

Алгоритм рейтингования при коллективных экспертизах

С. М. Абрамов, д. ф.-м. н. чл.-корр. РАН
И. Н. Григоревский, к. т. н.,
Институт программных систем им. А. К. Айламазяна РАН

Потребность учета мнений группы экспертов или просто рецензентов является актуальной проблемой в широком классе задач: начиная от простого голосования и заканчивая ранжированием научных статей или докладов участников конференций.

Чтобы судить о сложности учета индивидуальных мнений экспертов при проведении коллективной экспертизы, достаточно рассмотреть работу практически любой научной конференции, где участники делают свои доклады и по результатам докладов наиболее значимые работы должны быть отмечены какими-либо наградами. Обычно количество докладов на конференциях исчисляется десятками и рецензенты (эксперты, в нашей терминологии) физически не способны оценить и сравнить все доклады, к тому же тематика докладов бывает весьма широка, и кажется маловероятным, чтобы многие рецензенты являлись специалистами во всех темах, затронутых в докладах. В лучшем случае эксперт-рецензент может оценить 3–4 работы. Но здесь возникает другая проблема — какова шкала оценки? Один рецензент оценивает по 5-балльной системе, другой — по 3-балльной. Третий рецензент-оригинал — придумает свою собственную. Если ввести единую шкалу оценок и заставить экспертов пользоваться только ей, сравнение оценок от этого легче не станет, так

как начинает играть роль случайность полученных экспертом работ для оценки. Например, рецензенту 1 может достаться набор из слабых работ и он выберет «лучшую среди плохих» и высоко ее оценит, а рецензенту 2 попадутся сильные работы, такие что «худшая» из них будет на голову выше всех представленных на конференции. В итоге получится, что работа «худшая, среди лучших» будет оценена вторым рецензентом ниже, чем оценка «лучшей, среди худших» работ, доставшихся рецензенту номер 1. Встает вопрос — как соотносить оценки разных рецензентов друг с другом? Задача эта нетривиальна и простого решения не найдено до сих пор.

Метод, предложенный в данной работе, был сформулирован д. ф.-м. н., членом-корреспондентом РАН С. М. Абрамовым и предлагает отбросить шкалы оценок. Вместо этого экспертам предлагается сравнивать работы (назовем их «*сущностями*») попарно, оперируя понятием «*предпочтительнее*». Таким образом, от каждого эксперта мы будем получать множество работ, которые он рецензировал, упорядоченные по степени предпочтения. Правильнее даже сказать, что результатом работы экспертов будет множество пар элементов имеющегося множества, связанных оператором «предпочтительнее». Каждую такую пару будем называть «*правилом*». При этом множества у разных экспертов могут пересекаться и некоторые правила — противоречить друг другу. Суть алгоритма построения «общего» итогового мнения состоит в том, чтобы найти такой порядок сущностей, который бы нарушал наименьшее количество правил.

1. Существующие методы экспертных оценок

Методами экспертных оценок называют такие способы организации работы со специалистами-экспертами и обработки мнений экспертов, полученных в результате этой работы, которые можно выразить в количественной и/или качественной форме с целью получения коллективного экспертного мнения. Полученное коллективное мнение может быть использовано в дальнейшем процессе принятия решений.

На сегодняшний день придумано много различных методов получения экспертных оценок. В рамках некоторых из них работа с экспертами происходит раздельно, так, что эксперты не знают, кто еще является экспертом, таким образом, на их мнение не влияет авторитет других экспертов. В других методах предполагается, что эксперты будут собраны вместе и совместно подготовят материалы для лиц, принимающих решения. При этом происходит обмен мнениями, обсуждение и даже обучение. Некоторые методы подразумевают фиксированное число экспертов так, чтобы статистические методы проверки согласованности мнений и процедуры их усреднения позволяли принимать обоснованные решения. В дру-

гих — число экспертов может расти в процессе проведения экспертизы, как, например, в методе «снежного кома» [1].

На данный момент не разработана научно обоснованная классификация методов экспертных оценок, а тем более каких-либо однозначных рекомендаций по их применению. Большая часть методов, используемых в наше время, базируется на работах, опубликованных в 70–80-х годах прошлого века [2–6].

2. Постановка задачи

Прежде, чем описывать математическую постановку задачи консолидации мнения экспертов, введем несколько определений.

Проблема — объект, о котором требуется получить консолидированное мнение экспертов. Это может быть какая-то разновидность голосования или ранжирование тем НИР с целью определения приоритетов в финансировании, или ранжирование статей с целью составить сборник из самых лучших и т. п.

Отношение предпочтения — отношение вида « A предпочтительнее B ». Обозначать предпочтение будем знаком « $>$ ». Таким образом, выражение « A предпочтительнее B » будем записывать как « $A > B$ ».

