Нешитой Василий Васильевич
Петренко Борис Васильевич
Белорусский государственный университет культуры и искусств

МОДЕЛИ РАССЕЯНИЯ ПУБЛИКАЦИЙ

Математические модели, описывающие формализованный процесс функционирования системы информационных потоков, в состоянии охватить только основные, характерные ее закономерности. При этом необходимо иметь в виду, что системы информационных потоков имеют динамический характер и им свойственно множество состояний. Переход из одного состояния в другое также подчинен закономерностям, их описание не менее важно, чем познание системы в каком-либо из фиксированных состояний. Изучению потоков информации посвящено большое количество работ. При этом особое внимание уделяется ранговым распределениям и закону рассеяния публикаций.

Если упорядочить журналы по убыванию числа помещенных в них статей по данной тематике, то получим ранговое распределение. В первом приближении его можно описать законом Дж. К. Ципфа, который устанавливает связь между относительной частотой статей и порядковыми номерами в списке по убывающим частотам. Закон Ципфа по форме совпадает с формулировкой закона рассеяния публикаций С. Бредфорда в ранговой интегральной форме [5]:

$$X(r)=a+b\log r, \qquad (1)$$

где $X(r)$ – накопленное число статей в r первых журналах; a, b – параметры.

Таким образом, закон Бредфорда представляет собой не что иное, как закон Ципфа, примененный к периодическим изданиям. Другими словами, закон рассеяния в общем случае задается ранговым распределением. Отсюда следует вывод: для отыскания точной формулировки закона рассеяния публикаций необходимо найти такое теоретическое распределение, которое с высокой точностью описывает все разнообразие статистических ранговых распределений.

Свой закон С. Бредфорд проиллюстрировал кривой рассеяния: $X(r)=f(\ln r)$, которая на границе зоны ядра переходит в прямую (правда, неизвестно, как эту границу вычислять), и сформулировал его в следующем виде (цит. по: [1]): «Если научные журналы расположить в порядке уменьшения числа помещенных в них статей по какому-либо заданному предмету, то в полученном списке можно выделить ядро журналов, посвященных непосредственно этому предмету, и несколько групп, или зон, каждая из которых содержит столько же статей, что и ядро. Тогда числа журналов в ядре и последующих зонах будут относиться как $1 : N : N^2$». В исследованиях Бредфорда величина $N \approx 5$.

Последователи Бредфорда понимали, что закон рассеяния публикаций нуждается в уточнении, но не могли найти то ранговое распределение, которое могло бы с высокой точностью описывать статистические ранговые распределения журналов. Поэтому их «уточнения» не вносили ничего нового в закон рассеяния или даже искажали его. Модель С. Бредфорда пытались уточнить Б. Викери, Б. Брукс, А. Мицевич и другие, однако их модели далеки от реальных распределений.

Большинство исследователей вслед за Бредфордом принимали число статей в ядре журналов и зонах рассеяния одинаковым, что не согласуется с опытными данными. Так, Т. Викери предлагал вводить любое количество зон рассеяния с равным числом статей в ядре и зонах рассеяния. Ясно, что такой модели не соответствует ни одно теоретическое распределение.

Для описания статистических ранговых распределений необходимо использовать теоретические законы с убывающей плотностью, заданной на положительной полуоси, поскольку ранг журнала – положительное число.

Одним из простейших и подходящих законов для описания статистических ранговых распределений является закон В. Вейбулла. Функция распределения и плотность вероятности его задаются формулами $F(t)=1-e^{-\alpha t^{\beta}}$; $p(t)=\alpha\beta t^{\beta-1}e^{-\alpha t^{\beta}}$. При $\beta<1$ плотность $p(t)$ с ростом t убывает. Величина t в данном случае может обозначать ранг журнала, $F(t)$ – накопленную долю статей по заданной теме в t первых журналах, плотность $p(t)$ численно равна доле статей (из общего их числа) в журнале с рангом t.

Закон Вейбулла (так же, как и закон Ципфа) является частным случаем более общего закона рассеяния публикаций. Исследования показали, что закон Вейбулла в отличие от закона Ципфа может с высокой точностью описывать статистические ранговые распределения. Это позволяет выполнять все необходимые расчеты, касающиеся рассеяния публикаций. Более того, на основе закона Вейбулла можно получить математически точную формулировку закона рассеяния публикаций в смысле Бредфорда.

