

ИСПОЛЬЗОВАНИЕ ДИСКРИМИНАНТНОГО АНАЛИЗА ДЛЯ ПРОГНОЗА УДОБОЧИТАЕМОСТИ ШРИФТОВ

О.В. Токарь

Белорусский государственный технологический университет, Минск

Представлен результат применения дискриминантного анализа для выявления геометрических параметров, влияющих на удобочитаемость шрифта как одного из показателей качества восприятия печатных материалов, а также приведены функции, позволяющие прогнозировать удобочитаемость шрифтов только на основании их геометрии.

Известно, что удобочитаемость шрифтов и качество шрифтового оформления печатных и электронных источников информации связаны с качеством ее восприятия и усвоения читателем. Поэтому актуальна разработка новых методик для установления удобочитаемости материала и параметров, позволяющих ее регулировать. Для достижения этой цели может быть использован дискриминантный анализ, включающий несколько тесно связанных статистических процедур, которые можно разделить на методы интерпретации межгрупповых различий – дискриминации и методы классификации наблюдений по группам [1].

Исходная информация представляет собой данные о шрифтах, относящихся к группе с высокой и низкой удобочитаемостью. В данном случае удобочитаемость определялась методом парных сравнений. Для каждого шрифта рассчитаны геометрические параметры: отношение кегля шрифта к высоте знака, пропорциональность, контрастность и т. д. Методика оценки удобочитаемости и описание геометрии шрифтов приведены в [2].

Поскольку число дискриминантных функций меньше числа групп на единицу, в данном случае при помощи пакета Statgraphics была получена одна дискриминантная функция. Значение $\lambda = 0,21$ в интервале от 0 до 1 свидетельствует о неплохой дискриминации, т. е. о том, что задействованные переменные эффективно участвуют в различении групп. Большая величина канонической корреляции (0,89) соответствует тесной связи дискриминантной функции с группами. χ^2 имеет большое значение (27,0), что говорит о хорошей дискриминантивной способности данной функции.

Интерпретации различия между группами шрифтов с высокой и невысокой удобочитаемостью способствуют коэффициенты дискриминантной функции (табл. 1).

Таблица 1

Коэффициенты дискриминантной функции

Параметр	Коэффициент	Параметр	Коэффициент
Пропорциональность «н»	10,18	Контрастность «а»	-2,72
Контрастность «н»	1,85	Max / min «А»	-0,73
Засечки «н»	-3,55	Кегль / высота «а»	-16,22
Площадь «н»	-0,03	Штрих / просвет «а»	14,48
Периметр «н»	1,15	Площадь «а»	-0,0003
Кегль / высота «н»	18,28	Периметр «а»	-0,09
Штрих / просвет «н»	-3,77	Константа	-122,89
Пропорциональность «а»	-40,02	—	—

Наибольший вклад в дискриминацию вносят параметры пропорциональности и отношения кегля к высоте для знака «н», а также отношение основного штриха к внутрибуквенному просвету знака «а», т. е. эти параметры влияют на удобочитаемость шрифта.

Точность классификации шрифтов достаточно высокая (табл. 2). Из 26 шрифтов в группе с высокой удобочитаемостью изначально находятся 11 объектов, с низкой – 15. Точность распознавания объектов первой группы составила 100 % (распознаны все 11 шрифтов), второй – 93 % (из 15 шрифтов распознаны 14).

Таблица 2
Классификационная таблица

Исходная группа	Предсказанная группа		Итого
	1	2	
1	11 (100 %)	0	11
2	1 (7 %)	14 (93 %)	15

Для классификации новых объектов, т. е. прогнозирования удобочитаемости новых шрифтов только лишь по геометрическим параметрам, были получены следующие классификационные функции:

$$F_1 = -143389,0 + 3532,27 \cdot X_1 - 570,90 \cdot X_2 - 822,60 \cdot X_3 - 22,72 \cdot X_4 + 926,64 \cdot X_5 + 21773,7 \cdot X_6 + 1103,7 \cdot X_7 - 7104,98 \cdot X_8 - 371,21 \cdot X_9 - 63,0 \cdot X_{10} - 6757,98 \cdot X_{11} + 3415,81 \cdot X_{12} - 18,37 \cdot X_{13} + 716,99 \cdot X_{14};$$

$$F_2 = -143859,0 + 3571,28 \cdot X_1 - 563,80 \cdot X_2 - 836,19 \cdot X_3 - 22,82 \cdot X_4 + 931,04 \cdot X_5 + 21843,4 \cdot X_6 + 1089,25 \cdot X_7 - 7258,3 \cdot X_8 - 381,62 \cdot X_9 - 65,78 \cdot X_{10} - 6820,11 \cdot X_{11} + 3471,3 \cdot X_{12} - 18,39 \cdot X_{13} + 716,66 \cdot X_{14},$$

где X_1 – пропорциональность «н»; X_2 – контрастность «н»; X_3 – засечки «н»; X_4 – площадь «н»; X_5 – периметр «н»; X_6 – отношение кегля к высоте знака «н»; X_7 – отношение основного штриха к внутрибуквенному просвету «н»; X_8 – пропорциональность «а»; X_9 – контрастность «а»; X_{10} – отношение максимальной ширины штриха к минимальному для знака «а»; X_{11} – отношение кегля к высоте знака «а»; X_{12} – отношение основного штриха к внутрибуквенному просвету «а»; X_{13} – площадь «а»; X_{14} – периметр «а».

Полученный результат вполне удовлетворительный и позволяет предсказывать качество для вновь разрабатываемых или тестируемых шрифтов, являющихся средствами передачи информации как в электронном, так и печатном виде.

Список литературы

1. Дюк, В. Информационные технологии в медико-биологических исследованиях / В. Дюк, В. Эмануэль. – СПб. : Питер, 2003. – 528 с.
2. Токарь, О.В. Удобочитаемость современных текстовых шрифтов : науч. изд. / О.В. Токарь. – Минск : Современная школа, 2007. – 192 с.