

Данная работа выполнена в рамках более широкого сравнительного исследования восприятия звуков человеком и животными.

Литература

- Александров Ю. И. Системная психофизиология // Психофизиология / Под ред. Ю. И. Александрова. СПб.: Питер, 2006. С. 252–309.
- Александров Ю. И. Предисловие // В. Б. Швырков. Введение в объективную психологию. Нейрональные основы психики / Под ред. Ю. И. Александрова. М.: Изд-во ИП РАН, 1995. С. 7–12.
- Гаврилов В. В. Соотношение ЭЭГ и импульсной активности нейронов в поведении у кролика // ЭЭГ и нейрональная активность в психофизиологических исследованиях / Отв. ред.: В. Б. Швырков, В. М. Русалов, Д. Г. Шевченко. М.: Наука, 1987. С. 33–44.
- Максимова Н. Е., Александров И. О. Типология медленных потенциалов мозга, нейрональная активность и динамика системной организации поведения // ЭЭГ и нейрональная активность в психофизиологических исследованиях / Отв. ред.: В. Б. Швырков, В. М. Русалов, Д. Г. Шевченко. М.: Наука, 1987. С. 44–72.
- Харитонов А. Н. Палеогенез акустических структур: Дис. ... канд. психол. наук. М.: Изд-во ИП РАН, 2004.
- Швырков В. Б. Что такое ЭЭГ и нейрональная активность с системно-эволюционного подхода // ЭЭГ и нейрональная активность в психофизиологических исследованиях / Отв. ред.: В. Б. Швырков, В. М. Русалов, Д. Г. Шевченко. М.: Наука, 1987. С. 5–23.
- Швырков В. Б. Введение в объективную психологию. Нейрональные основы психики // Избранные труды / Под ред. Ю. И. Александрова. М.: Изд-во ИП РАН, 2006.
- Шевченко Д. Г., Александров Ю. И., Гаврилов В. В., Горкин А. Г., Гринченко Ю. В., Греченко Т. Н., Александров И. О., Максимова Н. Е., Безденежных Б. Н., Бодунов М. В. Закономерности формирования и реализации индивидуального опыта // Журнал высшей нервной деятельности. 1997. 47. № 2. С. 243–460.

ПРИМЕНЕНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБРАБОТКЕ ПСИХОЛОГИЧЕСКИХ ДАННЫХ¹

А. И. Панов

Институт системного анализа РАН (Москва),
МФТИ, ф-т управления и прикладной математики (Долгопрудный)
alex.panov.msk@gmail.com

В работе предлагается новый подход к обработке психологических данных, в основе которого лежат два метода интеллектуального анализа данных – метод АQ-покрытий и ДСМ-метод. На основе этого подхода реализован алгоритм, предназначенный для выявления причинно-следственных связей. Описано применение алгоритма в эксперименте и представлены его преимущества по сравнению со статистической обработкой.

Ключевые слова: интеллектуальный анализ данных, причинно-следственные связи, ДСМ-метод, АQ-покрытия.

¹ Работа выполнена при поддержке РФФИ (№07-06-80159, №09-07-12128-офи_м).

1 Введение

Одной из важнейших задач анализа данных является задача по выявлению и извлечению причинно-следственных (каузальных) зависимостей между свойствами объектов в массиве эмпирических данных. Однако само понятие каузальной зависимости требует предварительного уточнения с учетом особенностей предметной области, связанных с типом взаимосвязей входящих в нее объектов. Финн в своих работах (Финн, 2000) приводит описание трех «миров» (в нашем случае это классы предметных областей), с которыми работают интеллектуальные системы, отличающихся характером причинно-следственных связей. В первом «мире» все события случайны и связи между объектами носят стохастический, корреляционный характер. Во втором – наоборот, существуют только строго детерминистические отношения между объектами и причинами, которыми могут являться, например, проявления каких-либо свойств другого объекта. Третий мир – это одновременное сосуществование детерминации и случайных возмущений.

