

Модель диффузии информации

д.т.н. Д.В. Ландэ,

НТУУ «КПИ»

В настоящее время, когда объемы и динамика публикации информации в Интернет позволяют говорить об информационных потоках [1], актуальным становится изучение их динамики. Разнообразие поведения публикаций по различным тематикам и сложность взаимного влияния различных публикаций заставляют искать новые, ранее неизвестные в этой области методы. По-видимому, на данном этапе в области моделирования сложных информационных процессов успех может быть достигнут лишь путем синтеза достаточно простых алгоритмов и концепций.

Можно признать перспективными в этой области и теорию клеточных автоматов, впервые предложенную более тридцати лет тому назад Дж. фон Нейманом [2] и развитую С. Уолфрамом в фундаментальной монографии [3].

Клеточный автомат представляет собой дискретную динамическую систему, совокупность одинаковых клеток, образом соединенных между собой. Все клетки образуют сеть клеточных автоматов. Состояние каждой клетки определяется состоянием клеток, входящих в ее локальную окрестность. Окрестностью конечного автомата с номером j называется множество его «ближайших соседей». Состояние j -го клеточного автомата в момент времени $t + 1$, таким образом, определяется следующим образом:

$$y_j(t+1) = F(y_j(t), O(j), t), \quad (1)$$

где F – некоторое правило, которое можно выразить, например, языком булевой алгебры, $O(j)$ – окрестность, t – такт. Клеточные автоматы в традиционном понимании удовлетворяют таким правилам [4]:

- изменение значений всех клеток происходит одновременно (единица измерения времени – такт);
- сеть клеточных автоматов является однородной, т.е. правила изменения состояний для всех клеток одинаковы;
- на клетку могут повлиять лишь клетки из ее локальной окрестности;
- множество состояний клетки конечно.

Заметим, что теория клеточных автоматов была использована при анализе диффузии инноваций, процесса во многом близкого к распространению новостей в Интернет [5]. Подобная модель функционирует по следующим правилам: каждый индивид, способный принять инновацию, соответствует одной клетке, которая может находиться в двух состояниях: 1 - новинка принята; 0 - новинка не принята. Предполагается, что автомат, восприняв инновацию один раз, запоминает ее навсегда (состояние 1 – не может быть изменено). Автомат принимает решение о принятии новинки, ориентируясь на мнение восьми ближайших соседей, т.е. если в окрестности данной клетки имеется m сторонников новинки и, p - вероятность принятия новинки (генерируется по ходу работы модели), то при

$$pm > R, \quad (2)$$

где R - фиксированное пороговое значение, то клетка принимает инновацию (принимает значение 1).

Вместе с тем, динамике распространения информации присущи некоторые дополнительные свойства, которые и были учтены автором. При тех же условиях дополнительно предполагается, что клетка может быть в одном из трех состояний: 1 – «свежая новость» (клетка окрашивается в черный цвет); 2 – новость, устаревшая, но сохраненная в виде сведений (серая клетка); 3 – клетка не имеет информации, переданной новостным сообщением (клетка белая, информация не дошла или уже забыта). Правила распространения новостей следующие:

- изначально все поле состоит из белых клеток за исключением одной – черной, которая первой «приняла» новость (рис. 1 а);
- белая клетка может перекрашиваться только в черный цвет или оставаться белой (она может получать новость или оставаться «в неведении»);
- белая клетка перекрашивается, если выполняется условие, аналогичное (2), в модели диффузии инноваций: $pm > 1$;

- если клетка черная, а вокруг нее исключительно черные и серые, то она перекрашивается в серый цвет (новость устаревает, но сохраняется как сведения);
- если клетка серая, а вокруг нее исключительно серые и черные, то она перекрашивается в белый цвет (происходит забывание сведений при их общеизвестности).

Описанная система клеточных автоматов вполне реалистично отражает процесс распространения новостей среди отдельных информационных источников и их публикации. На поле размером 40 x 40 (размеры были выбраны авторами исключительно с целью наглядности) состояния системы клеточных автоматов полностью стабилизируются за ограниченное количество ходов, т.е. процесс эволюции - сходящийся. Пример работы модели приведен на рис. 1.

