Cognitive Temporal Document Priors (Abstract)*

Maria-Hendrike Peetz and Maarten de Rijke
ISLA, University of Amsterdam
{M.H.Peetz, derijke}@uva.nl

1. INTRODUCTION

Every moment of our life we retrieve information from our brain: we remember. We remember items to a certain degree: for a mentally healthy human being retrieving very recent memories is virtually effortless, while retrieving untraumatic memories from the past is more difficult [4]. Early research in psychology was interested in the rate at which people forget single items, such as numbers. Psychology researchers have also studied how people retrieve events. Chessa and Murre [1] record events and hits of web pages related to an event and fit models of how people remember, the so-called *retention function*. Modeling the retention of memory has a long history in psychology, resulting in a range of proposed retention functions. In information retrieval (IR), the relevance of a document depends on many factors. If we request recent documents, then how much we remember is bound to have an influence on the relevance of documents. Can we use the psychologists' models of the retention of memory as (temporal) document priors? Previous work in temporal IR has incorporated priors based on the exponential function into the ranking function [2, 3]—this happens to be one of the earliest functions used to model the retention of memory. Many other such functions have been considered by psychologists to model the retention of memory—what about the potential of other retention functions as temporal document priors?

Inspired by the cognitive psychology literature on human memory and on retention functions in particular, we consider seven temporal document priors. We propose a framework for assessing them, building on four key notions: *performance*, *parameter sensitivity*, *efficiency*, and *cognitive plausibility*, and then use this framework to assess those seven document priors. We show that on several data sets (newspaper and microblog), with different retrieval models, the exponential function as a document prior should not be the first choice. Overall, other functions, like the Weibull function, score better within our proposed framework.

2. METHODS

We introduce basic notation and then describe several retention functions serving as temporal document priors.

We say that document D in document collection $\mathcal{D}$ has time $time(D)$ and text $text(D)$. A query q has time $time(q)$ and text $text(q)$. We write $\delta_g(q,D)$ as the time difference between $time(q)$ and $time(D)$ with the granularity g.

We introduce a series of retention functions. The *memory chain models* ((1) and (2)) build on the assumptions that there are different memories. The Weibull functions ((3) and (4)) are of interest to psychologists because they fit human retention behavior well. In contrast, the retention functions *linear* and *hyperbolic* ((6) and (7)) have little cognitive background.

Memory Chain Model. The memory chain model [1] assumes a multi-store system of different levels of memory. The probability to store an item in one memory being μ,

$$f_{\text{MCM-1}}(D,q,g) = \mu e^{-a\delta_g(q,D)}. \tag{1}$$

The parameter a indicates how items are being forgotten. The function $f_{\text{MCM-1}}(D,q,g)$ is equivalent to the exponential decay in [2] when the two parameters (μ and a) are equal. In the two-store system, an item is first remembered in short term memory with a strong memory decay, and later copied to long term memory. Each memory has a different decay parameter, so the item decays in both memories, at different rates. The overall retention function is

$$f_{\text{MCM-2}}(D,q,g) = 1 - e^{-\mu_1\left(e^{-a_1\delta_g(q,D)} + \frac{\mu_2}{a_2-a_1}\left(e^{-a_2\delta_g(q,D)} - e^{-a_1\delta_g(q,D)}\right)\right)}, \tag{2}$$

where an overall exponential memory decay is assumed. The parameter μ_1 and μ_2 are the likelihood that the items are initially saved in short and long term memory, whereas a_1 and a_2 indicate the forgetting of the items. Again, t is the time bin.

One can also consider the Weibull function

$$f_{\text{BW}}(D,q,g) = \left(e^{-\frac{a\delta_g(D,q)}{d}^{d}}\right), \tag{3}$$

and its extension

$$f_{\text{EW}}(D,q,g) = b + (1-b)\mu e^{\left(-\frac{a\delta_g(D,q)}{d}\right)^{d}}. \tag{4}$$

Here, a and d indicate how long the item is being remembered: a indicates the overall volume of what can potentially be remembered, d determines the steepness of the forgetting function; μ determines the likelihood of initially storing an item, and b denotes an asymptote parameter.

The power function is ill-behaved between 0 and 1 and usual approximations start at 1. The *amended power function* is

$$f_{\text{AP}}(D,q,g) = b + (1-b)\mu(\delta_g(D,q)+1)^{a}, \tag{5}$$

where a, b, and μ are the decay, an asymptote, and the initial learning performance.

A very intuitive baseline is given by the linear function,

$$f_{\text{L}}(D,q,g) = \frac{-(a \cdot \delta_g(q,D) + b)}{b}, \tag{6}$$

where a is the gradient and b is $\delta_g(q, \text{argmax}_{D' \in \mathcal{D}}\, \delta_g(q,D'))$. Its range is between 0 and 1 for all documents in $\mathcal{D}$.

