

МЕЖДУНАРОДНЫЙ СОЛОМОНОВ УНИВЕРСИТЕТ



Дмитрий Владимирович ЛАНДЭ

Лекция 6

“Математические модели информационных потоков”



Моделирование информационных потоков

Общий характер временной зависимости числа тематических публикаций в сети определяется закономерностями, которые целиком допускают построение математических моделей.

Модель, аналогичная модели Бартона-Кеблера, учитывает статическую и динамическую составляющие от общих объемов сообщений по заданной тематике с учетом старения информации:

$$v(T) = 1 - ae^{-T} - be^{-2T}.$$

Баланс тем

Организации-генераторы новостной информации производят поток информации, в среднем постоянный по количеству сообщений.

Изменяются во времени лишь объемы сообщений, которые соответствуют той или другой теме. Таким образом, рост количества публикаций по одной теме сопровождается уменьшением публикаций по другим темам:

$$\int_0^T \sum_{i=1}^M n_i(t) dt = NT,$$

где $n_i(t)$ – количество сообщений по теме i в момент времени t , а M – общее количество всех возможных тем.



Линейная модель

В некоторых случаях динамика тематических информационных потоков реализуется линейно, то есть количество сообщений в момент времени t можно представить формулой:

$$y(t) = y(t_0) + v(t - t_0),$$

где $y(t)$ – количество сообщений на время t , v – средняя скорость увеличения (уменьшения) интенсивности тематического информационного потока во времени.

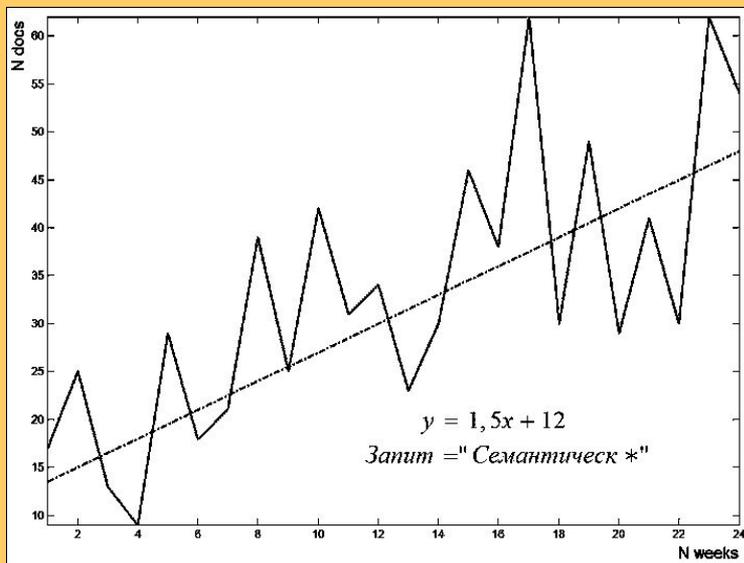
Содержательная составляющая информационного потока может быть оценена как флюктуация информационного потока – изменение стандартного отклонения $\sigma(t)$:

$$\sigma(t_i) = \sqrt{\frac{1}{i} \sum_{k=0}^i \{y(t_k) - (y(t_0) + v(t_i - t_0))\}^2}.$$

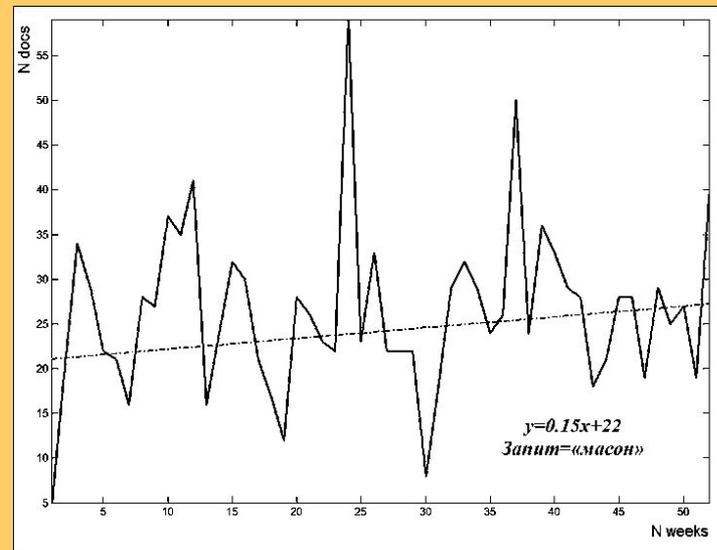
В случае поведения стандартного отклонения $\sigma(t) \propto t^\mu$, то чем большее значение μ , тем выше корреляция между текущими и предыдущими сообщениями. В этих случаях μ характеризует степень связи между случайными событиями и принимает значение от $\frac{1}{2}$ до 1.



