A memory model for internet hits after media exposure

Antonio G. Chessa\textsuperscript{a,}\ast, Jaap M.J. Murre\textsuperscript{a,}\textsuperscript{b}

\textsuperscript{a}Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands
\textsuperscript{b}University of Maastricht, The Netherlands

Received 16 May 2003

Abstract

We present a cognitive model, based on the mathematical theory of point processes, which extends the results of two studies by Johansen (Physica A 276 (2000) 338; Physica A 296 (2001) 539) on download relaxation dynamics. Responses from subjects are considered as single events, which are received from original listeners or readers and from a network of social contacts, through which a message may propagate further. We collected data on the number of daily visits at our web site after a radio interview with the second author, in which the name of the web site was mentioned. A model based on an exponential hit time distribution and a homogeneous point process for regular visitors fits our data and Johansen’s very well and is superior to both the power law and the logarithmic function. The fits suggest that hit data from different sources share the same cognitive mechanism, which are controlled merely by the encoding and retrieval of the target information memorised.

\textsuperscript{\ast}Corresponding author. Tel.: +31-20-525-6807; fax: +31-20-639-1656.
\textit{E-mail address:} achessa@fmg.uva.nl (A.G. Chessa).

\textcopyright 2003 Elsevier B.V. All rights reserved.

\textit{PACS:} 89.90.+n; 87.22.Jb

\textit{Keywords:} Message encoding; Memory; Information flow; Reaction time; Point processes; Internet

1. Introduction

The world-wide-web (WWW) provides one of the most efficient methods for retrieving information. However, statistics about the dynamics of information flow through the web, and its interaction with other types of media, still form a largely unexplored
field of research. Two recent papers published data about the response of a population on the WWW to the release of new information and, in particular, about the evolution of this response over time since the date of release [1,2]. In these two studies, the authors quantified the number of paper downloads from a website since the time of appearance of two interviews with the authors, in which the website was mentioned. This resulted in two data sets of the number of downloads as a function of time, the course of which was reported during periods of 70 and 100 days since the interviews.

In their analysis, the authors argued that the data were fitted reasonably well by a power law [1] and a logarithmic function [2]. Both these functions contain three parameters, which describe (1) the size of the Internet population that was interested in an interview, (2) the decline rate of paper downloads and (3) a ‘background rate’ that depends on the ease of finding the web page and the general interest of the subjects posted on the page. One of the future challenges posed by the authors is the “rationalisation” of the functions fitted. The process underlying such functions should include factors like the dynamics of information, rumour spreading and psychological decision. The latter is in fact the final stage of a memory process that starts with the initial encoding of information in the brain, then proceeds with the storage (that determines the forgetting rate) and finally with retrieval and recall. Final recall will depend on the encoding and storage characteristics that are at the disposal of ‘web surfers’ for retrieving and recalling the essential information, for example, its web address (URL), and on the effectiveness of the cue. The authors of the two papers do not make concrete attempts to formalise these processes, but hint at several possible stochastic models.

Our Memory Chain Model offers a comprehensive though concise formalisation of each of the above-mentioned four stages of memory [3,4]. The model is based on the theory of point processes [5,6], in which the ‘points’ of the process reflect ‘copies’, ‘representations’ or ‘features’ of a memorised item. Memory representations are generated after the presentation of an item, which may be transferred to one or more ‘memory stores’. These transfer or induction processes between stores characterise the underlying neurobiological processes that are responsible for the formation of a wide spectrum of short-term to long-term memories. The corresponding processes are active on time scales that range from milliseconds to tens of years. Beside the neurobiological plausibility, the Memory Chain Model offers a unified theory for different quantitative measures of learning and forgetting that have been introduced in the psychological literature over the past 100 years [7].

