

«Сейфуллин окулары – 12: Ғылым жолындағы жастар-болашақтың инновациялық әлеуеті» атты Республикалық ғылыми-теориялық конференция материалдары = Материалы Республиканской научно-теоретической конференции «Сейфуллинские чтения-12: Молодежь в науке - инновационный потенциал будущего" . – 2016. – Т.1, ч.3 – С.307-309

АНАЛИТИКА ДАННЫХ ПСИХОДИАГНОСТИКИ

Билялова М.

Статистические методы анализа данных повсеместно используются для обработки данных психодиагностики, как статистическая теория оценивания, корреляционный, факторный анализ и т.д. При проведении экспериментальных исследований возникают ситуации, когда интерпретировать результаты приходится при большом числе исходных психологических признаков (большая размерность данных) и недостаточном для полноценного статистического анализа числе респондентов. В таких условиях анализ взаимосвязей между психологическими признаками является трудной задачей.

Большинство методов диагностирования исходят из предположения, что данные поступают равномерным постоянным потоком. На практике такое встречается редко. Поэтому одна из самых востребованных областей применения хранилищ данных - диагностирование - оказывается реализованной некачественно или со значительными ограничениями. Для борьбы с этим явлением применяют следующие методы:

1. Аппроксимация. Если нет данных в какой-либо точке, берется ее окрестность и известным формулам вычисляется значение в этой точке. Соответствующая запись добавляется в хранилище. Это работает для упорядоченных данных.

2. Определение наиболее правдоподобного значения. Для этого берется не окрестность точки, а все данные. Этот метод применяется для неупорядоченной информации, т.е., когда мы не в состоянии определить, что же является окрестностью исследуемой точки.

Перечисленные проблемы могут быть минимизированы, если для анализа психодиагностических данных использовать интеллектуальный анализ, основанный на многомерной обработке данных. Инструменты для анализа данных и поддержки принятия решений в настоящее время считаются одним из важнейших типов приложений. Реализация подобного анализа нередко базируется на построении многомерных хранилищ данных (datawarehouses) и на аналитической обработке данных (On-Line Analytical Processing, OLAP) — популярной технологии многомерного анализа.

Таким образом, программное обеспечение может использоваться в практической психологии [1,2].

Постановка задачи. В настоящей работе приводятся результаты анализа психодиагностических данных группы людей. Исследовались взаимосвязи

психологических характеристик личности группы людей с их психодиагностической характеристикой [3].

Во время обучения различные наборы входных данных (из базы данных примеров) подаются на входной слой программного обеспечения, а полученные выходные сигналы сравниваются с известными целевыми значениями (из той же базы данных примеров), подсчитываются ошибки, модифицируются и дают наибольший вклад в ошибку (с помощью алгоритма «обратного распространения ошибки» [4]), и этот цикл повторяется много раз, пока не достигается приемлемая точность отображения входной информации в выходную. Число циклов может достигать 10 000.

Выбор данных для обучения сети и их предобработка являются самыми важными этапами решения задачи. Набор данных для обучения должен быть очищен, т.е. должны быть исключены противоречия, дубликаты, аномальные значения.

Существуют также другие теоретически и эмпирически установленные правила использования программного продукта при анализе данных [5]. Например, число вычислительных путей должно быть в 2-5 раз меньше, чем число примеров (различных наборов данных). Использование более сложной платформы (использование большего числа вычислительных путей) приводит к так называемому «переобучению» ПО. В этом случае продукт хорошо запоминает примеры, но теряет способность к обобщению. То есть он хорошо запоминает и восстанавливает примеры из базы данных, но плохо «вычисляет» другие наборы данных, отличающиеся от тех, которые были в базе данных примеров. Для того, чтобы контролировать, что делает ПО (просто запоминает примеры или обобщает существующие в примерах данных зависимости) поступают следующим образом, при работе базу данных примеров делят на две части. Большую (обучающую) используют для обучения, а меньшую (тестовую) – для проверки обученной платформы на точность восстановления примеров [6].