Для каждого типа предметной области существуют свои методы выявления каузальных зависимостей, основывающиеся на характерных особенностях этих зависимостей. Так, в первом «мире» применимы статистические инструменты. Во втором – широко известен развитый аппарат классического **ДСМ-метода** (Финн, 1991). Однако большое количество изучаемых предметных областей лишь приблизительно могут быть отнесены к первому или второму типу и на самом деле являются представителями третьего типа с разным соотношением детерминации и случайности, которое зачастую определяется разным уровнем исследованности связей между объектами. Характерным примером может служить психология, где взаимосвязь между психологическими характеристиками личности объяснима далеко не всегда, не смотря на ее очевидное наличие, и до сих пор главным математическим инструментом любого психолога является пакет программ по статистической обработке.

Одним из примеров попытки построить аппарат описания каузальных отношений в третьем «мире» служит расширение ДСМ-метода с помощью введения фальсификации по статистическим соображениям (Григорьев, 1996). Однако в этом и других методах корреляционные связи рассматриваются лишь как вспомогательный механизм для поправки функции сходства объектов. Показательными с точки зрения психолога могут служить работы А. А. Михеенковой в области социологии, которая, как предметная область, достаточно близка собственно психологии (Климова, Михеенкова, Панкратов, 1999). В этих работах строится так называемая диспозиционная теория регуляции социального поведения личности за счет выделения ситуации, недетерминировано влияющей на поведение субъекта. Т. е. происходит искусственное разделение влияющих факторов поведения на детерминированные и случайные, при этом, в основном, за счет знаний эксперта.

В разрабатываемом автором методе, представляющем данный подход, предлагается иное использование статистических соображений в процессе выявления каузальных связей. В основе метода лежат первоначальная статистическая обработка по построению корреляционного графа, вывод правил относительного описания объектов по методу **AQ-покрытий** (Michalski, 1973; Осипов, 2009) и первый шаг ДСМ-метода. AQ-обучение основывается на идее постепенного покрытия целевого класса объектов с помощью последовательно порождаемых правил. ДСМ-метод представляет собой индуктивный метод порождения гипотез о наличии причинно-следственных связей в объекте, описываемом набором значений атрибутов (свойств).

Основные шаги алгоритма, являющегося составной частью метода, были описаны в работе автора (Панов, 2010), здесь же мы остановимся на некоторых деталях применения метода в условиях реального эксперимента.

2 Подготовка данных

Входными данными для рассматриваемого метода являются данные психологического тестирования людей по некоторым признакам, значения которых задаются соответствующими шкалами измерений. Далее под объектами будем подразумевать конкретные данные, полученные от испытуемого. На объектах задается разбиение по классам (группам). Обычно такое разбиение определяется проводимым эмпирическим исследованием: тестирование людей разных профессий, возраста, вероисповедания и т. п. – или через выделение групп собственно психологического *классифицирующего признака* (уровня агрессивности, тревожности и т. п.). При достаточно большом количестве испытуемых возможно проведение предварительного **кластерного анализа**, по результатам которого строится так называемый *статистический классифицирующий признак*. В процессе полного цикла обработки данных одного тестирования предусматривается проведение *реклассификации* – выделение нового классифицирующего признака и соответствующее переразбиение объектов на классы.

Следует заметить, что наиболее продуктивной, в смысле находимых зависимостей, особенно при малых объемах выборок, является именно такая постановка эксперимента, когда разбиение на группы происходит естественным образом в ходе самого эксперимента. Это связано с тем, что в дальнейшем, в ходе работы алгоритма, происходит построение *отличительного описания* каждого класса, поэтому отличающиеся группы испытуемых в рамках эксперимента будут более кратко и емко (см. далее) описываться на этапе собственно анализа данных.

Еще одним предварительным и исключительно важным шагом является **дискретизация** шкал признаков. Выбор стратегии разбиения непрерывной шкалы на определенное количество непересекающихся интервалов во многом сказывается на получаемых результатах и определяется условиями предметной области и условиями эксперимента. Наиболее часто применяемые стратегии:

- *равномерное разбиение* используется при небольшом количестве объектов исследования, которые изначально равномерно распределены по всей шкале;
- *частотное разбиение* используется при достаточно большой выборке ($n > 50$);
- *разбиение по алгоритму ChiMerge* – наиболее универсально и использует метрику χ^2 для выявления интервалов с использованием информации о классах (Kerber, 1992).