Сущность — нечто, непосредственно оцениваемое экспертом. Оценка состоит в том, что эксперт сравнивает две сущности и выставляет отношение предпочтения между ними. Таким образом, для того чтобы оценка стала возможной, необходимо по меньшей мере две сущности. Примерами сущностей могут быть ответы на некий поставленный вопрос (тогда оценка эксперта может выглядеть как «*Да* $>$ *Нет*») или же, в случае сравнения статей, сами статьи (тогда оценка будет вроде «*Статья № 1* $>$ *Статья № 2*»).

Правило — выставленное экспертом отношение предпочтения. То есть это то, что мы называем оценкой эксперта (выражение типа « $A > B$ »). С точки зрения самого эксперта, это его оценка, с точки зрения же системы — мы будем говорить о правиле, а собрав все оценки всех экспертов, получим множество правил.

В общем случае, проблема состоит из описательной части и множества сущностей.

В процессе своей работы эксперты сравнивают сущности и выдают свои оценки, формируя таким образом множество правил.

Наконец, система ранжирует сущности таким образом, чтобы наименьшее количество правил оказалось нарушено.

В результате получаем отранжированные сущности и считаем, что задача получения консолидированного мнения экспертов решена.

То есть суть учета мнений всех экспертов заключается в том, чтобы найти такую итоговую последовательность сущностей, которая нарушает наименьшее количество правил.

Рассмотрим примеры, чтобы показать принцип, по которому происходит формализация задач.

Пример 1. Выбор одного из кандидатов.

Описательная часть (вопрос экспертам): Кого из кандидатов Вы хотели бы видеть в качестве мэра Цветочного города?

Сущности (кандидаты в мэры): Знайка, Незнайка, Пончик, Винтик, Шпунтик, Пилюлькин.

Оценки экспертов (правила):

Эксперт 1: Знайка > Пончик, Пилюлькин > Шпунтик.

Эксперт 2: Пончик > Знайка, Незнайка > Пончик.

Обратите внимание, что при сравнении Знайки и Пончика эксперты противоречат друг другу, Винтик вообще ни разу не упоминается. Теперь получим консолидированное мнение экспертов:

Если мэром станет	Нарушено правил	Выполнено правил
Знайка	1	1
Незнайка	0	1
Пончик	1	1
Винтик	0	0
Шпунтик	1	0
Пилюлькин	0	1

Лучший результат (одно правило выполнено и ни одно не нарушено) показали Незнайка и Пилюлькин. Это и есть консолидированное мнение экспертов.

Пример 2. Вопрос типа «Да»/«Нет».

Описательная часть (вопрос экспертам): Правомерны ли действия гражданки Сова по присвоению хвоста гражданина Иа-Иа?

Сущности (варианты ответов): Да, Нет.

Оценки экспертов (правила):

Эксперт 1: Да > Нет.

Эксперт 2: Нет > Да.

Теперь получим консолидированное мнение экспертов:

Действия гражданки Сова	Нарушено правил	Выполнено правил
Правомерны (Да)	1	1
Неправомерны (Нет)	1	1

Лучший результат в данном случае выделить не удалось, так как мнения экспертов противоположны.

Пример 3. Ранжирование статей.

Описательная часть (вопрос экспертам): Каким статьям и авторам Вы отдадите предпочтения?

Сущности (статьи):

1. «Пчеловодство, или Как не слипнуться с дуба?», автор Винни П.;

2. «Основные правила метания бисера», авторы: Пятачок, Посторонним В.;
3. «Современная философия жизни», автор Иа-Иа.

Оценки экспертов (правила):

Эксперт 1: Статья 3 > 2 > 1, 1 > 3.

Эксперт 2: Статья 2 > 3.

Автор лучшей статьи	Нарушено правил	Выполнено правил
Винни П.	1	1
Пятачок, Посторонним В.	1	2
Иа-Иа	2	1

То что правила, полученные в результате работы, могут противоречить друг другу — вполне нормально и адекватно должно обрабатываться механизмом учета. В данном примере, статья «Основные правила метания бисера», по консолитированному мнению экспертов, признана лучшей.

3. Компьютерная реализация алгоритма

Описанный выше алгоритм может быть легко реализован в виде программного обеспечения. Это особенно актуально для глобальных экспертных сетевых сообществ, где число оценок может быть велико. Архитектурно база данных для такой системы будет состоять из нескольких таблиц:

Таблица «Вопросы»

для хранения перечня вопросов, содержащая

- поле идентификатора вопроса,
- текст со смысловым описанием вопроса,
- текстовое название вопроса для отображения в списках и заголовках;

Таблица «Сущности»

для хранения списка сущностей, содержащая

- поле идентификатора сущности,
- ссылку на идентификатор вопроса,
- текст с описанием сущности,
- текстовое название сущности для отображения в списках и заголовках;

Таблица «Правила»

