А. А. Заславский

Теорема Эрроу

Алексей Александрович Заславский - старший научный сотрудник
Центрального экономико-математического института РАН,
учитель московской школы №1543

VIVOS VOCO

Образовательный сетевой выпуск
VIVOS VOCO! - ЗОВУ ЖИВЫХ!
http://www.accessnet.ru/vivovoco

ТЕОРЕМА ЭРРОУ

А. Заславский

Рассмотрим следующую задачу. Есть n объектов и N субъектов, оценивающих эти объекты. Результатом оценки является ранжировка, т. е. упорядоченный список объектов. На первом месте в списке стоит объект, который данный субъект считает наилучшим, на втором — следующий по предпочтительности,…, на n-м месте — наихудший из объектов. Нужно сформулировать правило, которое по N произвольным ранжировкам строит одну ранжировку, в наибольшей степени отражающую мнение всех N субъектов.

Прежде чем придать этой задаче строгую математическую формулировку, скажу несколько слов о ее практических применениях. С одной стороны, успешное решение задачи помогло бы сформировать идеальную избирательную систему (то, что существующие в настоящее время избирательные системы идеальными не являются, следует уже из того, что их много; впрочем, ниже недостатки некоторых избирательных систем будут рассмотрены подробно).

Другой областью применения данной задачи является многокритериальная оптимизация, т. е. выбор наилучшего из объектов, оцениваемых по различным критериям. Например, при обмене или покупке квартиры необходимо учитывать такие критерии, как местоположение квартиры, площадь, планировку, потребность в ремонте, стоимость и тому подобное. Каждый критерий как-то ранжирует имеющиеся альтернативы, и на основе этих частных ранжировок нужно построить одну итоговую. Разумеется, критерии не всегда являются равнозначными, тем не менее, хотелось бы при принятии решений учитывать их все.

Попробуем теперь выяснить, какими свойствами должно обладать решающее правило. Для этого выясним прежде всего, почему нас не устраивают правила, применяемые на практике.

Наиболее распространенным среди таких правил является правило простого большинства, применяемое, например, в большинстве стран для выбора депутатов парламента. Оно заключается в том, что каждый из голосующих называет наилучший с его точки зрения объект, и объект, набравший наибольшее число голосов, считается победителем.

Чтобы понять недостатки этого подхода, рассмотрим следующую ситуацию. Пусть оцениваются три объекта: A, B и C, причем среди голосующих 40% ранжируют их как (A, B, C), 30% — (B, C, A), 30% — (C, B, A). Тогда побеждает объект A, к которому 60% опрошенных отнеслись явно отрицательно. Впрочем, самое плохое не это, а то, что если исключить из голосования любой из объектов B или C, вроде бы не имеющих шансов на победу, то результат голосования оказывается другим! Таким образом, включая в список для голосования заведомо «непроходные» кандидатуры, можно влиять на результат.[1]

[1]Эта ситуация описана, например, в рассказе Г. К. Честертона «Белая ворона».

Другое часто применяемое правило состоит в том, что если при первом голосовании никто не получил более 50% голосов, то два лидера выходят во второй тур, победитель которого уже определяется простым большинством. Для анализа этого правила рассмотрим такую ситуацию. Пусть опять оценивают три объекта, а мнения голосующих распределены так: (A, B, C) — 40%, (B, C, A) — 29%, (C, A, B) — 31%. Тогда во второй тур выходят A и C, прич C побеждает, получив во втором туре 60% голосов. Однако если исключить из списка для голосования объект A, то победителем становится B!

Анализ этих двух примеров позволяет сформулировать первое требование к решающему правилу.

1) *Аксиома независимости.* Результат сравнения между двумя объектами A и B должен зависеть только от результатов сравнения их в индивидуальных ранжировках, но не от сравнений их с другими объектами.

Два следующих требования являются более очевидными и специальных пояснений не требуют.

2) *Аксиома монотонности.* Если один из голосующих изменил свое мнение в пользу объекта A, а мнения остальных не изменились, то в итоговой ранжировке положение объекта A не может ухудшиться.