The objective of this paper is to apply the Memory Chain Model to WWW download statistics. We will apply our model both to the previously published data and to a new data set regarding the number of daily visits at our web site measured from the time of a radio interview in which the web site was mentioned. This data set will be described in the fourth section. In Section 2 we will give a short description of the Memory Chain Model. In Section 3 we will derive expressions for the number of Internet hits from this model that are based on responses from subjects who hear or read an interview and subsequently memorise the target information in the interview. In Section 4 we will fit the expected number of hits derived from the Memory Chain Model to
our data set. We will also fit this model to the previously cited data, which we will compare to the fits obtained with the power law and the logarithmic function applied in the two studies [1,2]. In Section 5 we will extend the Memory Chain Model with a component that describes information flow through a network of social contacts that may receive a message from the subjects that first heard or read an interview (the cognitive component). We will also discuss some more complex examples of Internet hits to which the model applies.

2. The Memory Chain Model

The mathematical basis of the Memory Chain Model is the theory of point processes [5,6]. Point processes are defined in similar terms as random variables (a measurable mapping from a probability space), but the ‘outcome space’, which is the set of real numbers in the case of random variables, has a more complex structure through the involvement of a phase or state space. Such a space may refer to spatio-temporal domains in practice, which can be represented by (a subset of) Euclidean space. The outcome space of a point process is a space of counting measures, which are integer-valued set functions. That is, for subsets of the state space that belong to its \(\sigma\)-algebra, a counting measure assigns integer values to such subsets. Within the context of point processes, these values are samples of integer-valued random variables. The probability distribution of a point process is completely determined by the finite-dimensional or joint probabilities of these random variables. A counting measure thus counts the number of ‘points’ or ‘events’ in subsets of the state space, which are sampled according to the finite-dimensional distributions of a point process.

Point processes are used to describe a broad range of physical, biological and environmental phenomena. A phenomenon, or some of its characteristics, can often be represented as a number of events, which are mapped as points in a two- or three-dimensional space and/or time. A rich class of point process models is formed by the Gibbs point processes, which in statistical physics are used to describe the potential energy of a collection of particles [6]. A point process can, for example, also be used to model the distribution of trees in a forest. A point then represents the location of a tree, which is ‘marked’ by the diameter of a tree. Another example is the location of bird nests of one or more different species.

The structure of grains and pores in a sedimentary body, and the location, size, and orientation of oil and gas reservoirs in the subsurface are other examples of spatial applications of point processes [8]. In the latter case, the time-dimension could be included in the state space to describe the actual geological processes that contributed to the formation of sedimentary bodies. One could then use a point process as a model for the formation of deposits in a fluvial system, where the sedimentation starts from a source area that develops into a system of streams that together create an alluvial fan. Every stream may branch into new streams that lay down new deposits, which, however, may erode older ones. Processes of sedimentation and erosion have similarities with the cascading structure of memory processes, which further motivates the use of point processes to describe such processes.
A recurring concept in memory psychology is the distinction between different types of memories, such as iconic or echoic memory, working memory, and long-term memory. When some kind of information is presented to a subject it will be encoded if it passes through the sensory systems (e.g., by paying attention to the information). At a later stage, neurons will fire in a part of the brain that holds a working memory representation of the information for a few seconds to minutes [9,10]. Memory representations held by such neural groups rapidly decay, but these can be ‘salvaged’ by a series of induction processes that lead to stronger representations in neural systems such as the hippocampus and the neocortex (Fig. 1). From a neurobiological point of view, these processes can be viewed as a cascade of induction and decline processes that take place at different time scales [11].

Point processes have already been used to describe time-series of firing neurons [12]. The variable of interest in the Memory Chain Model is the number of memory representations (‘critical features’ of some learned item) in time-intervals, so that the real line is the state space of the model. The point processes of the Memory Chain Model have an onset that coincides with the time of exposure to an item that will be memorised, for instance, the time at which learning is initiated or, in the present setting, the broadcast time of an interview that contains the target item. A point process therefore has no points, with probability 1, at times before the onset. At the end of item exposure, the intensity of the point process has some value $\mu_1$, which is equal to the mean number of memory representations per time-unit. This parameter is called the initial encoding of a memory. Forgetting usually takes place after item exposure, which in our model is described as a point process with intensity function $r(t) = \mu_1 \tilde{r}(t)$, where $t$ denotes the time since item exposure and $\tilde{r}$ is a function that may both describe a weakening ($\tilde{r}(t) < 1$) and a strengthening ($\tilde{r}(t) > 1$) of the initial encoding.