Количество интервалов разбиения также играет большую роль. При проведении экспериментов было установлено, что наиболее интерпретируемые результаты при работе с психологическими данными получаются при разбиении на 3 интервала (высокое значение признака, среднее и низкое).

3 Определение меры однородности и структуры классов

После проведения предварительной обработки данных получается множество объектов тестирования, каждый из которых обладает набором *объектных свойств*,

разделенных на определенные классы. Свойство представляет собой пару: название признака (шкалы) и набор интервалов из множества всех интервалов, на которые разбита соответствующая шкала. В объектном свойстве набор интервалов представлен одним интервалом, в которое попадает значение данного признака у рассматриваемого объекта. Например, признак «Возраст» определяется непрерывной шкалой [0–100], которая разбивается на интервалы: «дети» (1-й интервал) [0–16], «взрослые» (2-й интервал) [16–55] и «пожилые» (3-й интервал) [55–100]. Объект «Иванов» имеет значение «40» по этому признаку, что означает, что он обладает свойством «взрослый» (или «Возраст=2», т. е. «Возраст – средний»).

При помощи метода АQ-покрытий строится описание каждого класса объектов, состоящее из *классовых свойств*, набор интервалов которых в общем случае состоит более чем из одного интервала (*дизъюнкция* интервалов). Например, группа людей может обладать следующим классовым свойством: «Возраст = IV2» (т. е., «Возраст – невысокий»). Это означает, что каждый представитель класса по признаку «Возраст» обладает либо свойством «ребенок» («Возраст – низкий»), либо свойством «взрослый» («Возраст – средний»). Классовые свойства каждой группы объединены в *правила* (*конъюнкция* свойств), каждое из которых описывает определенную подгруппу в данном классе. Набор правил класса обладает следующими свойствами:

- каждый объект класса описывается как минимум одним правилом из набора, т. е. он принадлежит как минимум одной подгруппе класса;
- ни одно правило набора не покрывает ни одного объекта из другого класса, т. е. ни один объект данного класса не описывается ни одним правилом из наборов других классов;
- каждое правило характеризуется **покрытием** (емкостью – количеством объектов исследования, подпадающих под это правило) и **сложностью** (например, длиной–краткостью).

Процесс построения набора правил является *расширением* объектных свойств до классовых и зависит от порядка встречающихся объектов и от порядка рассмотрения их свойств, что напоминает образование естественных понятий у человека в модели Дж. Брунера (Панов, Чудова, 2010).

Таким образом, каждый класс объектов эксперимента получает так называемое отличительное описание. В это описание попадают только те свойства, которые помогают отличить объект этого класса от объектов других классов. При этом общие характеристики набора, такие как количество правил в классе и их общая сложность, могут служить источником дополнительной информации о классе. Большое количество правил может говорить о разрозненности группы (крайний случай – количество правил равно количеству объектов, что говорит о том, что объединяющие признаки не выявляются в эксперименте). Наоборот, их малое количество – о большой однородности группы (крайний случай – одно правило, что говорит о большой схожести объектов). Так как правила строятся согласно принципу минимальности (критерий отбора правил при наличии альтернатив – наименьшая длина), то их сложность так же говорит о степени схожести объектов класса. Описываемые каждым правилом подгруппы определяют внутреннюю структуру класса.

4 Выявление причинно-следственных связей

Все уникальные классовые свойства набора правил образуют *универсум* свойств данного класса. В ходе экспериментов были выявлены пороги по размеру универсу-

