

УДК 519.2

КЛАСТЕРНЫЙ АНАЛИЗ КАК СРЕДСТВО УМЕНЬШЕНИЯ РАЗМЕРНОСТИ ПРИЗНАКОВОГО ПРОСТРАНСТВА ПРИ ИССЛЕДОВАНИИ НА ИНТЕРНЕТ-ЗАВИСИМОСТЬ

Иванов В. С., Колесникова М. А., Знаенко Н. С.

Ульяновский институт гражданской авиации

имени главного маршала авиации Б. П. Бугаева, г. Ульяновск

Цель исследования: используя средства многомерной статистики, снизить размерность признакового пространства для исследования на Интернет-зависимость.

В любой научной деятельности классификация является одной из составляющих, без которой невозможны построение и проверка научных гипотез и теорий. И так как при анализе и прогнозировании явлений природы исследователь довольно часто сталкивается с многомерностью их описания, то одним из средств достижения цели является многомерный статистический анализ. В многомерном статистическом анализе анализируемые объекты удобно интерпретировать в качестве точек (векторов) в соответствующем признаковом пространстве. Естественно предположить, что геометрическая близость двух или нескольких точек в этом пространстве означает близость «физических» состояний соответствующих объектов, их однородность. Большинство тестов, используемых в педагогике и психологии, ориентируются на конечную сумму баллов, но, имея одинаковый результат сложения баллов, респонденты могут иметь различные направления развития своих качеств. При этом число признаков, по которым исследуются объекты, бывает достаточно большим. Перед исследователем, как правило, ставится задача оптимизировать признаковое пространство таким образом, чтобы исходные данные были легко визуализированы, информативны и не дублировали друг друга.

Рассмотрим использование кластерного и корреляционно-регрессионного анализов как методов многомерной статистики для снижения признакового пространства при исследовании объектов на Интернет-зависимость. Традиционные методики определения Интернет-зависимости – это опросные методики. При таком методе исследования испытуемому предлагается ответить на ряд вопросов, представленных в анкете. Одним из наиболее распространенных является тест Кимберли Янг Internet Addiction Test (IAT) из 20 вопросов. [1] Для диагностики Интернет-зависимости в нашем исследовании был использован тест Л. Н. Юрьевой и Т. Б. Большот [2] (это адаптированный к нашей социально-общественной сфере деятельности тест Кимберли Янга), содержащий 11 вопросов, ответы на которые оценивались баллами от 1 до 4 и рассматривались как признаки. В эксперименте участвовала группа из 17 курсантов 2 курса. На основе этого была создана матрица «объект-признак». Используя евклидову метрику и метод Варда [1], получили первичное разбиение группы на кластеры (рис 1).

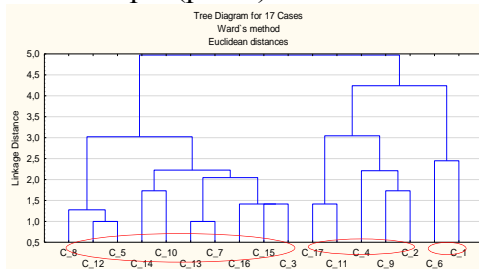


Рис. 1. Первичное разбиение группы на классы Интернет-зависимых объектов

Наблюдается три группы однородных объектов, которые условно, согласно авторам [2], можно охарактеризовать как: (C_1, C_6) – это группа подверженных риску развития компьютерной зависимости, (C_17, C_11, C_4, C_9, C_2) – находятся на стадии увлеченности, остальные имеют 0% риска развития компьютерной увлеченности. Наша задача состояла в том, чтобы сохранить данное разбиение, уменьшив при этом количество признаков. Для достижения поставленной цели были проклассифицированы признаки, к которым также были применены методы кластерного анализа. В результате все признаки разбились на два класса. Затем к признаковому пространству применили корреляционный анализ с тем, чтобы выявить коррелируемые признаки. На основе анализа корреляционной матрицы из каждой группы были отобраны признаки 4, 5, 8, 10, которые не коррелируют ни с какими признаками, и 2, 7, 9, которые мало коррелируют со всеми, а с отобранными ранее не коррелируют совсем. В итоге получилось аналогичное первому разбиению респондентов на классы, но при использовании меньшего количества признаков (рис 2).

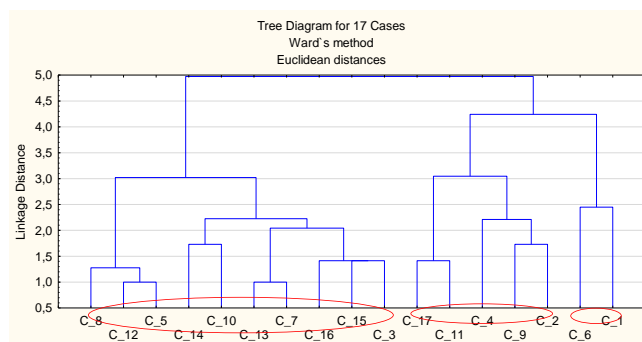


Рис. 2. Вторичное разбиение группы на классы Интернет-зависимых объектов, с использованием меньшего числа признаков

Подтверждением результатов разбиения послужили аналогичные кластеры, полученные методом k-средних.

Таким образом, вместо 11 вопросов, проверяющих респондентов на Интернет-зависимость можно использовать 7, получив при этом те же результаты. В дальнейшем, выявив из этих 7 признаков еще более информативные и создав функцию классификации, можно соотнести неизвестных респондентов к соответствующему классу.

Библиографический список

1. Айвазян, С.А. Прикладная статистика и основы эконометрики: учебник для вузов / С.А. Айвазян, Мхитарян В.С. – М. : ЮНИТИ-ДАНА, 1998. – 1022 с.
2. Internet Addiction: A Handbook and Guide to Evaluation and Treatment /Ed. by K.S. Young? C.N. Abreu. – Hoboken, New Jersey: John Wiley & Sons, 2011.
3. Юрьева, Л.Н. Компьютерная зависимость: формирование, диагностика, коррекция и профилактика: монография / Л.Н. Юрьева, Т.Ю. Больбот – Днепропетровск: Порогт, 2006. – 196 с