3) *Аксиома ненавязанности.* Для любой данной ранжировки существует такой набор индивидуальных ранжировок, что итоговая ранжировка совпадает с данной.

Из аксиом 2 и 3 вытекает, в частности, что если все индивидуальные ранжировки совпадают, то и итоговая совпадает с ними.

Итак, нам предстоит найти решающее правило, удовлетворяющее аксиомам 1–3. Прежде всего, отметим, что такие правила существуют. Например, можно выделить одного из голосующих и всегда выбирать его ранжировку в качестве итоговой. Такие правила естественно называть диктаторскими. Я докажу поразительный результат:

Теорема. *Любое решающее правило, удовлетворяющее аксиомам 1–3, является диктаторским.*

Доказательство. Пусть некоторое подмножество I множества субъектов обладает следующим свойством: если каждый субъект из I предпочитает объект A объекту B (в дальнейшем будем писать $A > B$), то и в итоговой ранжировке $A > B$. Тогда множество I назовем утверждающим для пары A, B.

Определение утверждающего множества корректно в силу аксиомы 1. При этом очевидно, что утверждающие множества существуют. Например, множество всех объектов является утверждающим для любой пары.

Исследуем свойства утверждающих множеств.

Лемма 1. Если множество I утверждающее для некоторой пары, то оно утверждающее для любой пары.

Доказательство. Пусть множество I утверждающее для пары (C, D). Докажем сначала, что оно утверждающее для пары (C, B) при любом B. Положим в индивидуальных ранжировках $C > D > B$ для всех субъектов из I, $D > B > C$ для всех субъектов, не входящих в I. Тогда в итоговой ранжировке $C > D$, так как I — утверждающее множество для пары (C, D), $D > B$ в силу аксиом 2, 3. Следовательно, $C > B$. Но для всех субъектов, не входящих в I, $B > C$, т. е. I — утверждающее множество для пары (C, B). Аналогично можно доказать, что если I — утверждающее множество для пары (C, B), то оно утверждающее и для пары (A, B) при любом A, что и доказывает лемму 1.

Согласно лемме 1 можно говорить просто об утверждающих множествах, не называя пары объектов.

Лемма 2. Пересечение двух утверждающих множеств — утверждающее множество.

Доказательство. Пусть I, J — два утверждающих множества. Положим для субъектов, входящих и в I, и в J $A > B > C$, для субъектов, входящих только в I $C > A > B$, входящих только в J $B > C > A$, и не входящих ни в I, ни в J $C > A > B$. Тогда в итоговой ранжировке A¿B, так как I утверждающее множество, $B > C$, так как J утверждающее множество, и значит, $A > C$. Но $A > C$ только для субъектов, входящих в пересечение I и J. Лемма доказана.

Из леммы 2 следует, что существует минимальное утверждающее множество, являющееся пересечением всех утверждающих множеств. Из аксиом 2 и 3 вытекает, что это множество не пусто.

Лемма 3. Минимальное утверждающее множество содержит один элемент.

Доказательство. Пусть элемент j принадлежит минимальному утверждающему множеству V. Положим для j $A > B > C$, для остальных субъектов из V $B > C > A$, для субъектов, не входящих в V $C > A > B$. Тогда в итоговой ранжировке $B > C$, так как V утверждающее множество, $A > B$, так как в противном случае оказалось бы, что множество $V \setminus \{j\}$ утверждающее, что противоречит минимальности V. Следовательно, $A > C$. Но для всех субъектов, кроме j, $C > A$. Лемма доказана.

Таким образом, существует такой субъект j, что для любых двух объектов их относительное расположение в итоговой ранжировке такое же, как и в ранжировке j. Следовательно, итоговая ранжировка совпадает с ранжировкой j. Теорема доказана.

С точки зрения построения идеальной избирательной системы получен весьма обескураживающий результат. Фактически он означает, что демократия является скрытой формой диктатуры.

Больший практический интерес представляет применение данного результата к задачам многокритериального выбора. Из теоремы следует, что для того чтобы делать непротиворечивый выбор, необходимо выбрать какой-нибудь один критерий и руководствоваться только им. Впрочем, есть возможность учитывать остальные критерии, превращая их в ограничения. Например, решающее правило для выбора квартиры может выглядеть так: выбрать квартиру наибольшей площади среди расположенных в определенном районе, не требующих капитального ремонта и стоящих не дороже заданной суммы.

