

ОЦЕНКА КАЧЕСТВА ИНФОРМАЦИОННОЙ СИСТЕМЫ СОЦИАЛЬНО-ПСИХОЛОГИЧЕСКОЙ ДИАГНОСТИКИ МЕТОДАМИ DATA MINING

© 2017 Макаров А.А., Китаев Д.Ф.

Самарский университет государственного управления
«Международный институт рынка», г. Самара, Россия

Статья посвящена проблеме оценки качества информационной системы, предназначенной для социально-психологического тестирования коллектива. Показано, что решение такой задачи может быть основано на принципах Data mining, т.е. совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Применение данных методов возможно в том случае, если есть большой массив числовых данных. Показано, что такой массив может быть сформирован с помощью рассматриваемой системы социально-психологического тестирования. Данная система состоит из совокупности модифицированных психологических тестов. Было показано, что результаты тестирования сохраняются в базе данных, а затем интерпретируются с целью формирования различных выходных показателей. Рассмотрены принципы анализа совокупности числовых данных с помощью системы Deductor Studio Academic 5.2. Показан процесс реализации кластеризации данных на основе алгоритма Кохонена.

Ключевые слова: коллектив, тестирование, психологический тест, информационная система, программный комплекс, психологическое тестирование, база данных, интегральный показатель, Datamining, нейронная сеть, алгоритм Кохонена.

В работах Китаева Д.Ф., Макарова А.А., Макаровой Л.В., Смольникова С.Д. приведено описание информационной системы позволяющей решать широкий круг задач:

- накопление, сертификация и модификация банка профессиональных тестовых систем;
- автоматизированное профессиональное тестирование персонала;
- сохранение результатов профессионального тестирования в базе данных;
- автоматизированное психологическое тестирование персонала;
- сохранение результатов психологического тестирования в базе данных;
- обработка результатов профессионального и психологического тестирования с целью получения обобщенных характеристик;
- получение, сохранение, выборка как индивидуальных качественных показателей (личности, специалиста и др.), так и обобщенных качественных характеристик коллектива [1, 7, 2, 4].

Таким образом, информационная система состоит из двух подсистем:

1. Подсистема комплексной социально-психологической диагностики личности сотрудников.

2. Подсистема диагностики профессиональной компетентности сотрудников на основе двухкомпонентной профессиональной тестовой системы узкой направленности [3].

Рассмотрим более подробно работу подсистемы комплексной социально-психологической диагностики личности сотрудников (КСПД).

В основе КСПД лежит социально-психологическое тестирование на основе батареи адаптированных психологических тестов:

- тест стандартными прогрессивными матрицами Равена;
- модифицированный 16-факторный личностный опросник Кеттелла;
- модифицированный тест Томаса;
- ситуационный тест на социальную зрелость;
- тест Кейрси;
- тест на уровень развития вербального интеллекта (авторская разработка).

Сам процесс тестирования возможен в двух режимах: off-line (тесты представлены

на бумажных носителях, текстовый файл с исходными данными формируется вручную) или on-line (текстовый файл с исходными данными формируется автоматически) (рис. 1).

(Образец, 25/10/16)
№:1
1.ФАМ:Иванова 2.ИМЯ:Ольга 3.ОТЧ:Ивановна ПОЛ:Жен
4.ГОРОД:Самара 5.ОРГАНИЗАЦИЯ:ООО "Ибис"
6.ПОДРАЗДЕЛЕНИЕ:Отдел закупок 7.ДОЛЖНОСТЬ:Менеджер
8.ГОД РОЖДЕНИЯ:1978 9.ОБЩ. ТРУД. СТАЖ:10 10.В ДАННОЙ ОРГ-ИИ С:2007
11.В ЗАНИМАЕМОЙ ДОЛЖНОСТИ С:2007 12.ОБРАЗОВАНИЕ:В
13.В:2008 ГОДУ ЗАКОНЧИЛ:Самарский государственный экономический университет
14.ПО СПЕЦИАЛЬНОСТИ:Экономика
15.ТЕЛЕФОН: РАБОЧИЙ *** ДОМАШНИЙ:***
16.ОСНОВНЫЕ ЦЕЛИ ДЕЯТЕЛЬНОСТИ И РЕШАЕМЫЕ ЗАДАЧИ:ВС
РАВЕН:451263621345261213551544823874517212343786551266768610000000
ВЕРБАЛЬНЫЙ:333113224132254442214000
СИТУАЦИОННЫЙ:a8 a10 65 v9 610 610 a10 a10 a6 68 67 610 610 a8
ПСИХОЛОГИЧЕСКИЙ:
1- 60:1133132133113233121133133111112112211331331111133333233133
61-120:31333333122333333113331121132111111131331131333112331313113
121-180:31311123131113111222112122221111222121122111112212111221211
181-235:212211111122122122212222222112121122211112212121122
**