для хранения сырых результатов-правил, содержащая

- поле идентификатора правила,
- ссылку на идентификатор эксперта,
- ссылку на идентификатор вопроса,
- ссылку на идентификатор сущности, которой было отдано предпочтение в данном правиле,

- ссылку на идентификатор сущности, которой НЕ было отдано предпочтение в данном правиле;

Таблица «Результаты»

для хранения уже обработанных результатов, содержащая

- поле идентификатора результата,
- ссылку на идентификатор вопроса,
- ссылку на идентификатор сущности,
- порядковый номер сущности в общем списке сущностей, ранжированных с учетом мнения всех экспертов, чьи правила были обработаны на данный момент,
- весовой показатель результата, позволяющий судить о том, насколько близко сущности отстоят друг от друга в общем ранжированном списке.

Наполнение указанных таблиц данными может быть реализовано разными способами: напрямую через интерфейсы СУБД; через создание сохраненных процедур, которые в свою очередь будут использоваться создателями интерфейса; через создание веб-сервисов.

Схема взаимодействия с интерфейсом представлена на рис. 11.1.

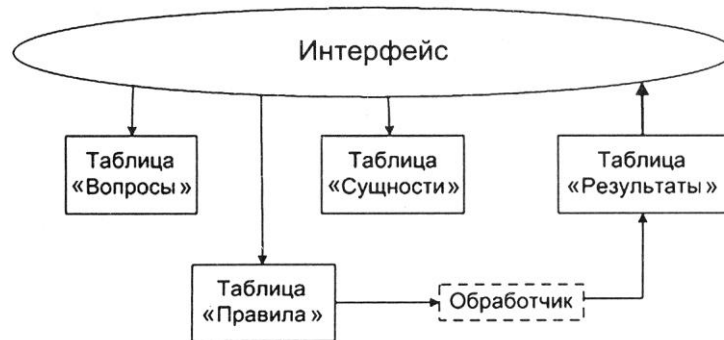


Рис. 11.1. Схема взаимодействия с интерфейсом

Стоит отметить, что инструмент разработки обработчика во многом зависит от среды, в которой предстоит работать системе. То есть от типа платформы, от выбранной операционной системы и т. д.

Необходимо отметить, что в обработчике планируется поддержка добавления и учета новых правил без полного перерасчета уже обработанных.

В качестве дальнейшего развития в функционал системы можно добавить некоторые характеристики экспертов:

- Вес эксперта — уровень доверия, показатель опытности и стажа в качестве эксперта;

- Оценка эксперта — характеристика, позволяющая судить о том, насколько правила, сформулированные данным экспертом, соотносятся с результирующей ранжированной последовательностью сущностей.

Более того, обычную систему оценок также можно модифицировать, чтобы она удовлетворяла описанному алгоритму. Для этого достаточно из оценок, поставленных одним преподавателем, формировать «правила».

Например, если эксперт поставил Знайке оценку 5, а Незнайке — оценку 2, это означает, что эксперт оценил знания Знайки выше, чем Незнайки. Именно это неабсолютное утверждение и будет правилом. Главное, чтобы не сравнивались абсолютные оценки от разных экспертов.

На самом деле любой экзаменатор, ставя абсолютную оценку, «мысленно» сравнивает ее с оценками, поставленными другим экзаменуемым, или с некоторым эталонным знанием. Используя абсолютные оценки по схеме создания «правил», можно учесть также и «время жизни» оценки, не создавая правила с «устаревшими» результатами. Даже у одного эксперта шкала оценок претерпевает изменения со временем.

Стоит также отметить, что в приведенных примерах рассматривается сравнительная оценка работ экспертами только по одному критерию. В рамках данной работы не предполагается создание многокритериальной системы учета экспертных оценок, хотя в принципе такая система может быть реализована.

4. Заключение

Предложенный к реализации механизм учета мнений экспертов для проведения коллективных экспертиз позволит получить близкий к «идеальной середине» объективный результат с учетом мнений всех участвующих экспертов.

Литература

1. Орлов А. И. Допустимые средние в некоторых задачах экспертных оценок и агрегирования показателей качества // Многомерный статистический анализ в социально-экономических исследованиях. М.: Наука, 1974. С. 388–393.
2. Панкова Л. А., Петровский А. М., Шнейдерман М. В. Организация экспертиз и анализ экспертной информации. М.: Наука, 1984. 120 с.
3. Китаев Н. Н. Групповые экспертные оценки. М.: Знание, 1975. 64 с.
4. Статистические методы анализа экспертных оценок. М.: Наука, 1977. 384 с.
5. Экспертные оценки // Вопросы кибернетики. Вып. 58. М.: Научный совет АН СССР по комплексной проблеме «Кибернетика», 1979. 200 с.
6. Экспертные оценки в системных исследованиях // Сборник трудов. Вып. 4. М.: ВНИИСИ, 1979. 120 с.