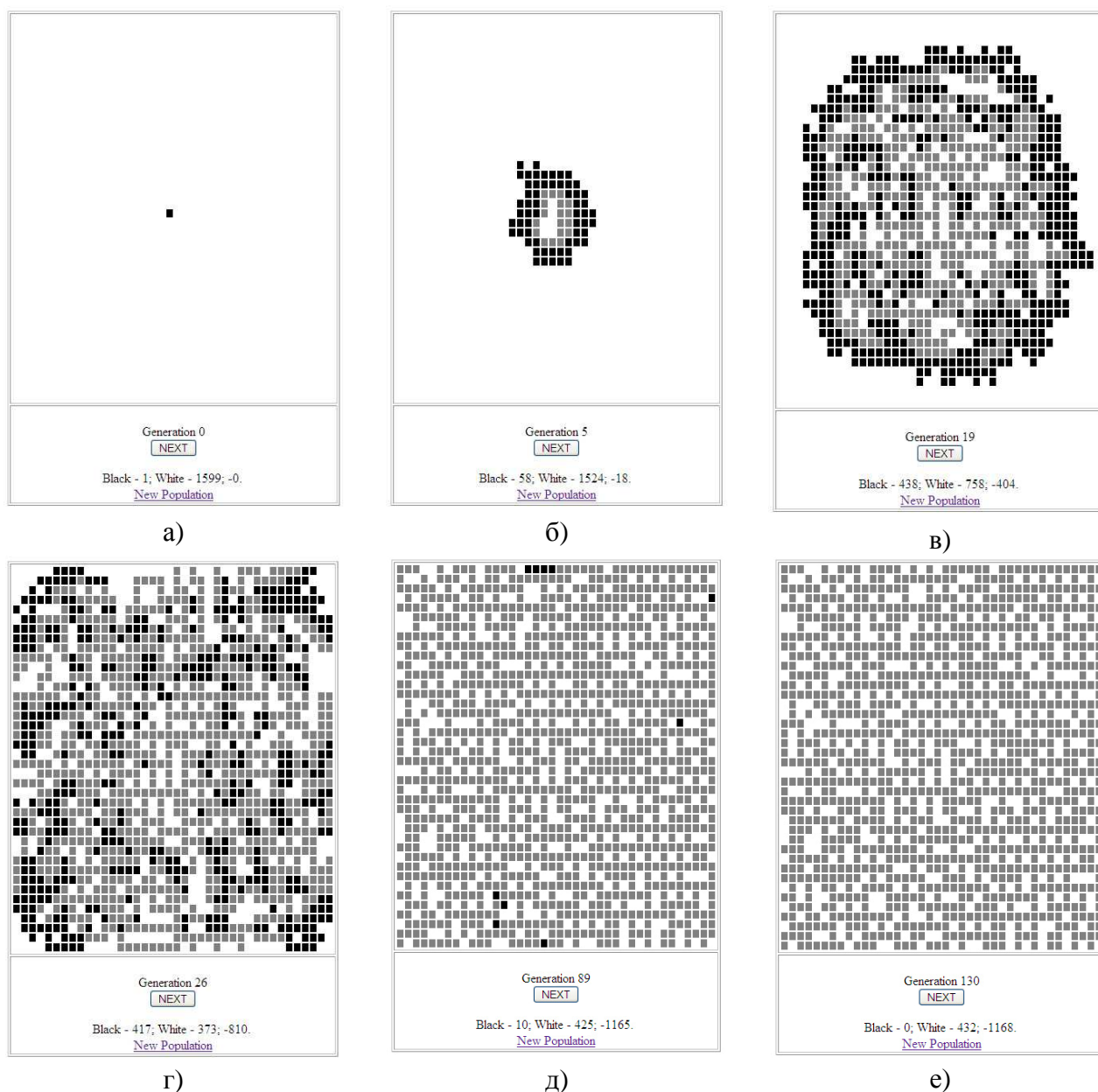


Рис.1. Процесс эволюции системы клеточных автоматов «диффузии новостей»: а – исходное состояние; б-д – промежуточные состояния; е – конечное состояние

Многочисленные эксперименты с данным клеточным автоматом, доступным в настоящее время в сети Интернет по адресу <http://edu.infostream.ua/newsk.pl> показывают, что период его сходимости составляет от 80 до 150 тактов. Типичные зависимости количества клеток (последовательности количества однотипных клеток), пребывающих в различных состояниях в зависимости от номера такта приведены на рис. 2. При анализе приведенных

графиков следует обратить внимание на такие особенности: 1 – суммарное количество клеток, пребывающих во всех трех состояниях на каждом шагу итерации постоянно и равно размеру поля, 2 – при стабилизации клеточных автоматов соотношения серых, белых и черных клеток приблизительно составляет: 0.75 : 0.25 : 0; существует точка пересечения кривых, определяемых всеми тремя последовательностями на уровне 33 % каждая.