Кривая распределения Вейбулла при $0 < \beta < 1$ является убывающей и не имеет никаких особых точек, которые можно было бы использовать в качестве границ ядра и зон рассеяния. Для нахождения этих границ преобразуем ранговое распределение Вейбулла в к форме $tp(t) = f(\ln t)$:

$$tp(t) = \frac{\alpha\beta e^{\beta\ln t}}{e^{\alpha e^{\beta\ln t}}}.$$

Произведение $tp(t)$ представляет собой плотность распределения случайной величины $\ln T$ [2].

Приняв обозначения $\ln t = x$, $tp(t) = p(\ln t) = p(x)$, последнюю формулу перепишем в виде:

$$p(x) = \frac{\alpha\beta e^{\beta x}}{e^{\alpha e^{\beta x}}}.$$

Плотность $p(x)$ обладает тем замечательным свойством, что кривая распределения (т. е. график плотности $p(x)$) имеет три характерные точки: моду C и две точки перегиба A и B с абсциссами соответственно xA, xC, xB. При этом точки перегиба находятся на равных расстояниях от моды, т. е.

$$x_C - x_A = x_B - x_C.$$

Переходя к плотности $p(t)$, с учетом равенства $x = \ln t$, можем записать:

$$\ln t_C - \ln t_A = \ln t_B - \ln t_C.$$

Из последнего равенства имеем:

$$\frac{t_C}{t_A} = \frac{t_B}{t_C} = n, \qquad (2)$$

откуда следует математически точная формулировка закона рассеяния в смысле Бредфорда

$$t_A : t_C : t_B = t_A\,(1 : n : n^2). \qquad (3)$$

Эта формула отличается от закона Бредфорда тем, что величины t_A, t_C, t_B в левой ее части обозначают число журналов от начала частотного списка соответственно до точек А, С, В, а не в ядре и зонах рассеяния.

Если по закону Вейбулла рассчитать число журналов в ядре и зонах рассеяния, то соответствующая формула будет иметь вид:

$$t_A : t_I : t_{II} = t_A\,(1 : (n-1) : (n-1)n). \qquad (4)$$

Здесь t_A – число журналов в ядре; t_I, t_{II} – число журналов соответственно в первой и второй зонах рассеяния.

Формулировка закона рассеяния в смысле Бредфорда в виде формул (3, 4) была предложена авторами настоящей статьи еще в 1974 г. [4].

Закон Вейбулла позволяет получить формулу для расчета координат точек А, С, В, т. е. для расчета границ ядра и зон рассеяния и доли статей в них. Эти формулы имеют вид:

$$\left.\begin{aligned} & t_C = \left(\frac{1}{\alpha}\right)^{1/\beta};\ t_A = \left(\frac{3-\sqrt{5}}{2\alpha}\right)^{1/\beta};\ t_B = \left(\frac{3+\sqrt{5}}{2\alpha}\right)^{1/\beta};\ n = \frac{t_B}{t_C} = \frac{t_C}{t_A} = \left(\frac{3+\sqrt{5}}{2}\right)^{1/\beta} \\ & F(t_C) = 1 - \frac{1}{e} = 0.6321;\ F(t_A) = 1 - e^{-(3-\sqrt{5})/2} = 0.3175;\ F(t_B) = 1 - e^{-(3+\sqrt{5})/2} = 0.9271 \end{aligned}\right\} \qquad (5)$$

Итак, в случае справедливости закона Вейбулла в ядро журналов входит примерно 32% от всех статей по данному предмету, в ядро и первую зону – 63%, а в ядро и первые две зоны – 93%. Следовательно, на первую зону приходится 31% статей, на вторую – 30%, а на третью лишь 7% статей. Как видно из приведенных формул, они зависят от двух параметров закона Вейбулла – α и β.

Величина t_B может характеризировать оптимальный объем фонда по заданной тематике с точки зрения полноты комплектования. Чтобы увеличить полноту комплектования статей, например, на 5% выше оптимального значения $F(t_B) \approx 0{,}93$, необходимо увеличить количество наименований журналов примерно в 2 раза (при $\alpha = 0{,}1$; $\beta = 0{,}5$), в то время как уменьшение полноты комплектования по статьям на 5% ниже значения $F(t_B)$ приводит к уменьшению количества наименований журналов в 1,5 раза.

Отсюда следует, что для более полного удовлетворения информационных потребностей специалистов в справочно-информационном фонде должны комплектоваться по крайней мере те журналы, которые образуют ядро и первые две зоны рассеяния, где содержится около 93% статей.

