

Взаимодействие логических и статистических методов анализа психологических данных

Александр Игоревич Панов

Отдел «Интеллектуальные динамические системы и когнитивные исследования»
Институт системного анализа
Федеральный исследовательский центр «Информатика и управление»
Российской академии наук

21 марта 2018

Семинар «Анализ больших данных в информатике, управлении и
психологии»



Основа статистических методов анализа данных (факторного, регрессионного и др.) — **корреляционная связь**.

Корреляция между двумя признаками A и B в наборе данных X — не есть каузальная зависимость между признаком A и B .

Классический пример трех признаков: A, B, C . Пусть $A \rightarrow B, A \rightarrow C$, тогда при уменьшении или увеличении проявления признака A будет в меньшей или большей степени соответственно проявляться и признаки B и C , т.е. B и C будут изменяться согласованно.

Основа статистических методов анализа данных (факторного, регрессионного и др.) — **корреляционная связь**.

Корреляция между двумя признаками A и B в наборе данных X — не есть каузальная зависимость между признаком A и B .

Классический пример трех признаков: A, B, C . Пусть $A \rightarrow B, A \rightarrow C$, тогда при уменьшении или увеличении проявления признака A будет в меньшей или большей степени соответственно проявляться и признаки B и C , т.е. B и C будут изменяться согласованно.

Между ними будет определен высокий уровень корреляции, что на самом деле не отражает реально имеющихся причинно-следственных связей.

Интеллектуальный анализ данных → логические методы анализа — моделирование индуктивных и дедуктивных рассуждений, рассуждений по аналогии.

Интеллектуальные методы позволяют смоделировать процесс принятия решений специалистом в данной предметной области.

ДСМ-метод — метод порождения гипотез о наличии причинно-следственных связей.

Интеллектуальный анализ данных → логические методы анализа — моделирование индуктивных и дедуктивных рассуждений, рассуждений по аналогии.

Интеллектуальные методы позволяют смоделировать процесс принятия решений специалистом в данной предметной области.

ДСМ-метод — метод порождения гипотез о наличии причинно-следственных связей.

Интеллектуальный анализ данных → логические методы анализа — моделирование индуктивных и дедуктивных рассуждений, рассуждений по аналогии.

Интеллектуальные методы позволяют смоделировать процесс принятия решений специалистом в данной предметной области.

ДСМ-метод — метод порождения гипотез о наличии причинно-следственных связей.

Существенные преимущества: интерпретируемость и работа в сложном каузальном пространстве со многими классами объектов.

Существенные недостатки: комбинаторная сложность (низкая эффективность) и проверка гипотез (высокая доля ложных гипотез).

Основная идея метода: проверять логические гипотезы статистическими методами.

- 1 Построение логического описания объектов различных классов.
- 2 Формирование базы фактов вида Объекты *times* Свойства.
- 3 Применение первого этапа ДСМ-метода для выдвижения гипотез о наличии причинно-следственных связей.
- 4 Фальсификация гипотез путем структурного моделирования.
- 5 Отбор прошедших предыдущий этап гипотез.

Дано:

- Множество $X = \{x_i\}$ — множество из m объектов (испытуемых).
- Отображение $f_i : X \rightarrow D_j$ — признаки объекта, $D = \{d_k^j\}$ — шкалы методик и опросников.
- $F = \|f_j(x_i)\|_{m \times n}$ — матрица объекты-признаки.
- $C = \{c_l\}$ — множество классов, на которые были разбиты объекты.

Предобработка:

- Дискретизация интервальных признаков: $(d_1^j, d_k^j) = w_1 \cup w_2 \cup w_3$.
- Переход к свойствам объектов $h_k^i = x_i \in w_k$.
- Удаление дублирующих строчек матрицы объекты-свойства $\|f_j(x_i)\|_{m \times n}$.

AQ-обучение:

- Для каждого класса c_t строим набор AQ-правил вида

AQ-обучение:

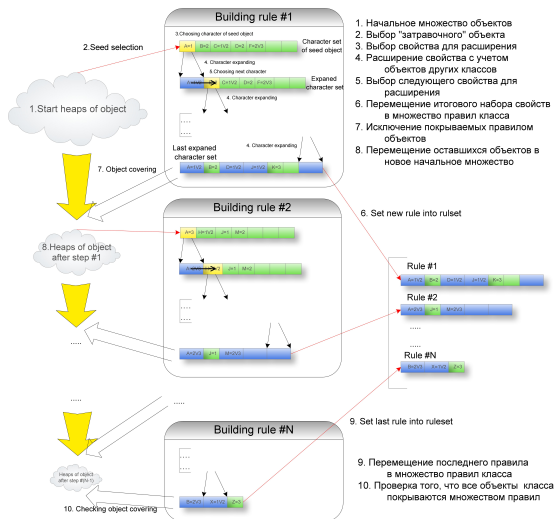
- Для каждого класса c_t выбираем затравочный объект и строим базу правила вида $R_i = \{d_i^1 = w_{i_1}, \dots, d_i^j = w_{i_j}\}$.
- Перебираем примеры класс итерационного обобщая правило, получаем набор AQ-правил вида