Рисунок 1- Образец исходных данных психологического тестирования

Система обрабатывает массив исходных данных и результаты обработки сохраняет в базе данных «Кадры». Эта база данных формата Microsoft Access содержит множество таблиц, запросов, форм и отчетов (рис. 2).

Таблицы

factorB

factorC

factorE

factorF

factorG

factorH

factorM

factorO

factorQ1

factorQ3

factorQ4

Информация о шка...

Кадры

Образование

Основная

Имя поля	Тип данных
Код	Счетчик
Дата	Дата/время
ном	Числовой
Кто это	Текстовый
Фам	Текстовый
Имя	Текстовый
Отч	Текстовый
Пол	Текстовый
Город	Текстовый
Орг	Текстовый
Под	Текстовый
Дол	Текстовый
Грожд	Числовой
Стаж	Числовой
В_орг	Числовой
В_долж	Числовой
Образ	Текстовый
Когда	Числовой

Рисунок 2 - Фрагмент базы данных «Кадры»

Первичные баллы результатов тестирования группируются в интегральные показатели с целью получить итоговый результат – количественную и качественную оценки конкурентоспособности тестируемых.
За несколько лет эксплуатации данной системы был накоплен большой статистиче-

ский материал - несколько тысяч протестированных. В связи с этим появилась принципиальная возможность оценить качество функционирования системы КСПД методами Data mining.
В такой постановке задача может быть сформулирована следующим образом: необ-

ходимо определить максимально возможный набор независимых параметров для оценки адекватности модели тестирования.

Для целей исследования массива данных на автокорреляцию и наличие скрытых закономерностей из набора данных были исключены качественные параметры, а также интегральные параметры, представляющие собой компиляцию других параметров.

В результате в наборе данных остались следующие поля:

- числовая оценка формальных параметров: возраста, образования, стажа, формулировка цели (поле «Фор»);
- числовая оценка интеллектуального потенциала (поле «Инт»);
- числовая оценка способности к руководству (поле «Рук»);
- числовая оценка исполнительных качеств (поле «Исп»);
- числовая оценка коммуникативных способностей (поле «Комм»);
- числовая оценка психологической устойчивости (поле «Эмоц»).

Массив данных содержал 3395 записей по 6 параметров в каждой, т.е. 20375 значений.

Стандартный корреляционный анализ позволяет найти матрицу корреляций между полями, но не дает возможность «увидеть» более сложные (скрытые) закономерности, например, совместное влияние двух и более факторов на третий.

Такого рода задачи решаются с помощью многофакторного дисперсионного анализа и кластерного анализа. В последние годы появились мощные аналитические методы и инструментальные средства для кластерного анализа больших массивов данных, в частности, нейронные сети. Кластерный анализ использовался нами для решения поставленной задачи.

Нейронная сеть представляет собой совокупность искусственных нейронов, организованных по слоям. Каждый нейрон суммирует с определенными весовыми коэффициентами $w_1, w_2, \dots, w_n$ информацию $x_1, x_2, \dots, x_n$ с множества входов и преобразует эту информацию с помощью функции активации f в выходной сигнал y :

$$y = f\left(\sum_i x_i w_i\right).$$

Для первого слоя нейронов входные сигналы $x_1, x_2, \dots, x_n$ – это непосредственная входная информация, а для последующих слоев – это выходы нейронов предыдущего слоя. Нейронная сеть (в дальнейшем – просто сеть) начнет выполнять свои функции только тогда, когда определены все весовые коэффициенты w_i . Процесс определения этих коэффициентов называется «обучением сети».