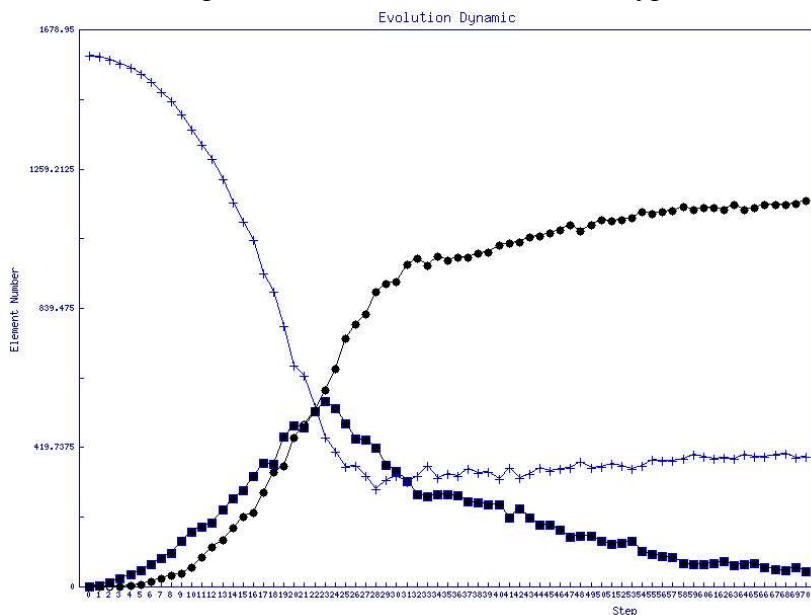


Рис. 2. Количество клеток каждого цвета в зависимости от шага эволюции: белые клетки - (+); серые клетки - (•); черные клетки - (■)

Особое внимание на графике следует обратить на зависимость, образованную черными клетками. Вид данной кривой вполне согласуется с «жизненной» динамикой новости, сначала она динамично распространяется, захватывая все новые уголки информационного пространства, затем происходит своеобразное насыщение, и информация для большинства реципиентов перестает быть новостью, переходя в разряд сведений или попросту забываясь.

Детальный анализ полученных зависимостей позволил провести аналогии данной модели «диффузии информации» со следующими аналитическими рассуждениями. Результаты моделирования позволяют предположить, что эволюция серых клеток описывается некоторой непрерывной функцией:

$$x_g = f(t, \tau_g, \gamma_g), \quad (3)$$

где t - время (шаг эволюции), τ_g - сдвиг по времени, обеспечивающие получение требуемого фрагмента аналитической функции, γ_g - параметр крутизны данной функции.

Соответственно, динамика белых клеток x_w (количество клеток в момент t) может моделироваться «перевернутой» функцией x_g со своими аналогичными параметрами:

$$x_w = 1 - f(t, \tau_w, \gamma_w). \quad (4)$$

Поскольку, как было сказано выше, всегда выполняется условие баланса, т.е. общее число клеток в конкретный момент времени всегда постоянно, то условие нормировки можно записать:

$$x_g + x_w + x_b = 1, \quad (5)$$

где x_w - количество черных клеток в момент времени t .

Таким образом, в соответствии с (5):

$$x_b = 1 - x_g - x_w = f(t, \tau_w, \gamma_w) - f(t, \tau_g, \gamma_g). \quad (6)$$

Вид зависимости, представленной на рис. 2. позволяет предположить, что в качестве функции $f(t, \tau, \gamma)$ может быть выбрано следующее выражение:

$$f(t, \tau, \gamma) = \frac{C}{1 + e^{\gamma(t-\tau)}}, \quad (7)$$

где C - некоторая нормирующая константа.

На рис. 3 приведены графики зависимостей x_g, x_w, x_b от шага эволюции системы клеточных автоматов, полученные в результате аналитического моделирования, выраженного формулами:

$$x_g = \frac{0.75}{1 + e^{-0.15(t-30)}},$$

$$x_w = 1 - \frac{0.75}{1 + e^{-0.25(t-20)}},$$

$$x_b = 0.75 \left(\frac{1}{1 + e^{-0.25(t-20)}} - \frac{1}{1 + e^{-0.15(t-30)}} \right).$$
(8)

Для экспериментальной проверки рассматриваемой модели были проведены исследования динамики объемов публикаций по двум резонансным тематикам, имеющим прямое отношение к избирательным процессам в Украине (досрочным выборам в Верховную Раду в 2007 году). Исследования проводились на базе системы интеграции информационных ресурсов сети Интернет InfoStream [6], охватывающей в настоящее время свыше 3000 источников новостной информации, обеспечивающей информационный поток в 50000 сообщений в сутки.

