	0,006	Bravery	0	Agility	0
Penalty Taking	0	Creativity	0	Balance	0,207
Jumping	0,103	Decisions	0,019		
Pace	0,002	Determination	0,003		

Таблица 3. Полученные коэффициенты значимости критериев

Футболист	Суммарный балл
Павел Недвед	14,78
Тьерри Анри	14,34
Паоло Мальдини	13,43
Андрей Шевченко	13,35
Зинедин Зидан	13,34
Руд ван Нистелрой	13,27
Рауль Гонсалес	13,27
Роберто Карлос	13,27
Девид Бекхем	13,21

Таблица 4. Итоговый рейтинг и суммарный балл футболистов

Критерий	Коэффициент значимости	Критерий	Коэффициент значимости
Aggression	0,12	Balance	0,16
Decisions	0,2	Jumping	0,17
Flair	0,16	Strength	0,17
Acceleration	0,01		

Таблица 5. Полученные коэффициенты значимости критериев

Футболист	Суммарный балл
Павел Недвед	14,3
Тьерри Анри	16,1
Паоло Мальдини	13,9
Андрей Шевченко	13,9
Зинедин Зидан	13,9
Руд ван Нистелрой	15,4
Рауль Гонсалес	13,9
Роберто Карлос	13,9
Девид Бекхем	12,9

Таблица 6. Итоговый рейтинг и суммарный балл футболистов

Методом идентификации ранжирующих алгоритмов в форме линейной свертки был получен ранжирующий алгоритм с погрешностью 2 (максимальная погрешность – 8).

Методом идентификации ранжирующих алгоритмов в форме метода ПРИНН-У был получен ранжирующий алгоритм с погрешностью 3 (максимальная погрешность – 8).

- к третьей группе важности был отнесен один критерий – Strength;
- ко второй группе – 3 критерия: Acceleration, Balance, Jumping;
- к первой – 3: Aggression, Decisions, Flair.

Футболист	Суммарный балл
Павел Недвед	292
Тьерри Анри	339
Паоло Мальдини	291
Андрей Шевченко	287
Зинедин Зидан	279
Руд ван Нистелрой	342
Рауль Гонсалес	277
Роберто Карлос	312
Девид Бекхем	258

Таблица 7. Итоговый рейтинг и суммарный балл футболистов

Ранжирование стран по степени готовности к внедрению электронного обучения

Другим примером применения предложенного подхода является восстановление алгоритма ранжирования стран по степени готовности к внедрению электронного обучения (e-learning). Для примера было выбрано 15 стран (табл. 8) из рейтинга по оценкам готовности стран к электронному обучению в 2003 году, составленным организацией The Economist Intelligence Unit в сотрудничестве с корпорацией IBM [5]. В качестве критериев были приняты индикаторы, представленные в [6] (табл. 9).

Страна	Рейтинг по [5]
Швеция	1
США	2
Финляндия	3
Дания	4
Великобритания	5
Ирландия	6
Голландия	7
Франция	8
Австрия	9
Германия	10
Бельгия	11
Италия	12
Испания	13
Греция	14
Португалия	15

Таблица 8. Рейтинг по оценкам готовности стран к электронному обучению 2003 года

На основе идентификации ранжирующих алгоритмов в форме метода ПРИНН-У ИС был определен ранжирующий алгоритм с погрешностью 4 (максимальная погрешность – 15) и получены следующие результаты:

- к третьей группе важности было отнесено 5 критериев; DIDIX, государственные расходы на НИР, общие расходы на НИР, дистанционный труд, электронная коммерция;
- ко второй группе – 2 критерия: пользователи Интернета, компьютеры, имеющие доступ в Интернет;
- к первой – 3: DSL выделенные линии, электронное обучение, COQS.