Fig. 1. (a) Storage systems for memories at different time scales, with feedforward induction between and decline within stores. (b) Abstract representation used in the Memory Chain Model.
The decline function $\tilde{r}$ describes the storage of a memory after its initial encoding. The shape of this function is determined by the number and linking of memory structures, which we call stores. A neurobiologically motivated assumption of our model is that these memory stores are linked in a feedforward manner (Fig. 1).

The initial encoding takes place in the first store of the chain where memory representations occur in time according to a point process with intensity function $r(t) = \mu_1 \exp(-a_1 t)$. We assume exponential decline functions for every store, where $a_i$ denotes the decline rate parameter for memory representations in store $i$. Store $i$ may induce a point process in the next store $i + 1$ of the chain. The intensity function of store $i + 1$ then arises as a convolution of the intensity function of store $i$ and the decline function of store $i + 1$. A further characteristic of the stores is that these are ordered such that $a_i > a_{i+1}$ for all stores $i$. This means that memory representations formed in higher-order stores have increasing expected lifetimes, which agrees with the concept of consolidation of short-term memories into longer-term representations.

When retrieval is attempted (e.g., at a test for recall), a cue will be used to access the target item stored. A subspace of the brain will be searched for memory representations, the size of which denotes the effectiveness of a retrieval cue. If cue effectiveness, which we denote by $q$, does not depend on time $t$, then the intensity function at retrieval is a thinned version of $r$, namely, $\mu_1 \tilde{r}(t)q$, where $0 \leq q \leq 1$. The initial encoding and cue effectiveness act as a single parameter when these are not affected separately. We will continue to write $\mu_1$ for $\mu_1 q$.

The Memory Chain Model has been applied to a broad range of memory and retention data, including learning and forgetting curves for normal subjects, data of different types of amnesia, and data on the proven recall of advertisements [3,4,13,14]. The model has been used to derive and fit different retention functions to these data, such as the probability of recall, measures for recognition memory, and probability distributions for ages of autobiographical memories. Retention is determined by the presence of memory representations aggregated over all the stores, so that we merely have to consider the intensity function of the sum of the point processes over all the stores.

In the next section we will use the intensity function of a single-store model to derive long-term reaction time distributions, which will be fitted to the Internet hit time data presented in the fourth section.

3. Hit time distributions

The derivation of reaction time distributions can be considered as a classical application of point process theory. Such distributions can be obtained by calculating the first contact distribution function, which is the distribution of the distance from an arbitrarily fixed point to the nearest point of the process [6]. In terms of the Memory Chain Model, this can be interpreted as the distribution of the time between item exposure, which we will fix at time zero, and the occurrence of the first memory representation that allows retrieval of a memory, thereby permitting a subject to take action. Let us denote the reaction or hit time for the Memory Chain Model by $T$, its distribution function by $F$ and the underlying point process by $M$. The distribution function $F$
follows from the relation \( F(t) = P\{ T \leq t \} = P\{ M([0,t]) \geq 1 \mid M([0,\infty)) \geq 1 \} \), where the random variable \( M([0,t]) \), which denotes the number of memory representations in the time interval \([0,t]\), has expectation \( E[M([0,t])] = \int_0^t r(z) \, dz \). Notice that the conditioning on the event \{ \( M([0,\infty)) \geq 1 \) \} is necessary in order to obtain \( F \), since this event does not have probability 1 for the point process \( M \). We assume that \( M \) is a Poisson point process, so that the random variable \( M([0,t]) \) has a Poisson distribution. Another important property of Poisson processes is the independent scattering property, which means that numbers of points in disjoint subsets of the state space are independent. By working out the conditional probability that determines \( F \), and by making use of the aforementioned properties of the Poisson process, it follows that

\[
F(t) = \frac{1 - \exp(-\int_0^t r(x) \, dx)}{1 - \exp(-\int_0^\infty r(x) \, dx)}.
\]