Если имеется набор образов (шаблонов)

$$\bar{X}^M = (x_1^M, x_2^M, \dots, x_i^M),$$

для которых известен результат, $(\bar{y}^M)$, то такой набор можно использовать для обучения сети, после чего она становится способной распознавать и классифицировать образы из потока данных, для которых результат неизвестен. Однако далеко не всегда такое обучающее множество известно. В этих случаях используются алгоритмы самообучения сети. Одним из самых популярных таких алгоритмов является алгоритм, предложенный финским математиком Т. Кохоненом.

Суть алгоритма Кохонена в следующем: во входном потоке данных сеть улавливает особенности (закономерности) и разделяет этот поток на несколько групп (кластеров). Это происходит следующим образом: в многомерном пространстве параметров вычисляются декартовы расстояния от данного образа (строки данных) до каждого нейрона входного слоя, и соответствующий образ относится к тому нейрону, для которого это расстояние минимально.

Для лучшей визуализации результатов описанным выше способом строятся самообучающиеся сети Кохонена особого вида – карты Кохонена, на которых соответствующие декартовы расстояния обозначены различными цветами. Такие карты дают емкий образ анализируемого многомерного набора данных. В заключение заметим, что на входе сети данные должны быть нормированы.

В качестве инструментального средства был использован пакет Deductor Studio Academic 5.2, в котором реализован алгоритм Кохонена. Все обозначенные выше поля представлялись как входные, все множество записей делилось на группу самообучающихся (90%) и группу тестовых (10%).

Карта Кохонена для каждого из полей планировалась в виде 192 шестиугольных ячеек, в каждой из которых в среднем было порядка 100 значений, что обеспечивало достаточную статистическую устойчивость к случайным изменениям данных.

На уровне достоверности 95% (рис. 3) во входном потоке данных не было выявлено разделение на кластеры, что свидетельствует о том, что в этом наборе отсутствуют скрытые многофакторные закономерности и корреляции между выбранными полями.