Критерий
Пользователи Интернетом
DSL Выделенные линии
Компьютеры, имеющие доступ в Интернет
Электронное обучение
Индекс цифровой грамотности COQS
Коэффициент неравенства DIDIX
Государственные расходы на НИР
Общие расходы на НИР
Дистанционный труд
Электронная коммерция

Таблица 9. Индикаторы и критерии эффективности образовательных систем

Страны	Суммарный балл
Швеция	1497
США	1480
Финляндия	1416
Дания	1465
Великобритания	1368
Ирландия	1110
Голландия	1281
Франция	842,9
Австрия	1281
Германия	1177
Бельгия	835,6
Италия	755,7
Испания	715,3
Греция	640,2
Португалия	574,9

Таблица 10. Итоговый рейтинг и суммарный балл стран

Критерий	Коэффициент значимости
Пользователи Интернетом	0,114
DSL Выделенные линии	0
Компьютеры имеющие доступ в Интернет	0,005
Электронное обучение	0
Индекс цифровой грамотности COQS	0
Коэффициент неравенства DIDIX	0,016
Государственные расходы на НИР	0,804
Общие расходы на НИР	0,044
Дистанционный труд	0,018
Электронная коммерция	0

Таблица 11. Полученные коэффициенты значимости критериев

Страны	Суммарный балл
Швеция	3,147
США	3,146
Финляндия	2,930
Дания	2,707
Великобритания	2,582
Ирландия	1,963
Голландия	2,144
Франция	1,963
Австрия	2,702
Германия	2,247
Бельгия	1,961
Италия	1,554
Испания	1,552
Греция	1,415
Португалия	1,412

Таблица 12. Итоговый рейтинг и суммарный балл стран

На основе метода идентификации ранжирующих алгоритмов в форме линейной свертки был определен ранжирующий алгоритм с погрешностью, равной 3 (максимальная погрешность – 15). В табл. 11 приведены результаты.

Выделим наиболее важные критерии из решений, полученных обоими методами (табл. 13).

Критерий	Коэффициент значимости методом ЛС	Коэффициент значимости методом ПРИНН-У
Государственные расходы на НИР	0,804	9
Пользователи Интернетом	0,114	3
Общие расходы на НИР	0,044	9
Дистанционный труд	0,018	9
Коэффициент неравенства DIDIX	0,016	9
Компьютеры, имеющие доступ в Интернет	0,005	3
DSL Выделенные линии	0	1
Электронное обучение	0	1
Индекс цифровой грамотности COQS	0	1
Электронная коммерция	0	9

Таблица 13. Наиболее важные критерии из решений, полученных обоими методами

Из табл. 13 видно что коэффициенты значимости, найденные разными методами, практически совпадает, что говорит о правильности и точности найденного решения.

Обоими методами нам не удалось найти решение с нулевой погрешностью, но погрешность определяется только входными данными, поэтому следует провести полный функциональный анализ тех стран, которые не подчинились найденному алгоритму, что приведет к увеличению числа критериев.

Заключение

- Исследованы методы идентификации (восстановления) ранжирующих алгоритмов.
- Показана возможность использования методов.

- Разработана универсальная информационная система.
- На примере определены ранжирующие алгоритмы.
- Проведен сравнительный анализ методов идентификации (восстановление) ранжирующих алгоритмов на примере применения.

Литература

1. Пиявский С.А. Методы оптимизации и оптимального управления: Учебное пособие. СГАСУ. Самара, 2004. 160 с.
2. Ларичев О. И. Теория и методы принятия решений. М.: Логос, 2000.
3. Ларичев О. И. Методы поддержки принятия решений. М.: УРСС, 2001.
4. Sports Interactive «Championship Manager», 2004.
5. Рейтинг готовности стран к электронному обучению. // Learning World. 2004. №5–6. С. 30–32.
6. Advancement of the knowledge society: Comparing Europe, the US and Japan, Wyattville Road, Loughlinstown, Dublin 18, Ireland.
<http://www.eurofound.eu.int/publications/files/EF0403EN.pdf>