For a single-store point process with intensity function \( \mu_1 \exp(-a_1 t) \) we obtain, for decline rate parameters \( a_1 > 0 \),

\[
F(t) = \frac{1 - \exp(-\mu_1/a_1)(1-e^{-a_1 t})}{1 - \exp(-\mu_1/a_1)}.
\]

If \( a_1 = 0 \), the single-store point process is a homogeneous Poisson process with intensity \( \mu_1 \), which gives rise to the exponential distribution function

\[
F(t) = 1 - \exp(-\mu_1 t).
\]

The form of the distribution function \( F \) depends on two aspects: (1) the number of memory stores, and (2) the number of memory representations that is required to retrieve and recall the memory of an item. In order to derive \( F \) we made use of first contact distributions, which implies that we implicitly used the psychological assumption that one memory representation will suffice to recall the full item. We used this recall threshold as a reference value in each of our model fits to experimental data \[3,4\]. Departures from this base value were needed in a limited number of studies, where recall thresholds greater than 1 seem to be related to the complexity of the structure of items to be memorised. During initial learning, new items are then memorised as distinct ‘chunks’ of information, which are gradually assembled into a single chunk after repeated learning of the same item (leading to a series of recall thresholds that decrease to 1). The item that had to be memorised in the present study is the name of our web site, which has a simple structure, so that we set the recall threshold equal to 1.

### 4. Fits to Internet hit time data

A 3-min radio interview with the second author (JM) that was broadcast nation-wide mentioned our URL during primetime (4:50 p.m.). This allowed us to record in the subsequent days a large, isolated peak in hits to our web site (Fig. 2). Our site is located at the easy-to-remember address (for the Dutch) ‘memory.uva.nl’ and contains a Daily News Memory Test that has been completed spontaneously nearly 20,000 times in the past two years. Participants are sent an electronic mail message upon completion
in which they are invited to bring our site to the attention of their acquaintances. The observed number of daily hits consists of two components: the hits from listeners and their social contacts, and hits coming from visitors to our web site who did not hear about the interview. The latter group of subjects give rise to a ‘base rate process’, which appears in Fig. 2 as the number of daily hits plotted before the interview. The number of hits after the interview tends to stabilise around this base rate with increasing time lags. We recorded the number of daily hits for 21 days until the day the interview was broadcast and for 28 days after the interview.

Expressions for the expected number of daily hits can be derived as follows from the hit time distributions (1) and (2). The number of daily hits in the base rate process can be thought of as a homogeneous point process with (constant) intensity $\beta$ (per day), which is equal to the expected number of daily base rate hits. We also introduce a parameter $n$ for the total number of hits from original listeners after the interview. The expected cumulative number of daily hits in the time interval $[0, t]$, where time 0 fixes the time of broadcast, is equal to

$$\beta t + nF(t).$$

This expression contains four parameters for hit time distribution (1) and three for distribution function (2): the base rate $\beta$, the initial encoding $\mu_1$, the decline rate parameter $a_1$ of the single-store Memory Chain Model (which appears only in (1)), and the parameter $n$.

We estimated the base rate $\beta$ from the number of daily hits before the interview, which is obtained by dividing the total number of hits over this period by 21 days, which gives an estimate of 2622.81. The total number of hits reported during the first 28 days after the interview is 155,216, so that the parameter $n$ must satisfy the equation $nF(28) = 155,216 - 28\beta$. This implies that we are left with one or two free parameters to fit the shape of the curve representing the number of hits after the radio interview. Fig. 2 shows an excellent fit of expression (3) to these data, with exponential cognitive hit time distribution function $F$ given by (2), which corresponds to a single-store point
process with zero decline rate and with initial encoding $\mu_1 = 0.414$. The fit does not improve for a hit time distribution function with nonzero decline rate.