$$R_{ij} = \bigcap_j \left(d^j = \bigcup_q w_q \right).$$

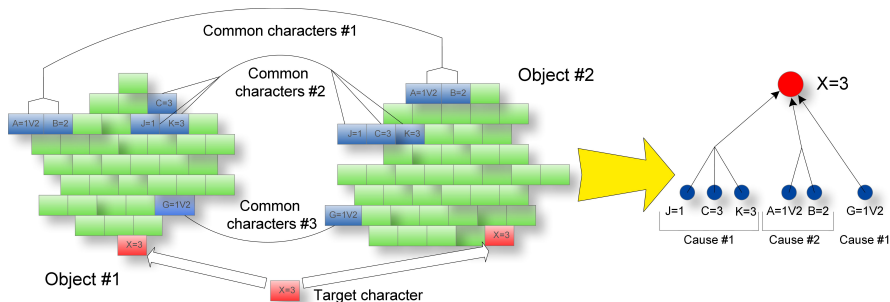
Этап ДСМ:

- Составление базы фактов ДСМ-метода — битовая матрица $B_{m \times 3n}$.
- Сокращение пространства поиска — удаление противоречивых примеров.
- Использование гипотез для выявления каузальных связей:
 - *Гипотеза сходства.*
 - *Гипотеза различия.*
 - *Гипотеза абдукции.*
- Получаем множество гипотез вида:

$$H(h_g, c_l) = \bigcap_{j \neq g} \left(d^j = \bigcup_q w_q \right).$$

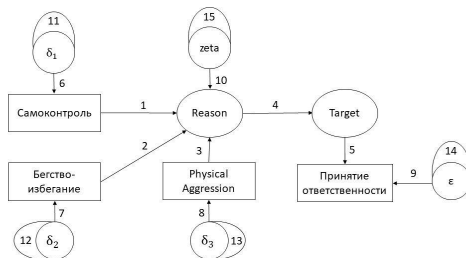


Этап ДСМ анализа



Алгоритм перевода ДСМ гипотез в диаграмму путей:

- Наблюдаемые переменные $X' = \{x_1, x_2, \dots, x_m\}$ — множество признаков f_1, f_2, \dots, f_n , участвующих в гипотезе, в том числе и целевая.
- Латентные переменные — признаки из комплексных свойств причин d^j .
- Наличие интервалов учитывается при определении параметров модели a_1, a_2, \dots
- Запуск итерационного процедуры проверки соответствия дисперсий и ковариаций имеющимся данным.



- Chudova, N. V. and A. I. Panov. "Causal Inference in Psychological Data in the Case of Aggression". *Scientific and Technical Information Processing*. 2017.
- Boiko, A. A. et al. "A framework for automated meta-analysis: Dendritic cell therapy case study". *2016 IEEE 8th International Conference on Intelligent Systems (IS)*. 2016.
- Чудова, Н. В. и А. И. Панов. «Извлечение причинно-следственных отношений из данных психологического исследования на материале изучения агрессивности». *Искусственный интеллект и принятие решений*. 2016.
- Lupatov, A. Yu. et al. "Assessment of Dendritic Cell Therapy Effectiveness Based on the Feature Extraction from Scientific Publications". *Proceedings of ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods*. 2015.
- Panov, A. I., A. V. Shvets, and G. D. Volkova. "A Technique for Retrieving Cause and Effect Relationships from Optimized Fact Bases". *Scientific and Technical Information Processing*. 2015.
- Бойко, А. А. и др. «Система анализа данных из научных источников на примере текстов по дендритноклеточным вакцинам». *Труды Института Системного Анализа РАН*. 2015.
- Панов, А. И., А. В. Швец и Г. Д. Волкова. «Метод извлечения причинно-следственных связей с использованием оптимизированных баз фактов». *Искусственный интеллект и принятие решений*. 2015.
- Panov, A. I. "Extraction of Cause – Effect Relationships from Psychological Test Data Using Logical Methods". *Scientific and Technical Information Processing*. 2014.
- Панов, А. И. и А. В. Швец. «Эволюционный метод покрытий для составления базы фактов ДСМ–метода». *Четырнадцатая национальная конференция по искусственному интеллекту с международным участием КИИ–2014 (24–27 сентября 2014г., г. Казань, Россия): Труды конференции*. 2014.
- Панов, А. И. «Выявление причинно-следственных связей в данных психологического тестирования логическими методами». *Искусственный интеллект и принятие решений*. 2013.

Спасибо за внимание!

pan@isa.ru

<https://github.com/cog-isa/aqjsm>