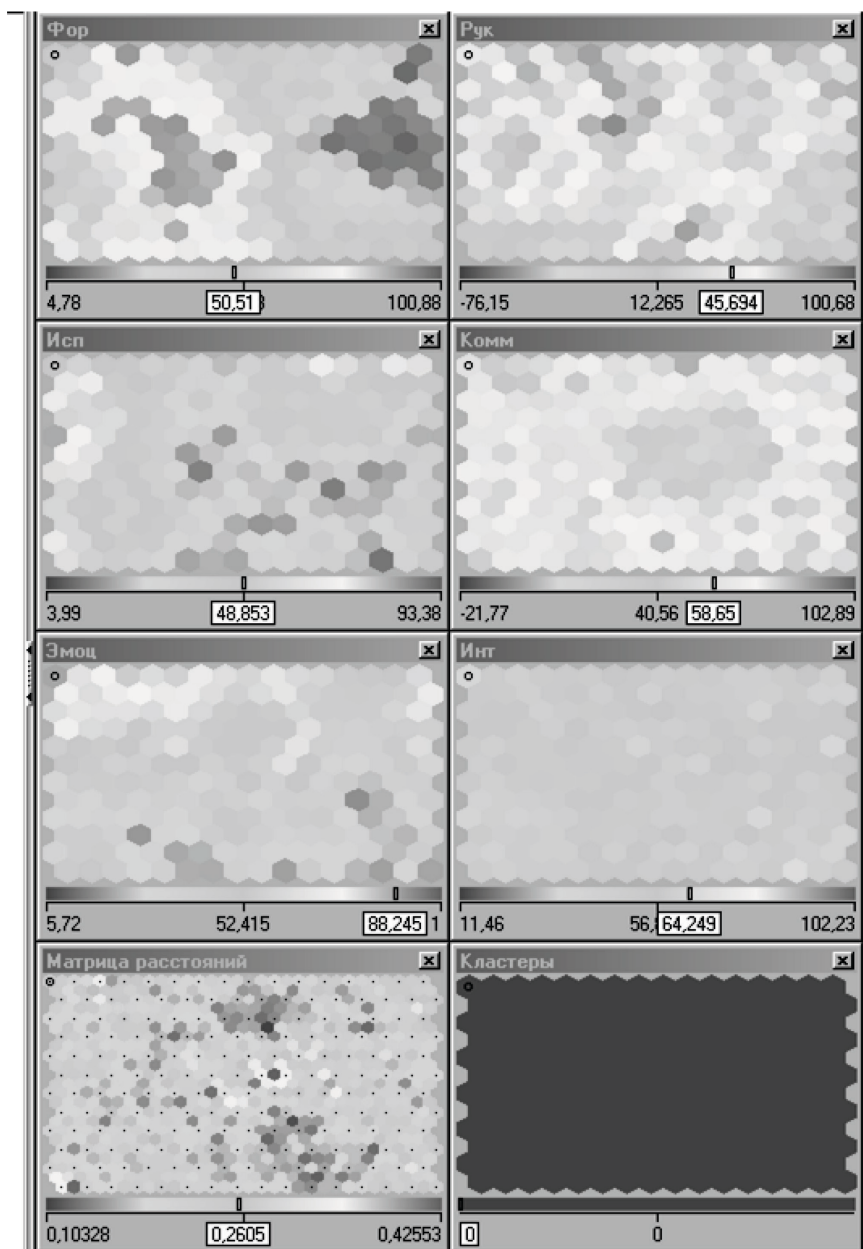


Рисунок 3 - Результат кластеризации

Таким образом, эти поля являются независимыми параметрами и могут адекватно использоваться для представления данных тестирования.

СПИСОК ИСТОЧНИКОВ

1. Китаев Д.Ф., Макаров А.А., Макарова Л.В., Смольников С.Д. Информационная система для разработки имитационной модели оценки коллектива// Вестник Международного института рынка. - 2016. - № 1. С. 190-198.
2. Китаев Д.Ф., Макаров А.А., Макарова Л.В., Смольников С.Д. Имитационная модель оцен-

ки эффективности обучения в группе // Вестник Международного института рынка. 2015. № 2. С. 92-98.

3. Макаров А.А., Смольников С.Д. Вестник Самарского государственного технического университета. Серия: Экономические науки. 2011. № 2. С. 90-95.

4. Китаев Д.Ф., Макаров А.А., Смольников С.Д. Алгоритм формирования многоагентной модели группового образовательного процесса // Современные проблемы науки и образования. – 2015. – № 2-3. С.2

QUALITY RATING OF INFORMATION SYSTEM OF SOCIAL AND PSYCHOLOGICAL DIAGNOSTICS BY DATA MINING METHODS

© 2017 Alexey A. Makarov, Dmitriy F. Kitaev

Samara University of Public Administration “International Market Institute”, Samara, Russia

The article deals with the quality assessment of the information system for social and psychological testing of staff. It is shown that the solution of this problem can be based on the principles of Data mining, i.e. all detection methods in data previously unknown, nontrivial, practically useful and accessible interpretation of the knowledge necessary for decision-making in various spheres of human activity. The application of these methods is possible, if there is a large array of numerical data. It is shown that such an array can be formed using the system of social and psychological testing. This system consists of a set of modified psychological tests. It was shown that the test results are stored in the database and then interpreted with the aim of formation of various output indicators. The article considers the principles of the analysis of a set of numerical data through a system with Deductor Studio Academic 5.2. The authors show the process of implementing a data clustering algorithm based on Kohonen.

Keywords: staff, test, psychological test, information system, software system, psychological testing, database, integrated indicator Data-mining, neural network, Kohonen algorithm