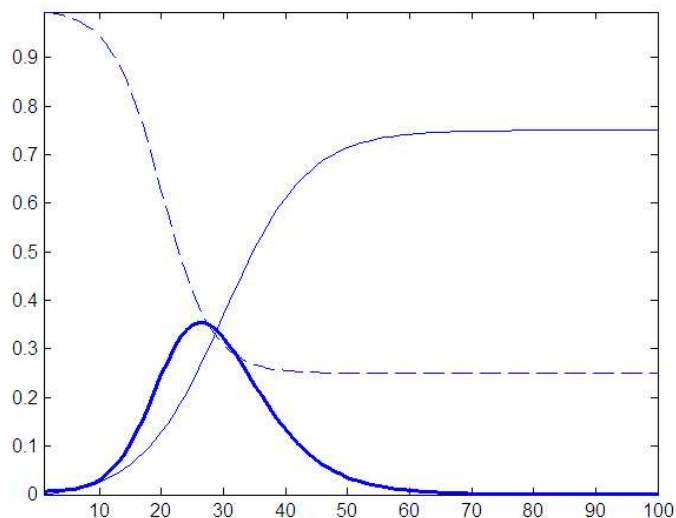


Рис.3. Непрерывные зависимости, полученные в результате аналитического моделирования, в зависимости от шага эволюции: сплошная линия – серые (x_g); пунктирная линия – белые (x_w); сплошная жирная линия – черные (x_b)

Для получения необходимых для эксперимента информационных потоков было сформировано два запроса к системе InfoStream [6]. Первый запрос касался регистрации избирательного блока БЮТ, решение о которой, как известно, было принято в результате судебного рассмотрения. Второй запрос позволил отобрать документы с информацией о четвертом указе Президента Украины, относящемся к досрочным выборам в Верховную Раду. На рис. 4 и 5 приведена реальная динамика объемов публикаций в Интернет по дням по соответствующим тематикам. Сравнение экспериментальных зависимостей с результатами аналитического моделирования позволяют говорить о высокой точности аппроксимации экспериментальных данных модельными кривыми.

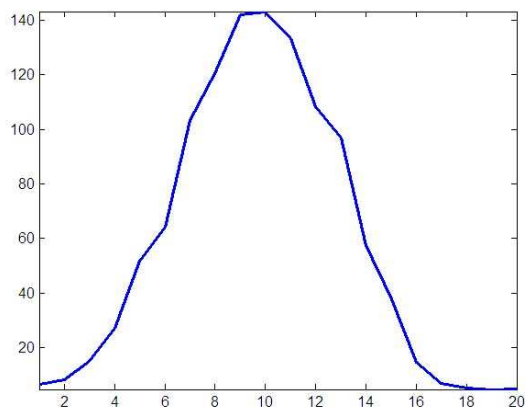


Рис. 4. Динамика понятия «ЦВК, БЮТ, регистрация, но не референдум», в системе интеграции новостей InfoStream

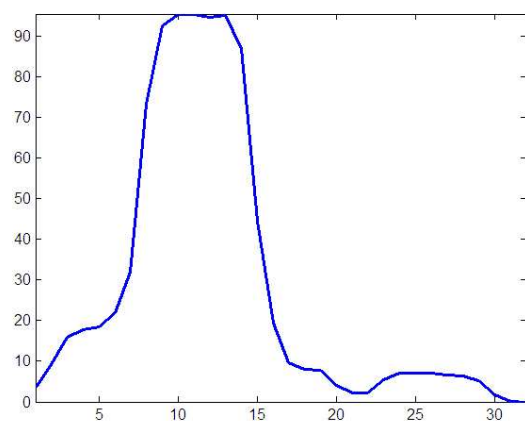


Рис.5. Динамика понятия «четвертый указ», в системе интеграции новостей InfoStream

В заключении следует отметить, что полученная колоколообразная зависимость диффузии новостей на интернет-источниках (веб-сайтах) хорошо согласуется с реальным поведением тематических информационных потоков, а на локальных временных промежутках с известными моделями, например, экспоненциальной и логистической.

Литература

- [1] Ландэ Д.В., Фурашев В.Н., Брайчевский С.М., Григорьев А.Н. Основы моделирования и оценки электронных информационных потоков. - К.: Инжиниринг, 2006. - 176 с.
- [2] Нейман Дж. Теория самовоспроизводящихся автоматов. - М.: Мир, 1971. – 382 с.
- [3] S. Wolfram. A New Kind of Science. Champaign, IL: Wolfram Media Inc., 2002. – 1197 pp.
- [4] Плотинский Ю.М. Модели социальных процессов. – Изд. 2-е. – М.: Логос, 2001. – 296 с.
- [5] S.C. Bhargava, A. Kumar, A. Mukherjee. A stochastic cellular automata model of innovation diffusion // Technological forecasting and social change. – 1993. Vol. 44. № 1. - P. 87-97.
- [6] Григорьев А.Н., Ландэ Д.В., и др. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис. Научно-методическое пособие. - Киев: ООО "Старт-98", 2007. - 40 с.