

© О. В. Токарь, к.т.н., УО «Белорусский государственный технологический университет», Минск, Республика Беларусь

УРАВНЕНИЕ РЕГРЕССИИ ДЛЯ ПРОГНОЗА УДОБОЧИТАЕМОСТИ ШРИФТА НА ОСНОВЕ ЕГО ХАРАКТЕРИСТИК

В статье излагается опыт построения уравнения на основе регрессионно-корреляционного анализа, связывающего ряд понятий, отражающих характеристики шрифта, например, «изящный», «плотный», «легкий», «выразительный», «удобный», «читаемый», «красивый», а также высокую или низкую удобочитаемость, которое может использоваться для прогнозирования качества шрифта.

Ключевые слова: шрифт; регрессионный анализ; удобочитаемость; восприятие; прогноз.

Постановка проблемы

Разработка прогнозирования удобочитаемости шрифта является актуальным инструментарием для получения новых знаний о качестве шрифта, как средства передачи информации. Использование различных методов помогает сформулировать целостное содержание понятия качества для данного объекта.

Для выявления зависимости удобочитаемости шрифта от нескольких факторов может быть использован регрессионный анализ. Для разработки уравнения регрессии, позволяющего прогнозировать удобочитаемость шрифта как изучаемый показатель, необходимо отобрать ряд факторов, влияющих на искомый показатель.

В данной работе излагается предположение, что качество шрифтов может определяться через восприятие потребителя-

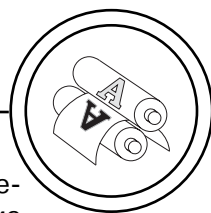
ми информации специальных характеристик, помогающих шрифту выполнять его функциональное предназначение. Данные, количественно отражающие восприятие пользователями этих характеристик, предлагаются в качестве независимых переменных.

Цель работы

Цель работы заключается в разработке уравнения регрессии, описывающего зависимость удобочитаемости от ряда понятий, выражающих восприятие шрифта его потребителями.

Результаты проведенных исследований

Для количественного оценивания характеристик шрифта были использованы 11 понятий: изящный, удобный, плотный, легкий, читаемый, красивый, стильный, емкий, выразительный, привлекательный, стройный. Обос-



нование выбора понятий приведено в источнике [1]. Оптимальным вариантом выбрано 3-шкальное деление (1 — испытуемый согласен с предложенным признаком, 2 — не совсем согласен, 3 — не согласен).

Путем опроса пользователей шрифта (читателей) были установлены количественные значения данных понятий для 23 шрифтов: 1. BookmanCTT; 2. Bell Gothic CTT BT; 3. Eras LightCTT; 4. Geo Slb712CTT Md BT; 5. KabelCTT Medium; 6. KisSCCTT BT; 7. New BaskervilleCTT; 8. NewtonCTT; 9. OCRF-RegularCTT; 10. Octava CTT; 11. RaleighCTT BT; 12. Orenburg; 13. Parangon 330C; 14. Univers Cyr; 15. CooperCTT Lt BT; 16. Franklin GothicBookCTT; 17. TextBookCTT; 18. Gothic725CTT Bd BT; 19. Humanist531CTT BT; 20. ZapfElliptical711CTT BT; 21. SabonCyr; 22. OfficinaSansMedium CTT; 23. Original GaramondSCCTT BT.

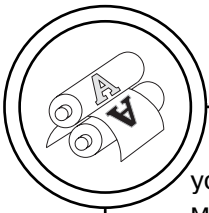
В качестве зависимой переменной определено разделение шрифтов на группу с высокой (группа 1) и низкой (группа 2) удобочитаемостью, определенной по методу парных сравнений [2]. Частично зависимая и независимые переменные представлены в табл.