Сделаем теперь небольшой исторический обзор. Доказанную выше теорему часто называют теоремой Эрроу по имени известного американского математика, лауреата Нобелевской премии по экономике. На самом деле, результат, полученный Эрроу, является более общим. Он относится к агрегированию не только ранжировок, но и произвольных транзитивных отношений.

Предположим, например, что наша цель не ранжировать объекты, а разбить их на некоторое, не заданное заранее число классов. В этом случае все решающие правила, удовлетворяющие требованиям, аналогичным аксиомам 1–3, имеют следующий вид: выберем некоторое подмножество I множества субъектов и будем включать объекты A и B в один класс тогда и только тогда, когда их включают в один класс все субъекты из I. Для доказательства достаточно незначительно изменить доказательства лемм 1 и 2. Подробнее об этом можно прочитать в [1]. Что касается доказанного частного случая, то его вполне можно связать с именем знаменитого французского философа 18 века Кондорсе, который первым сформулировал парадокс возникновения нетранзитивности при применении правила большинства. Например, если три субъекта дают следующие ранжировки трех объектов: $A > B > C$, $B > C > A$, $C > A > B$, то применение правила большинства дает $A > B$, $B > C$, но $C > A$. Нетрудно заметить, что именно парадокс Кондорсе был основным инструментом доказательства теоремы. Интересно отметить, что когда во Франции произошла революция и установилась демократия, Кондорсе был казнен.

Ситуации, схожие с парадоксом Кондорсе, возникают в различных игровых задачах. Вот несколько примеров.

На железной дороги, имеющей форму окружности, в вершинах правильного треугольника расположены станции A, B, C. Ездящий по дороге единственный поезд проходит их именно в таком порядке. Два пассажира одновременно входят на две разные станции, и тот, кто первым входит в поезд, выигрывает. Очевидно, что выгоднее входить на станции A, чем на станции B, на станции B, чем C, на станции C, чем A.

Три команды состоят из 3 теннисистов каждая. Силы теннисистов в команде A равны 1, 6 и 8; в команде B — 3, 5 и 7; в команде C — 2, 4 и 9. Матч между двумя командами состоит из 9 поединков каждого игрока первой команды с каждым игроком второй. Если более сильный игрок всегда выигрывает у более слабого, то A побеждает B, B — C и C — A со счетом 5:4.

Если 3 раза бросить монету, то возможны 8 различных серий — ООО, ООР, ..., РРР. Два игрока выбирают по одной серии каждый, и монета бросается до тех пор, пока в 3 последовательных бросках не выпадет одна из выбранных серий. Игрок, чья серия выпала, выигрывает. Оказывается, что какую бы серию ни выбрал первый игрок, второй всегда может обеспечить себе не меньше 60% шансов на выигрыш. Более подробно эта игра разобрана в [2].

После приведенных примеров уже не должна вызывать удивления нетранзитивность, например, реальных спортивных турниров. Однако, возникает вопрос, нельзя ли эту нетранзитивность измерить. Покажем, как можно решить эту задачу.

Рассмотрим сначала турнир без ничьих, в котором за победу присуждается 1 очко, а за поражение — 0 очков. Естественно определить степень его нетранзитивности, как число нетранзитивных троек. Однако, число различных троек быстро возрастает с ростом числа участников турнира (при n участниках оно равно $n(n-1)(n-2)/6$), поэтому возникает вопрос, существует ли более простой способ вычисления нетранзитивности. Ответ дает следующая формула

$$NT = n(n-1)(2n-1)/12 - \sum s_i^2/2 \qquad (1)$$

где NT — число нетранзитивных троек n — число участников турнира s_i — количество очков, набранных i-ым участником.