The hyperbolic discounting functionhas been used to model how humans value rewards: the later the reward the less they consider

*The full version of this paper appeared in ECIR 2013 [5].

Table 1: Assessing temporal document priors; # improved queries is w.r.t. MCM-1.

Condition	MCM-1	MCM-2	BW	EW	AP	L	HD
# impr. queries (temp.)	n/a	14 (58%)	5 (20%)	16 (67%)	5 (20%)	2 (8%)	6 (25%)
# impr. queries (non-temp.)	n/a	27 (35%)	35 (46%)	26 (34%)	38 (50%)	36 (47 %)	33 (43%)
# impr. queries (Tweets2011)	n/a	16 (32%)	17 (34%)	22 (44%)	0 (0%)	17 (34 %)	21 (42%)
MAP	+	–	+	0	0	–	0
P10	–	–	0	–	0	0	0
Rprec	0	±	+	±	0	0	0
MRR	0	0	+	0	+	+	+
Sensitivity of parameters	–	–	+	–	+	+	+
Efficiency: # parameters	2	4	2	4	3	2	1
Plausibility: fits human behav.	+	++	+	++	+	n/a	n/a
Plausibility: neurobiol. expl.	+	+	–	+	–	–	–

the reward worth. Here,

$$f_{\mathrm{HD}}(D,q,g) = \frac{1}{-(1+k*\delta_g(q,D))}, \tag{7}$$

where k is the discounting factor.

3. EXPERIMENTS

We propose a set of three criteria for assessing temporal document priors and we determine whether the priors meet the criteria.

A framework for assessing temporal document priors.

Performance. A document prior should improve the performance on a set of test queries for a collection of time-aware documents. A well-performing document prior improves on the standard evaluation measures across different collections and across different query sets. We use the *number of improved queries* as well as the *stability of effectiveness* with respect to different evaluation measures as an assessment for performance, where stability refers to that improved or non-decreasing performance over several test collections.

Sensitivity of parameters. A well-performing document prior is not overly sensitive with respect to parameter selection: the best parameter values for a prior are in a *region* of the parameter space and not a single value.

Efficiency. Query runtime efficiency is of little importance when it comes to distinguishing between document priors: if the parameters are known, all document priors boil down to simple look-ups. We use the *number of parameters* as a way of assessing the efficiency of a prior.

Cognitive plausibility. We define the cognitive plausibility of a document prior (derived from a retention function) with the goodness of fit in large scale human experiments [4]. This conveys an experimental, but objective, view on cognitive plausibility. We also use a more subjective definition of plausibility in terms of *neurobiological background* and how far the retention function has a biological explanation.

Discussion. To ensure comparability with previous work, we use different models for different datasets: TREC-2 and TREC-{6,7,8} for news and Tweets2011 for social media. On the news data set, we analyse the effect of different temporal priors on the performance of the baseline, query likelihood with Dirichlet smoothing [2]. We optimize parameters for different priors on TREC-6 using grid search. On the Tweets2011 data set, we analyse the effect of different temporal priors incorporated in the query modeling [3].

Table 1 gives an overview of the assessment of different document priors. We find that all but BW, AP, and L are stable in the parameter optimisation. Of those functions, BW and L have only few parameters, and BW performs best.

4. CONCLUSION

We have proposed a new perspective on functions used for temporal document priors used for retrieving recent documents. We showed how functions with a cognitive moti- vation yield similar, if not significantly better results than others on news and microblog datasets. In particular, the Weibull function is stable, easy to optimize, and motivated by psychological experiments.

Acknowledgments. We thank Jessika Reissland for her inspiration. This research was partially supported by the European Union's ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme, CIP ICT-PSP under grant agreement nr 250430, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements nr 258191 (PROMISE Network of Excellence) and 288024 (LiMoSINe project), the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.061.814, 612.061.815, 640.004.802, 727.011.005, 612.001.116, HOR-11-10, the Center for Creation, Content and Technology (CCCT), the BILAND project funded by the CLARIN-nl program, the Dutch national program COMMIT, by the ESF Research Network Program ELIAS, and the Elite Network Shifts project funded by the Royal Dutch Academy of Sciences.

References

[1] A. G. Chessa and J. M. Murre. A memory model for internet hits after media exposure. *Physica A Statistical Mechanics and its Applications*, 2004.

[2] X. Li and W. B. Croft. Time-Based Language Models. In *CIKM '03*, 2003.

[3] K. Massoudi, E. Tsagkias, M. de Rijke, and W. Weerkamp. Incorporating Query Expansion and Quality Indicators in Searching Microblog Posts. In *ECIR 2011*, 2011.

[4] M. Meeter, J. M. J. Murre, and S. M. J. Janssen. Remembering the news: modeling retention data from a study with 14,000 participants. *Memory & Cognition*, 33:793–810, 2005.

[5] M.-H. Peetz and M. de Rijke. Cognitive temporal document priors. In *34th European Conference on Information Retrieval (ECIR'13)*, 2013.