Закон Вейбулла позволяет также находить величину t при любых значениях $F(t)$:

$$t = \left(\frac{1}{\alpha} \ln \frac{1}{1 - F(t)} \right)^{1/\beta}.$$

Таким образом, распределение Вейбулла позволило получить математически точную формулировку закона рассеяния в смысле Бредфорда, а также формулы для вычисления доли статей в ядре и зонах рассеяния. Однако это распределение представляет собой частный случай некоторого более общего семейства распределений. Если ранговое распределение журналов не подчиняется закону Вейбулла, то для вычисления границ ядра, зон рассеяния и доли статей в них полученные формулы оказываются непригодными. Вот почему усилия многих исследователей не привели к нахождению общих формул для вычисления границ ядра и зон рассеяния, а также вычисления доли статей в них. Для этого требовалась разработка **теории обобщенных** (универсальных) **распределений.**

Такая теория была создана (ее автор В. Нешитой). Она включает три системы непрерывных распределений, заданные четырехпараметрическими плотностями, систему дискретных распределений, взаимосвязанную с системой кривых роста новых событий, методы вычисления типа выравнивающей кривой и точечных оценок параметров (универсальный метод моментов и общий устойчивый метод), номограммы для графического определения типа выравнивающей кривой и оценок параметров и серию компьютерных программ для работы с указанными системами [2].

С точки зрения теории обобщенных распределений универсальным законом рассеяния публикаций является вторая система непрерывных распределений, заданная тремя плотностями [2]:

$$\left.\begin{aligned}
p(t) &= Nt^{k\beta-1}\left(1-\alpha u t^{\beta}\right)^{1/u-1}\\
p(y) &= \frac{N(\ln y - l)^{k-1}}{y}\left[1-\alpha u(\ln y - l)\right]^{1/u-1}\\
p(y) &= \frac{N}{y}\left[1-\alpha u\left(\ln y-\overline{\ln y}\right)^{2}\right]^{1/u-1}
\end{aligned}\right\} \tag{6}$$

Рассмотренное выше распределение Вейбулла следует из первой плотности при $\kappa = 1,\ u \to 0$.

Если ранговое распределение журналов описывается первой плотностью, то границы ядра и зон рассеяния вычисляются по формулам [3]:

$$t_C = \left(\frac{k}{\alpha(1+ku-u)}\right)^{1/\beta}; \quad t_A = t_C/n; \quad t_B = t_C \cdot n, \tag{7}$$

где величина n равна:

$$n = \left[1+\frac{1-u+\sqrt{\left[4k(1+ku-u)+(1-u)\right](1-u)}}{2k(1+ku-u)}\right]^{1/\beta}. \tag{8}$$

Доли статей в ядре и зонах рассеяния вычисляются с помощью компьютерной программы через функцию распределения.

Формулировка закона рассеяния публикаций в смысле Бредфорда в общем случае остается прежней, т. е. задается формулами (3) и (4), полученными на базе распределения Вейбулла.

Вторая система непрерывных распределений описывает практически все многообразие статистических ранговых распределений, позволяет вычислять границы ядра и зон рассеяния, доли статей не только в ядре и зонах рассеяния, но и для любого числа журналов от начала частотного списка, а также величину n, входящую в формулы (3) и (4).

Поэтому вторая система непрерывных распределений, заданная тремя плотностями, по праву может считаться универсальным законом рассеяния публикаций [3].

ЛИТЕРАТУРА

1. **Михайлов, А. И.** Основы информатики / А. И. Михайлов, А. И. Черный, Р. С. Гиляревский. – Москва : Наука, 1968. – С. 93.

2. **Нешитой, В. В.** Элементы теории обобщенных распределений : монография / В. В. Нешитой. – Минск : РИВШ, 2009. – 204 с.

3. **Нешитой, В. В.** Универсальные законы рассеяния и старения публикаций / В. В. Нешитой // Веснік Бел. дзярж. ун-та культ. і маст. – 2007. – № 8. – С. 128–133.

4. **Петренко, Б. В.** Применение закона Вейбулла для расчета полноты комплектования справочно-информационного фонда / Б. В. Петренко, В. В. Нешитой // Проблемы оптимального комплектования и использования справочно-информационного фонда для принятия решений / Общество «Знание» Украинской ССР. – Киев, 1974. – С. 6–8.

5. **Bradford, S. C.** Documentation. – London, 1948. – 156 p.

В статье анализируются различные формулировки закона рассеяния публикаций С. Бредфорда по периодическим изданиям, дается математически точная его формулировка и предлагается универсальный закон рассеяния публикаций.

Various formulations of S. Bredford's law of publications dispersion are analyzed in the article by the periodic editions, its precise mathematic formulation is offered and a universe law of the publications dispersion is proposed.