Доказательство. Пусть все партии кроме партии участников с номерами 1 и 2 сыграны. Очевидно, что результат последней партии влияет на транзитивность только тех троек, в которые входят оба ее участника. Пусть в партиях с остальными участниками 1 и 2 набрали соответственно s_1 и s_2 очков, причем число участников, у которых они оба выиграли равно x_{11}, число участников, у которых 1-ый выиграл, а 2-ой проиграл — x_{12}, 2-ой выиграл, а 1-ый проиграл — x_{21}, оба проиграли — x_{22}. Ясно, что выполняются соотношения:

$$x_{11} + x_{12} = s_1$$

$x_{11} + x_{21} = s_2$

$x_{11} + x_{12} + x_{21} + x_{22} = n - 2$

Эти соотношения показывают, что если s_1, s_2 известны, то, зная одно из чисел x_{11}, x_{12}, x_{21}, x_{22}, можно определить остальные. В частности $x_{12} = n-2-s_2-x_{22}$, $x_{21} = n-2-s_1-x_{22}$.

Рассмотрим теперь различные исходы последней партии.

1) Победа игрока 1. В этом случае 1-й набирает $s_1 + 1$ очко, у 2-го остается s_2. Правая часть формулы (1) принимает вид:

$$C_1 - (s_1+1)^2/2 - s_2^2/2$$

где C_1 — полусумма квадратов очков остальных участников, не зависящая от исхода последней партии.

Тройка с участием 1-го и 2-го будет нетранзитивной тогда и только тогда, когда третий ее участник выигрывает у 1-го и проигрывает 2-му. Поскольку число таких троек равно x_{21}, левую часть формулы можно записать в виде $C_2 + x_{21}$, где C_2 не зависит от исхода партии.

2) Победа игрока 2. Рассуждая аналогично, получаем для правой части (1)

$$C_1 - (s_2+1)^2/2 - s_1^2/2,$$

а для левой $C_2 + x_12$.

Используя найденные ранее выражения для x_{12}, x_{21}, получаем, что разность левой и правой части в обоих случаях равна:

$$C_2 - C_1 + s_1^2/2 + s_2^2/2 + 1/2 + n - 2 - x_{22}$$

Таким образом эта разность не зависит от результата произвольной партии, и, значит, одинакова для всех турниров. Рассмотрим теперь турнир, в котором первый участник выиграл у всех остальных, второй у всех кроме первого и т.д. Для этого турнира число нетранзитивных троек равно нулю, а количества очков $s_1 = n-1$, $s_2 = n-2$, $\ldots$,$s_n = 0$. Используя формулу суммы квадратов натурального ряда, убеждаемся, что формула (1) для этого турнира верна. Следовательно она верна для всех турниров.

Из формулы (1) следует, что максимальное число нетранзитивных троек равно $(n^3-n)/24$ при нечетном n и $(n^3-4n)/24$ при четном.

Перейдем теперь к турнирам с ничьими. Прежде всего заметим, что знать количество очков, набранных каждым участником, явно недостаточно. Действительно, рассмотрим два турнира с $n = 3$ и такими таблицами:

-	1	0		-	1/2	1/2
0	-	1		1/2	-	1/2
1	0	-		1/2	1/2	-

Количества набранных участниками очков в этих турнирах совпадают, но в первой тройке транзитивность нарушена, а во второй нет.

Поэтому будем считать, что для каждого участника нам известно не только количество набранных очков, но и количество выигрышей, ничьих и поражений, соответственно v_i, n_i, p_i. Разумеется, для любого i — $v_i + n_i + p_i = n - 1$.

Теперь необходимо уточнить понятие степени нелогичности. Действительно, в отличие от турнира без ничьих в турнире с ничьими возможны три типа нетранзитивных троек, а именно:

-	1	0
0	-	1
1	0	-

-	1	1/2
0	-	1
1/2	0	-

-	1/2	1
1/2	-	1/2
0	1/2	-

Считать вклад всех этих троек в общую нелогичность одинаковым вряд ли разумно. Первая тройка представляется наиболее нетранзитивной, а третья — наименее. Поэтому имеет смысл приписать тройкам первого типа вес 1, второго x, третьего y, где $y \leqslant x \leqslant 1$. Выбор этих весов может диктоваться различными соображениями, но оказывается, что степень нелогичности выражается через v_i, n_i, p_i лишь в одном случае. Чтобы убедиться в этом, рассмотрим две пары турниров:

1)

-	1/2	1/2	1/2	1
1/2	-	1	0	1/2
1/2	0	-	1	1/2
1/2	1	0	-	1/2
0	1/2	1/2	1/2	-

-	1	1/2	1/2	1/2
0	-	1	1/2	1/2
1/2	0	-	1	1/2
1/2	1/2	0	-	1
1/2	1/2	1/2	0	-

2)

-	1/2	1/2	1	1
1/2	-	1	0	1
1/2	0	-	1	1/2
0	1	0	-	1/2
0	0	1/2	1/2	-

-	1	1	1/2	1/2
0	-	1/2	1	1
0	1/2	-	1	1/2
1/2	0	0	-	1
1/2	0	1/2	0	-

В обеих парах числа v_i, n_i, p_i совпадают для всех i. Непосредственно перебрав все тройки и приравняв для каждой пары суммарные нетранзитивности, получим:

$1 + 9y = 3x + 6y$

$1 + 2x + 3y = 4x + 3y$

Решив эту систему, получим $x = 1/2$, $y = 1/6$.

Осталось выяснить, существует ли при таких весах формула, аналогичная (1). Ответ на этот вопрос положительный. Приведем результат:

$$NT = ((n^3 - n) - \sum n_i(n_i + 2) - 3\sum(v_i - p_i)^2)/24 \qquad (2)$$

Формула (2) доказывается аналогично (1). Подробнее об этом можно прочитать в [3].

Исследовав формулу (2), можно увидеть, что максимальная нетранзитивность для турнира без ничьих, такая же, как для турнира с ничьими. При нечетном n она достигается, если при всех i $n_i = 0$, $v_i = p_i = (n-1)/2$, при четном, если при всех i либо $n_i = 0$, $(v_i - p_i)^2 = 1$, либо $n_i = 1$, $v_i = p_i = (n-2)/2$.

В заключение несколько слов о практической значимости приведенных результатов. Конечно, измерять противоречивость спортивных турниров никому не нужно. Однако, есть область прикладной математики, реально использующая эти результаты. Это теория экспертных оценок. Во многих прикладных задачах требуется выбрать наилучший из некоторых объектов, причем выбор является неформализованным и может быть осуществлен только с помощью экспертов. Опыт показывает, что эксперту легче ответить на вопрос, какой из двух объектов является лучшим, чем сравнивать сразу несколько объектов. Поэтому, когда число сравниваемых объектов не очень велико, целесообразно пользоваться так называемым методом парных сравнений, при котором эксперту предъявляются все пары возможных объектов и его ответы заносятся в таблицу типа турнирной. При этом эксперту может разрешаться или не разрешаться объявлять объекты равноценными, что соответствует турнирам без ничьих и с ничьими. Чтобы на основании ответов эксперта можно было сделать обоснованный выбор, необходимо, чтобы они были достаточно логичны. Для проверки логичности и используется формула (1), которую можно найти в большинстве книг по методу парных сравнений. Как правило, при проведении экспертиз опрашиваются несколько экспертов, после чего определяется коллективное мнение. В методе парных сравнений эксперты, у которых число нетранзитивных троек превышает некоторое критическое значение исключаются из рассмотрения, а мнения остальных учитываются с различными весами (предполагается, что большое число нетранзитивных троек свидетельствует о невысокой компетентности эксперта, и, следовательно, его мнение имеет небольшой вес). Применяется ли для сравнений с ничьими формула (2), мне неизвестно. Впрочем, следует отметить ее искусственность, связанную с выбором весов для троек разных типов. Действительно, то что нетранзитивность тройки второго типа вдвое меньше, чем первого, представляется достаточно разумным, но почему тройка третьего типа должна иметь нетранзитивность $1/6$, остается малопонятным. Если удастся найти доводы в пользу такого веса, ценность формулы (2) значительно возрастет.

ЛИТЕРАТУРА

1. Б.Г.Миркин. Проблема группового выбора. М., Наука, 1974.
2. М.Гарднер. Путешествие во времени. М., Мир, 1990.
3. А.Заславский. О логичных и нелогичных турнирах. Квант, 1997, № 5.