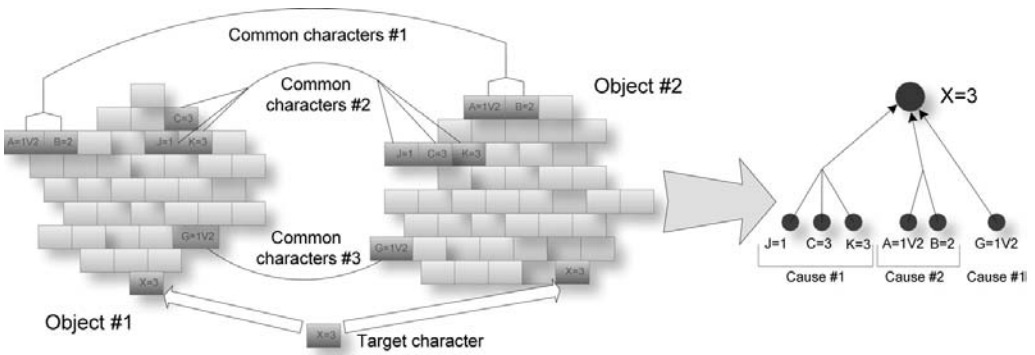


Рис. 1. Выявление причинно-следственных связей

ма, классифицирующие группу по возможности дальнейшего анализа (выявления причинно-следственных связей). При небольшом размере универсум является основой для **базы фактов** ДСМ-метода, которая содержит только «+» – элементы (объект обладает свойством) и «-» – элементы (объект не обладает свойством). *Целевым свойством* (для которого ищутся причины) выбирается поочередно каждое классовое свойство из универсума группы, все остальные свойства являются возможными компонентами причин (**атомами**). Объектами в базе фактов являются все объекты тестирования, в том числе и не принадлежащие анализируемому классу. Т. е. на данном этапе информация о классах переносится в состав универсума – для каждого класса будет свой универсум и, соответственно, свой тип структуры объектов (свой набор атомов).

Процесс выявления причинно-следственных связей, представленный на рисунке 1, основывается на понятии структурного сходства объектов, состоящих из атомов, входящих в универсум рассматриваемого класса, т. е. характерных для данного класса.

В качестве примера можно привести следующий эксперимент. После тестирования учителей и журналистов каждая группа описывается характерными для нее свойствами: («Вера в доброжелательность» – высокая, «Конфликтность» – низкая, «Напряженность» – низкая) – универсум учителей и («Ценность знания» – высокая, «Отчужденность» – низкая, «Напряженность» – низкая) – универсум журналистов. Затем выявляются причины наличия таких свойств, например, у учителей. Для этого каждый объект тестирования оценивается на наличие у него характерных для учителей свойств (оценка объекта с точки зрения его сходства с типичным учителем).

Таблица 1
База фактов

Объект	Вера в доброжелательность – высокая	Конфликтность – низкая	Напряженность – низкая
Алиса (учитель)	+	-	+
Вера (учитель)	+	+	+
Иван (учитель)	-	+	-
Алекс (журналист)	-	-	-
Анна (журналист)	+	-	-

Далее, уже вне зависимости от того, относится ли объект к учителям, ищется то общее между всеми объектами, обладающими целевым свойством (например, «Напряженность – низкая»), что и называется причиной (в данном примере, причина – «Вера в доброжелательность – высокая», так как именно это является общим между Алисой и Верой) этого целевого свойства.

Заключение

Выявляемые причинно-следственные связи обладают следующими отличиями от корреляционных статистических связей:

- наличие связи задается структурным сходством объектов, в то время как корреляционная связь отражает лишь совместное изменение признаков при поочередном рассмотрении объектов;
- наличие связей также определяется характерными свойствами анализируемого класса, в то время как корреляционные связи не учитывают классификацию;
- процесс выявления причинно-следственных связей методами искусственного интеллекта моделирует рассуждения эксперта при работе с данными.

Стоит заметить, что применение методики дает возможность учета как детерминированных связей, так и статистических соображений.

Метод был реализован в качестве программной системы, которая осуществляет поддержку по обработке экспериментальных данных и предоставляет средства по визуализации результатов. Было проведено пилотажное исследование на материале данных 67 испытуемых, характеризующихся по 44 признакам 7 опросников. Как показало это исследование, преимуществом разрабатываемой системы является то, что получаемые результаты более легко интерпретируются и более наглядно визуализируются, чем классическое представление стохастических связей в виде корреляционного графа.