We also used expression (3) to fit the download rate data of Johansen [2]. The fitted expression, which is shown in Fig. 3, has the same analytical form as the function fitted in Fig. 2. The fit indicates a long-term, asymptotic daily base rate of $\beta = 18.91$ downloads, an initial encoding $\mu_1 = 0.181$, zero decline rate $a_1$, and $n = 1589$. As was the case with our data, a model with nonzero decline rate does not improve the fit. An interesting finding that emerges from the above results is that, apart from the difference in the base rates between the fits in Figs. 2 and 3, the only parameter that controls the shape of the two fitted functions is the initial encoding. The value of $\mu_1$ is higher in the fit of Fig. 2, which may suggest either that our URL (www.memory.uva.nl) was better encoded than the URL of Johansen’s [2] study (www.wallstreetuncut.com), for instance, because of a longer exposure to the URL, or that cues used to retrieve encoded memory representations were more effective for our URL. We did not fit our model to the data of the first study by Johansen [1], because this data set has considerable variability.

The results obtained with our model for the two fitted data sets invite us to make a comparison with Johansen’s proposed power law and logarithmic function [1,2]. In particular, we are able to use the Internet hit data to investigate whether the cognitive processes underlying the hit rates are better matched by the assumptions of the Memory Chain Model, a power law or logarithmic function. The expected cumulative hit rates for these three models are given by $\beta t + n(1 - \exp(-\mu_1 t))$, $\beta t + nt^{1-b}$ and $c + \beta t + n \ln(t + 1)$, respectively. Each of these functions has three parameters, so that a fair comparison can be made. The power law fitted to our hit data is plotted as a dashed curve in Fig. 2. (The logarithmic function gave a similar fit, and is not shown.) The fit of the function that follows from our single-store Memory Chain Model is considerably better than the power law fit; the sum of squared differences for our model is about six times lower. The same result holds for the fits to Johansen’s [2] data. This confirms
one of the important findings in our previous studies, where the Memory Chain Model was found to be superior to frequently used functions, such as the power law and logarithmic functions, in practically all fits to memory retention data [3,4]. In Figs. 2 and 3 it can be seen that the power law fails to give a satisfactory fit to the data for the smaller time lags; it declines too rapidly in comparison with the behaviour shown by the data. This characteristic was also observed in many fitting studies on memory and retention in our previously cited papers.

5. Hit times for a socio-cognitive model

In a more general setting, hits may be obtained both from subjects who were directly exposed to an item and from their acquaintances, who constitute a network of social contacts. In order to model hit times in such situations, we assume that publicised information (i.e., our URL) flows through a network that encompasses first the neurobiological processes of memory in the original listeners. If the information and intention to pass it on are still available, this may also give rise to the activation of a network of social contacts.

In this section we propose an extended model for the expected number of hits described by (3), with the additional contribution of hits from persons belonging to a social network of contacts, who may receive messages from the original recipients. We consider hits from social contacts as a nonhomogeneous point process with intensity function \( \alpha(t - \tau)^m e^{-\lambda(t-\tau)} \), where \( t \geq \tau \) and \( \alpha, m, \lambda \geq 0 \), which is induced by a hit from an original recipient at time \( \tau \). The parameter \( \alpha \) controls the number of messages sent by an original recipient to his acquaintances and the willingness of social contacts to visit a website, \( \lambda \) controls the rate at which hits from social contacts decline in time, while \( m \) is a measure for the degree of connectivity of a social network. The intensity function of the point process describes different kinds of hit time behaviour. The number of daily hits may strictly increase, strictly decrease, remain constant at value \( \alpha \), or reach a maximum at time \( m/\lambda \), for \( \lambda > 0 \).

The expected cumulative number of ‘social hits’ \( S(t) \) in a time-interval \([0, t]\) induced by one original recipient follows from the expression

\[
S(t) = \int_0^t \int_\tau^t \alpha(z - \tau)^m e^{-\lambda(z-\tau)} f(\tau) \, dz \, d\tau ,
\]

where \( f \) is the probability density function of the cognitive hit time distribution function \( F \) given by (1) or (2). Expression (4) can be calculated analytically for integer values of \( m \) and exponential density \( f \). A full analytical account of this function requires a separate treatment of different special cases of \( S \), of which we will only discuss the function that resulted as the best fit to our data. We fitted the expected cumulative number of hits that satisfies the formal expression

\[
\beta t + n(F(t) + S(t)) .
\]