Примеры, для которых линейная модель адекватна



*Динамика количества
откликов на запрос
«семантическ*»*



*Динамика появления
документов в
информационном потоке,
содержащих слово «масон»*



Экспоненциальная модель

В некоторых случаях процесс увеличения (роста) актуальности или старения информации описывается экспоненциальной зависимостью, которую можно аппроксимировать такой формулой:

$$N(t) = N(t_0)e^{\lambda(t-t_0)},$$

где λ - среднее относительное изменение интенсивности информационного потока.

Относительное изменение интенсивности в определенный момент времени исчисляется по формуле:

$$\lambda(t_i) = (N(t_i) - N(t_{i-1})) / N(t_{i-1}).$$

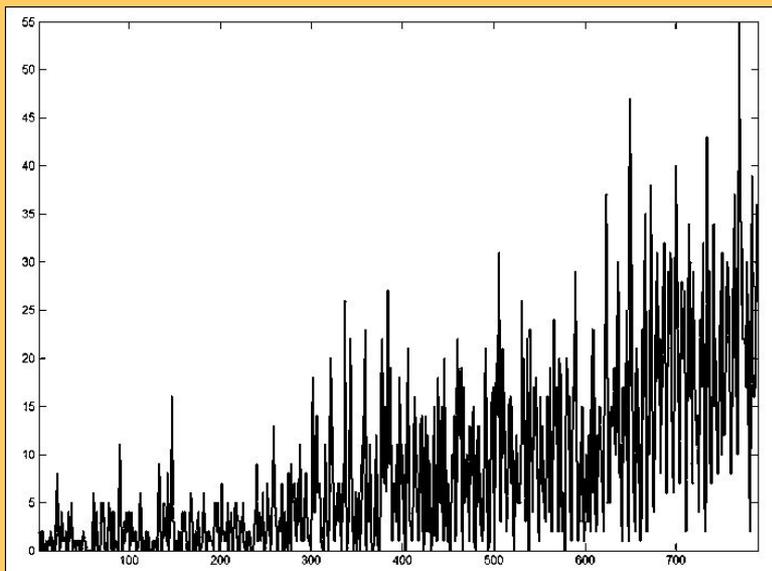
Изменение флуктуаций величины $\lambda(t_i)$ относительно среднего значения может быть оценена по формуле:

$$\sigma(t_i) = \sqrt{\frac{1}{i} \sum_{k=0}^i \{\lambda(t_k) - \lambda\}^2}.$$

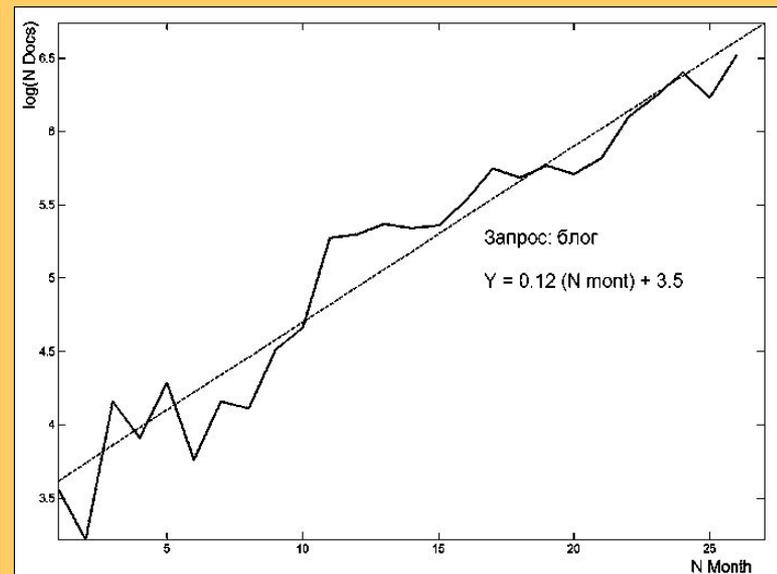
Если $\sigma(t)$ изменяется как корень квадратный из времени, то можно говорить о процессе с независимыми приращениями. В случае наличия значительной доли зависимых сообщений справедливо: $\sigma(t) \propto t^\mu$, причем $\mu < 1$. Значение $\mu \gg 1/2$, говорит о наличии долгосрочной памяти системы.