Литература

- Григорьев П. А. Sword-системы или ДСМ-системы для цепочек, использующих статистические соображения // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 1996. № 5–6. С. 45–51.
- Климова С. Г., Михеенкова М. А., Панкратов Д. В. ДСМ-метод как метод выявления детерминант социального поведения // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 1999. № 12. С. 3–14.
- Осинов Г. С. Лекции по искусственному интеллекту. М., 2009.
- Панов А. И. Методика интеллектуального анализа результатов психологического тестирования // Труды I Всероссийской научной конференции молодых ученых. 2010. Т. 1. С. 39–45.
- Панов А. И., Чудова Н. В. Моделирование процесса образования естественных понятий методами искусственного интеллекта // Четвертая международная конференция по когнитивной науке: Тезисы докладов. В 2 т. Томск, 22–26 июня 2010 г. 2010. Т. 2. С. 455.
- Финн В. К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ // Итоги науки и техники. 1991. Т. 15. С. 54–101.
- Финн В. К. Каузальный анализ данных в интеллектуальных системах // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 2000. № 11. С. 1–5.

- Kerber R. ChiMerge: Discretization of Numeric Attributes // In Proc. AAAI-92, Ninth.
- Michalski R. S. AQVAL/1-Computer Implementation of Variable-Valued Logic System VL1 and Examples of its Application to Pattern Recognition // Proc. Of the First Int. Joint Conf. on Pattern Recognition. Washington, DS, 1973. P. 3–17.
- National Conference Artificial Intelligence. AAAI Press / The MIT Press. 1992. P. 123–128.

ЗАВИСИМОСТЬ РЕЗУЛЬТАТОВ ПСИХОФИЗИЧЕСКОГО ЭКСПЕРИМЕНТА ОТ ВЫБРАННОЙ МЕРЫ СВЕТОВОГО И ЭЛЕКТРИЧЕСКОГО СТИМУЛА

А. П. Пахомов

Российский университет дружбы народов (г. Москва)

Pakhomoff44@yandex.ru

В статье проверяется предположение о том, что функции, полученные в психофизическом эксперименте со световой и электрической стимуляцией, зависят от меры, выбранной для репрезентации величины физического стимула. Показано, что в зависимости от выбранной меры указанных физических стимулов результаты классических психофизических экспериментов могут иметь неклассическую интерпретацию.

Ключевые слова: психофизическое шкалирование; кривые Фехнера и Стивенса, отношение Вебера, мера физического стимула.

Измерение физического стимула обычно имеет несколько альтернатив. Отсюда выбор шкалы стимула может определять: а) величину экспоненты в показательной функции; б) величину константы k в логарифмической функции (функции Фехнера); в) будет ли общая форма психофизической функции показательной, логарифмической или какой-либо другой; г) величину отношения Вебера и степень постоянства, которую оно имеет; д) междимензиональные корреляции между показателем степенной функции и динамическим диапазоном стимула.

Стимулами в психофизике являются основные физические величины длины, массы, времени и такие величины как вес, плотность, частота стимула, сила, энергия и т. д., которые обычно можно выразить через базисные величины с помощью простых математических функций. Приписывание чисел основным физическим величинам в общем случае рассматривается как пример процесса экстенсивного измерения (Krantz, Luce, Suppes, Tversky, 1971).

Как показано в работе (Myers, 1982), психофизические функции могут зависеть от того, как определен физический стимул. Ведет ли «прямое» психофизическое шкалирование к показательной, логарифмической или к любой другой функции, может определяться тем, какой аспект стимула выбран для того, чтобы представить его «интенсивность». Кроме того, величина экспоненты в показательной функции может зависеть от принятой меры стимула, а, если функция рассматривается как логарифмическая по результатам «прямого» субъективного шкалирования или же из-за принятия предположений Фехнера, величина константы k в выражении $\psi = k \log \phi$ может зависеть от определения шкалируемого стимула.

Абсолютная величина отношения Вебера также может изменяться как функция меры стимула. Кроме того, отношение Вебера может быть константой при одном варианте измерения ϕ , но изменяться систематическим образом при переходе к другим вариантам измерения ϕ в той же самой модальности стимула. Отсюда вытекает,