The model with an extended social component gives a better fit to our data than the model that only contains the base rate and the reaction time distribution derived from the Memory Chain Model. Function (5) gives the best fit with parameter \( m \) equal to
Fig. 4. Fit of expression (5) (solid line), with parameters $\mu_1 = 0.479$, $a_1 = 0$, $n = 27000$, $\alpha = 7.97$, $m = 0$, $\lambda = 3.93$, $\beta = 2622.81$, to the daily Internet hits of Fig. 2 (dots).

zero, that is, for a point process of hits from social contacts that has intensity function $\alpha e^{-\lambda(t-\tau)}$, for a given hit time $\tau$ from an original recipient. Fig. 4 shows the fit of expression (5) with this intensity function to our data, which is given by

$$\beta t + n(1 - e^{-\mu_1 t}) + n \frac{\alpha}{\lambda}(1 - e^{-\lambda t}) - n \frac{\alpha}{\mu_1 - \lambda}(e^{-\mu_1 t} - e^{-\lambda t})$$

(6)

The parameter values of the fit ($\alpha = 7.97$, $\lambda = 3.93$) suggest a large transfer of messages from original recipients to acquaintances, who respond quickly but are less willing to transfer information on our URL among their acquaintances. The fitted function improves the sum of squared differences between model and data by 20% with respect to function (3) shown in Fig. 2. The largest improvement is obtained in the fits to the number of hits on the first two days after the radio interview. The decrease in the expected number of hits on the second day is delayed in comparison with the function in Fig. 2 because of the transfer of messages from the original recipients.

Function (5) did not improve the fit for Johansen’s data, which suggests that Johansen’s download rates are merely the result of responses from a cognitive process that applies to the original recipients.

6. Concluding remarks

Following the two studies by Johansen [1,2], we collected similar Internet hit data, but with much larger sample sizes. We fitted these data with our Memory Chain Model to hits induced by cognitive processes among original listeners, and with a point process for hits from a network of social contacts. The result was a more explicit description of the processes underlying Internet hits, which proved to give much better fits to our data set and also to the data set described in Johansen [2] than the power law and the logarithmic function applied in the two studies by Johansen, for the same number of parameters in Figs. 2 and 3. Our data show evidence for transfer of information from the original recipients of the information on our URL (Fig. 4), while the data
in Johansen [2] suggest that hits from a social network are negligible. A reason for this difference may be that our participants receive e-mail messages that invite them to notify friends.

It is also important to note that the two data sets could be fitted by the same expression for the expected number of hits per day, in which only the expected number $\mu_1$ of encoded memory representations among the original listeners (combined with cue effectiveness at retrieval), the base rate $\beta$ and the transfer rate $\alpha$ are varied over the two different data sets. The base rates are different because we are dealing with different web sites. The encoding-retrieval and transfer values are different because the two web sites have been visited by different persons. Encoding depends on learning speed, time, interest and other factors that may vary among different subjects. Alternatively, the initial encoding may be the same, but the retrieval processes may have been characterised by cues with a different effectiveness for the two data sets.

The construction of a valid model for some form of response after an exposure to a target item is very important in different situations. Johansen [2] already mentioned the implications for the modelling of social systems such as financial markets. Other situations, in which both a cognitive and a social component are represented, are, for instance, the effect of advertising commercials on consumer buying behaviour. In this example, a hit time would represent the time at which a consumer bought a certain product. Though a single exposure has considerable scientific merit, in most advertising campaigns a large number of media exposures are distributed in time. The response to a commercial then is the result of a superposition of single exposures. In previous work, we extended and successfully applied the Memory Chain Model to situations for describing the learning and forgetting of commercials during and after repeated exposures to the same commercial [13,14]. This model allows maximisation of the effectiveness of a campaign by optimising the distribution of media exposures in time. The same model would also be useful, for example, in situations with hits reported after different media publications that contain a URL, where the times of publication are chosen in such a way that the expected number of hits on a certain day or time interval is maximised.

**Acknowledgements**

This research was supported by NWO, the Netherlands Society for Scientific Research. We are very grateful to Anders Johansen for providing us his paper download data sets, to Eric Maryniak for organising the Internet statistics and to Eric Postma for pointing out this research field.

**References**