Методика обработки данных. Для принятия правильных решений при создании моделей интеллектуального анализа данных необходимо понимать данные. Методы исследования данных включают в себя расчет минимальных и максимальных значений, вычисление средневероятного и стандартного отклонения и изучение распределения данных. Например, по максимальному, минимальному и среднему значениям можно заключить, что выборка данных не является репрезентативной для имеющихся процессов, и поэтому необходимо получить более сбалансированные данные или изменить предположения, лежащие в основе ожидаемых результатов. Стандартное отклонение и другие характеристики распределения могут сообщить полезные сведения о стабильности и точности результатов. Большая величина стандартного отклонения может свидетельствовать о том, что добавление новых данных поможет усовершенствовать модель. Данные, которые сильно отклоняются от стандартного распределения, могут оказаться искаженными или

представлять точную картину реальной проблемы, которая делает сложным подбор соответствующей модели для данных. Для анализа результатов диагностики развития интеллектуальной сферы использовалась специальная аналитическая платформа многомерного анализа. Реализованная архитектура, позволяет добиться максимальной гибкости при создании законченного решения. Благодаря данной архитектуре можно собрать в одном аналитическом приложении все необходимые инструменты анализа и реализовать автоматическое выполнение подготовленного сценария [7].

Технологическая платформа включает средства, позволяющие максимально сократить время разработки, быстро создавать и выводить на рынок новые прикладные решения и в короткие сроки адаптировать их в соответствии с изменяющимися требованиями предприятий. Возможности платформы обеспечивают не только высокую скорость первоначальной разработки продукта, но и его быструю адаптацию в дальнейшем.

Создание законченного решения занимает очень мало времени. Достаточно получить данные, определить сценарий обработки и задать место для экспорта полученных результатов. Наличие мощного набора механизмов обработки и визуализации позволяет двигаться по шагам, от наиболее простых способов анализа к более мощным. Таким образом, первые результаты пользователь получает практически сразу, но при этом можно легко наращивать мощность решения [8, 9].

Анализ результатов. Переход информационной отрасли производственного управления с учетного на новый, аналитический уровень — важный шаг в развитии современного предприятия. Применение технологий анализа данных позволяет существенно повысить оперативность и качество принятия решений, расширить информационный охват и точность оценки работы пользователей, восстановить недостающие сведения и получить прогноз развития личности [8, 9].

Заключение

Полученные результаты сразу позволяют выделить личностные качества, сильно влияющие на психофизиологический портрет личности.

Если оценивать значимость личностных качеств по их влиянию на коэффициент психофизиологического состояния, то достаточно большой диапазон изменения критериев наблюдается не для всех личностных качеств (признаков).

Таким образом, полученные при помощи данного программного обеспечения, данные согласуются с результатами психологических исследований, проведенными ранее с использованием корреляционного и факторного анализа. Но применение платформы позволяет выявлять и непосредственно вычислять нелинейные связи психологических признаков, которые не удастся обнаружить традиционными статистическими методами обработки данных психодиагностики.

Список литературы

1. Величковский Б.М. Когнитивная наука: Основы психологии познания: в 2 т. М.: Академия, 2006.
2. Воробьев А.В. Обзор применения математических методов при проведении психологических исследований [Электронный ресурс] // Психологические исследования: электрон. науч. журн. 2010. №2(10). URL: <http://www.psystudy.ru/index.php/num/2010n2-10/311-vorobiev10.html>.
3. Baxt W.G. Complexity, chaos and human physiology: the justification for non-linear neural computational analysis // Cancer Lett. 1994. Vol. 77, №2-3. P. 85-93.
4. Cattell R.B. Advanced in Cattellian Personality Theory. Handbook of Personality. Theory and Research. N.Y.: The Guilford Press, 1990.
5. Hebb D. Organization of behavior. N.Y.: Science Edition, 1961.
6. Ларичев О.И., Петровский А.В. Системы поддержки принятия решений. Современное состояние и перспективы их развития. // Итоги науки и техники.- Сер.: Техническая кибернетика. Т.21. М. ВИНТИ, 1987.
7. Анохин П. К. Очерки по физиологии функциональных систем. - М.: Медицина, 1974.
8. Астанин С.В., Захаревич В.Г. Информационно-советующие комплексы систем гибридного интеллекта.- Таганрог: Изд. Таганрогского государственного радиотехнического университета, 1997.
9. Астанин С.В., Чепиков Э.В. Принципы построения тренажно-моделирующих комплексов в среде виртуальной реальности/ Материалы всероссийской НПК «Человеческое измерение в информационном обществе».- М.: ВВЦ, 2003.-С.116