Известно, что регрессия изучает форму связи и позволяет выявить аналитическую зависимость между признаками и представить в виде уравнения [3]. С помощью регрессионного анализа может быть установлена теснота связи между показателями, т. е. устанавливается следующее:

- 1) степень зависимости резуль- тативного показателя от каж- дого фактора (оценивается с по- мощью коэффициента корреля- ции);
- 2) характер связи, т. е. по зна- чению коэффициента регрессии

Зависимая и независимые переменные

Номер шрифта	Изящный	Удобный	Плотный	Легкий	Читаемый	Красивый	Стильный	Емкий	Выразительный	Привлекательный	Стройный	Группы удобо- читаемости
1	1,75	1,83	2,75	1,90	1,73	1,55	1,53	2,38	1,45	1,55	2,48	1
2	1,97	2,45	1,83	2,55	2,45	2,21	1,97	2,24	2,10	2,31	1,72	2
3	2,30	2,53	3,00	2,03	2,20	1,88	1,65	2,55	1,95	1,85	2,63	2
4	2,35	2,35	2,23	2,58	2,30	1,85	1,60	2,40	1,45	1,55	2,33	1
5	2,28	2,60	2,25	2,58	2,33	1,75	1,53	2,45	1,60	1,73	2,33	2
...
20	1,86	1,66	2,72	1,62	1,45	1,62	1,66	2,34	1,55	1,48	2,34	1
21	1,70	1,44	2,59	1,26	1,33	1,59	1,93	2,30	1,56	1,63	2,15	1
22	2,30	2,44	1,74	2,44	2,15	2,33	2,07	1,93	1,81	1,96	1,74	2
23	2,27	2,67	2,78	2,29	2,58	1,98	1,84	2,51	2,18	2,13	2,44	2



устанавливается насколько изменяется величина результирующего показателя при изменении факторного показателя.

Ниже представлено уравнение, отражающее зависимость удобочитаемости (группа) от изменения предложенных факторов (X):

$$Y = 2,01823 - 0,594218 \cdot X_1 + 0,132001 \cdot X_{10} + 0,376137 \cdot X_{11} + 1,31772 \cdot X_2 - 0,768664 \cdot X_3 - 2,14906 \cdot X_4 + 1,10646 \cdot X_5 + 1,70357 \cdot X_6 - 0,597095 \cdot X_7 + 0,143383 \cdot X_8 - 0,818031 \cdot X_9,$$

где Y — группа удобочитаемости; X₁ — понятие «изящный»; X₂ — понятие «удобный»; X₃ — понятие «плотный»; X₄ — понятие «легкий»; X₅ — понятие «читаемый»; X₆ — понятие «красивый»; X₇ — понятие «стильный»; X₈ — понятие «емкий»; X₉ — понятие «выразительный»; X₁₀ — понятие «привлекательный»; X₁₁ — понятие «стройный».

Доля влияния факторов, включенных в модель, на результирующий показатель устанавливается с помощью коэффициента множественной детерминации (R²). В данной модели R² = 85,4 %, т. е. на 85,4 % независимые переменные (выбранные понятия) объясняют распределение шрифтов на две группы. Этот результат допустим для анализа данных. Также в уравнении присутствует свободный член, который не зависит от факторов X.

Поскольку коэффициент корреляции показывает, что наблюдается слабая связь между группой удобочитаемости и понятиями «стильный» (R = 0,16), «емкий» (R = -0,06), «стройный» (R = -0,02), эти понятия реко-

мендовано исключить из числа независимых переменных модели. В свою очередь, так как связь между понятиями «красивый» и «привлекательный» (R = 0,87) очень велика, то последнее понятие также можно исключить, поскольку его связь (R = 0,48) с результирующим показателем ниже, чем у понятия «красивый» (R = 0,53).