Пример, для которого экспоненциальная модель адекватна



*Посуточный график
появления сообщений,
содержащих термин «блог»*



*Помесячный график в
полулогарифмической шкале
появления сообщений,
содержащих термин «блог»*



Логистическая модель

Логистическую модель можно рассматривать как обобщение экспоненциальной модели Мальтуса, которая, предусматривает пропорциональность скорости роста функции ее значения в каждый момент времени:

$$\frac{dn(t)}{dt} = kn(t),$$

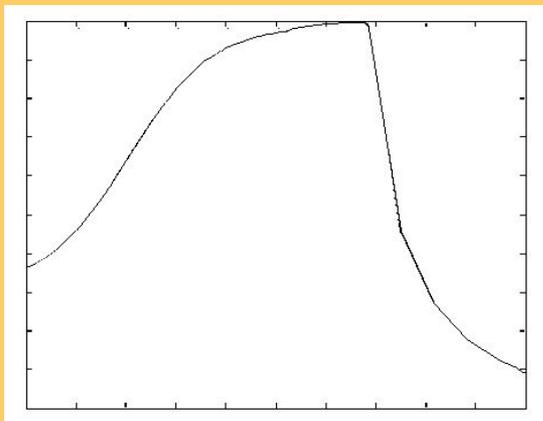
где k – некое

В случае логистической модели идея заключается в том, чтобы сделать коэффициент в уравнении Мальтуса функцией времени. Наиболее распространенным является использование константы, которая в явном виде ограничивает рост решения. В нашем случае с этой целью используем емкость N . Тогда правая часть соответствующего выражения представляется в виде:

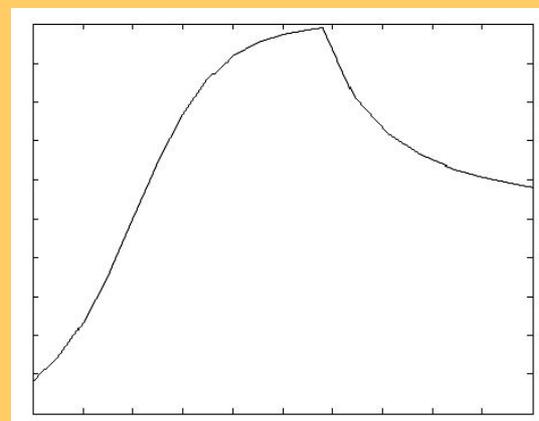
где k – коэффициент, который описывает отрицательные процессы, связанные с внутренними факторами.



Логистическая модель: примеры



Динамика объемов публикаций в Интернет по тематике болезни и отхода от деятельности известного политического деятеля



Динамика объемов публикаций в Интернет с упоминанием фамилии сенсационно избранного мэра большого города (до выборов и после)



Логистическая модель: детализация

На формальном уровне сопоставим с темой два параметра: продолжительность (характерное “время жизни”) λ и интенсивность D .

Продолжительность - промежуток времени, в течение которого тема имеет выраженную актуальность. Интенсивность - величина, которая характеризует порожденное соответствующей темой количество публикаций, усредненное по промежутку λ .

Вклад интенсивности D определяется следующим образом:

$$y(t) = \begin{cases} D, 0 < t \leq \lambda \\ 0, t < 0, t > \lambda \end{cases} \text{ в области:}$$

Соответственно, рассм $0 < t \leq \lambda$ с $D > 0$ и $t > \lambda$ с $D = 0$, для которых решениями являются функции $u(t)$ и $v(t)$. Полное решение получается путем “сшивки” на границе в точке λ :

$$n(t) = \begin{cases} u(t), 0 < t \leq \lambda \\ v(t), t > \lambda \end{cases}$$

$$u(\lambda) = v(\lambda)$$



Логистическая модель: уравнения

После нормирования параметров пороговой величины N , уравнение для первой области имеет вид:

$$\frac{du(t - \tau)}{dt} = pu(t - \tau)(1 - qu(t - \tau)) + Du(t - \tau),$$

Решение этого уравнения:

$$u(t) = \frac{u_s}{1 + \left(\frac{u_s}{n_0} - 1\right) \exp[-(p + D)(t - \tau)]}$$