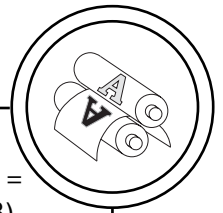
После исключения данных факторов уравнение имеет следующий вид:

$$Y = 1,44751 - 0,34048 \cdot X_1 + 1,31679 \cdot X_2 - 0,466306 \cdot X_3 - 1,85551 \cdot X_4 + 1,11346 \cdot X_5 + 1,09313 \cdot X_6 - 0,734826 \cdot X_9.$$

Расчеты показали, что для данной модели R² = 82,2 %, т. е. при исключении из модели факторов X₇, X₈, X₁₀, X₁₁, коэффициент детерминации снизился только на 3,2 %. Значение критерия Фишера (F = 9,87) больше табличного (F_{табл.} = 2,93), поэтому уравнение статистически значимо. Их двух полученных уравнений для практического применения более ценно второе, как не имеющее лишних факторов при прочих практически равных условиях.

R-значения почти для всех независимых переменных не превышают стандартного уровня значимости 0,1. Пограничные значения имеет данный параметр для фактора «изящный» и свободный член уравнения (1,44751), показывающий, что на значение анализируемого параметра влияют и другие факторы, не описанные в модели.

Коэффициенты регрессии уравнения показывают, что более



всего изменяется значение резуль- тативного показателя в зави- симости от фактора «легкий» (X_4), а менее всего от фактора «изящный» (X_1), хотя разница и не очень велика. Т.е. в рамках данной модели на степень удо- бочитаемости степень легкости, свойственной шрифту, влияет с весом — 1,85551, а степень изящности с весом — 0,34048.

Знак «минус» указывает на отрицательное влияние, а знак «плюс» — на положительное. Знаком «минус» отмечены коэф- фициенты уравнения при факто- рах «изящный», «плотный», «лег- кий», «выразительный», поэтому, чем меньше численно выраже- ны эти факторы, тем выше будет удобочитаемость. Знаком «плюс» отмечены коэффициен- ты при факторах «удобный», «читаемый», «красивый», поэто- му, чем больше выражены эти факторы, тем выше будет удо- бочитаемость.

Можно сделать вывод, что уравнение показывает связь между высокими показателями шрифтов по факторам «удоб- ный», «читаемый», «красивый» и классификацией шрифтов в группу с высокой удобочитае- мостью. В эту же группу попа- дают шрифты со слабо выражен- ными факторами «плотный», «изящный», «легкий», «вырази- тельный». Последние три факто- ра можно отнести к эстетиче- ской характеристике шрифта, а фактор «плотный» — к емкости.

Для прогнозирования удобо- читаемости нового абстракт- ного шрифта были с помощью опроса определены значения факторов «изящный» ($X_1 = 1,65$), «удобный» ($X_2 = 1,95$), «плот-

ный» ($X_3 = 2,45$), «легкий» ($X_4 = 1,80$), «читаемый» ($X_5 = 1,73$), «красивый» ($X_6 = 2,01$), «вырази- тельный» ($X_9 = 1,59$). Подста- новка этих данных во второе уравнение дает результат 1,93, который позволяет отнести дан- ный объект во вторую группу с низкой удобочитаемостью.

Следует отметить, что изна- чальное деление шрифтов на две группы вместо использования рангового списка в качестве зависимой переменной делает результат прогноза приблизи- тельным.

Выводы

В результате опыта регрес- сионного анализа данных, по- лученных путем опроса пользо- вателей шрифта по предложен- ной методике выявлено, что:

1) Может быть получено ста- тистически значимое регрессион- ное уравнение, связывающее ряд понятий, отражающих характери- стики шрифта, и группы с высо- кой и низкой удобочитаемостью, определенной отдельно методом парных сравнений.

2) Согласно алгоритму из уравнения на основе корреля- ционного анализа могут быть исключены следующие незави- симые переменные, мало влия- ющие на конечный результат: «стильный», «емкий», «строй- ный», «привлекательный».

3) Выявлено, что прямо про- порциональную связь с зависи- мой переменной имеют факторы «удобный», «читаемый», «краси- вый», причем вес коэффициен- та в рамках данной модели для них средний.

4) Уравнение может исполь- зоваться для прогнозирования



при наличии данных, полученных в результате опроса по изложенной в статье методике.

Данная работа расширяет методологическое поле оценки количественными методами качества шрифта, доказывает воз-

можность построения уравнений регрессии, прогнозирующих удобочитаемость не только по геометрическим параметрам шрифта [4], но и на основе данных из области восприятия шрифта пользователями.

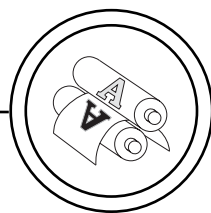
Список использованной литературы

1. Токарь О. В. Оценка восприятия печатных шрифтов / О. В. Токарь // Технологія і техніка друкарства : збірник наукових праць. — Київ : НТУУ «КПІ». — 2015. — № 3(49). — С. 39–46. — Режим доступу : <http://ttdruk.vpi.kpi.ua/article/view/54876>.
2. Токарь О. В. Комплексная оценка удобочитаемости современных типографских шрифтов на допечатной стадии полиграфического производства : автореф. дис. ... канд. тех. наук : 05.02.13 / МГУП. — М., 2006. — 22 с.
3. Ускевич Т. Г. Методические рекомендации по проведению корреляционно-регрессионного анализа в Microsoft Excel. — Минск : БГЭУ, 2010. — 20 с.
4. Токарь О. В. Прогнозирование удобочитаемости шрифта методом дискриминантного анализа / О. В. Токарь // Технологія і техніка друкарства : збірник наукових праць. — Київ : НТУУ «КПІ». — 2016. — № 2(52). — С. 13–19. — Режим доступу : <http://ttdruk.vpi.kpi.ua/article/view/59897>.

References

1. Tokar', O. V. (2015). Otsenka vospriyatiya pechatnyih shriftov [Evaluation of the perception of typefaces]. *Journal of Tekhnolohiia i tekhnika drukarstva — Technology and Technique of Typography*, 3, 39–46. Retrieved from <http://ttdruk.vpi.kpi.ua/article/view/54876> [in Russian].
2. Tokar', O. V. (2006). *Kompleksnaja ocenka udobochitaemosti sovremennyh tipografskikh shriftov na dopechatnoj stadii poligraficheskogo proizvodstva* [Comprehensive assessment of the readability of modern typographic fonts for printing production]. Moscow: MGUP [in Russian].
3. Ushkevich', T. G. (2010). *Metodicheskie rekomendatsii po provedeniyu korrelyatsionno-regressionnogo analiza v Microsoft Excel* [Guidelines for conducting the correlation and regression analysis in Microsoft Excel]. Minsk: BSEU [in Russian].
4. Tokar', O. V. (2016). Prognozirovaniye udobochitaemosti shrifta metodom diskriminantnogo analiza [The prediction of the readability of the font by the method of discriminant analysis]. *Journal of Tekhnolohiia i tekhnika drukarstva — Technology and Technique of Typography*, 2, 13–19. Retrieved from <http://ttdruk.vpi.kpi.ua/article/view/59897> [in Russian].

У статті викладено досвід побудови рівняння на основі регресійно-кореляційного аналізу, що пов'язує ряд понять, що відображають характеристики шрифту, наприклад, «витончений», «щільний», «легкий», «виразний», «зручний», «читабельний», «гарний» і високу або низьку читабельність, які можуть використовуватися для прогнозування якості шрифту.



Ключові слова: шрифт; регресійний аналіз;
читабельність; сприйняття; прогноз.

The article describes the experience of building an equation based on the regression-correlation analysis linking a number of concepts reflecting the characteristics of a font, for example, 'elegant', 'dense', 'easy', 'expressive', 'comfortable', 'readable', 'beautiful', and a high or low readability, which can be used to predict the quality of the font.

Keywords: font; regression analysis; readability; perception; prediction.

Рецензент — О. В. Зоренко, к.т.н., доцент,
КПІ ім. Ігоря Сікорського

Надійшла до редакції 27.02.17