Уравнение для второй области имеет вид:

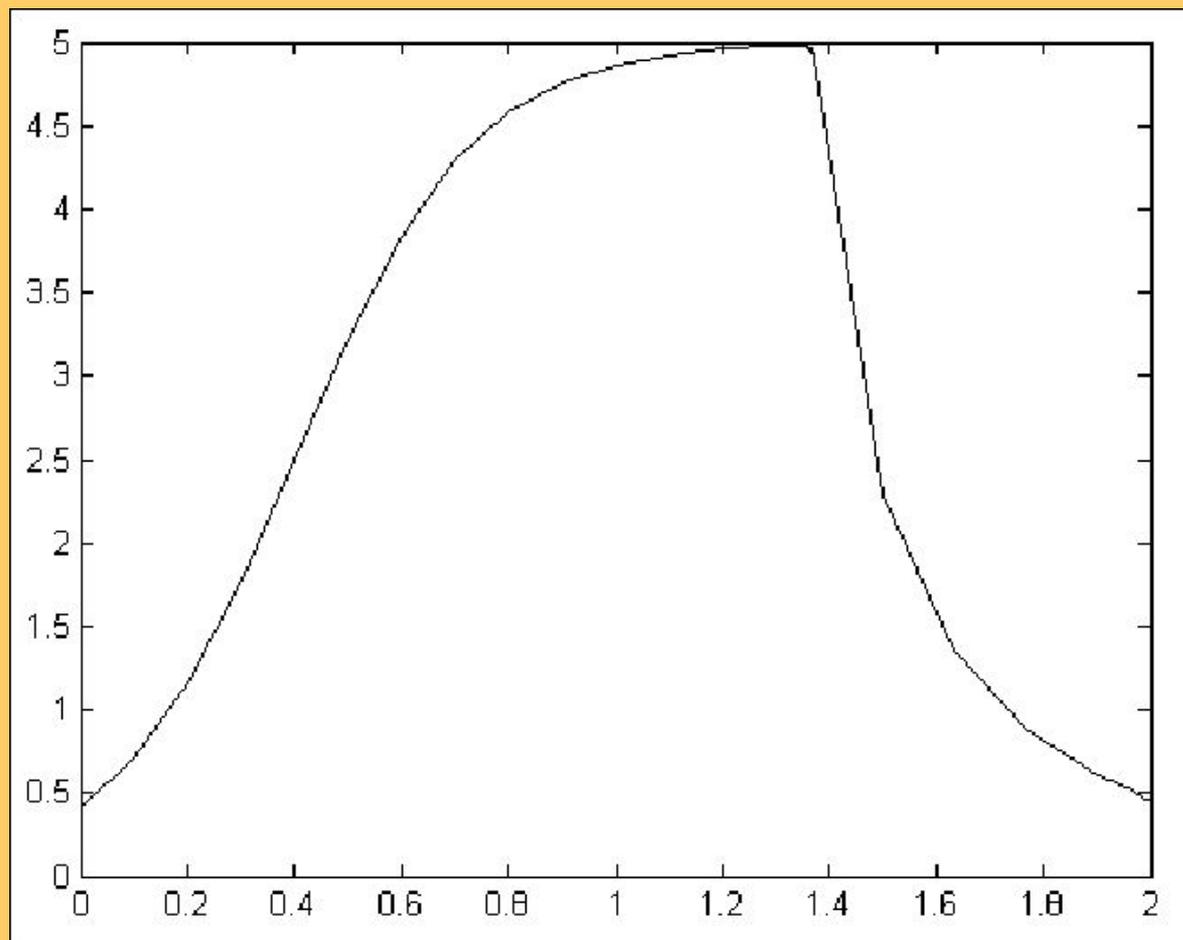
$$\frac{dv(t - \lambda)}{dt} = pv(t - \lambda)(1 - qv(t - \lambda)),$$
$$v(\lambda) = u(\lambda)$$

Решение второго уравнения:

$$v(t) = \frac{u(\lambda)}{qu(\lambda) + (1 - qu(\lambda)) \exp[-p(t - \lambda)]}.$$



Логистическая модель: обобщенный график информационного потока





Спасибо за внимание!

Ландэ Д.В

dwl@visti.net

<http://poiskbook.kiev.ua>

**МЕЖДУНАРОДНЫЙ СОЛОМОНОВ
УНИВЕРСИТЕТ
Киев, Украина**