Adaptive Temporal Query Modeling

Maria-Hendrike Peetz, Edgar Meij, Maarten de Rijke, and Wouter Weerkamp

ISLA, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
{m.h.peetz,edgar.meij,derijke,w.weerkamp}@uva.nl

Abstract. We present an approach to query modeling that uses the temporal distribution of documents in an initially retrieved set of documents. Such distributions tend to exhibit bursts, especially in news-related document collections. We hypothesize that documents in those bursts are more likely to be relevant and update the query model with the most distinguishing terms in high-quality documents sampled from bursts. We evaluate the effectiveness of our models on a test collection of blog posts.

1 Introduction

In this paper, we consider chronologically ordered document collections, such as those related to news. Here, discussions around a certain topic have a transient character and relevant documents are likely to appear in bursts. Previous approaches to exploiting the transient and bursty nature of relevance in temporally ordered document collections assume either that recent documents are more relevant [3] or they compute a temporal similarity to retrieve recent or diverse documents [5].

We automatically identify one or more bursts in the result set and find the most discriminating terms in the top ranked documents of the burst. Those terms are then used as a basis for query modeling.

2 Related Work

A typical example of query modeling is based on (pseudo-)relevance feedback [6]. More recent examples include Meij and de Rijke [10], who perform semantic query modeling by linking queries to Wikipedia, or Balog et al. [1], who also incorporate information from an entity's category in the setting of entity retrieval.

News corpora are inherently temporal. Early work by Li and Croft [7] tries to make use of this feature under the assumption that recent documents are more likely to be read and deemed relevant, creating an exponential recency prior. Efron and Golovchinsky [3] expand upon this and incorporate an exponential decay into the query likelihood. Dakka et al. [2] propose a more general framework to incorporate time into a language model. Jones and Diaz [4] classify queries according to the temporal distribution of result documents. Focusing on the blogosphere, the number of approaches to blog (post) retrieval that make specific use of temporal aspects is limited. Weerkamp and de Rijke [12] use timeliness of a blog post as an indicator for determining credibility of blog posts. Under the assumption that more recent documents are more relevant, Massoudi et al. [9] use an exponential decay for query expansion on microblogs.

R. Baeza-Yates et al. (Eds.): ECIR 2012, LNCS 7224, pp. 455–458, 2012.

3 Adaptive Temporal Query Modeling

Our approach to query modeling is based on pseudo-relevance feedback, which aims to improve the language model of the query by first retrieving a set R of N top-ranked documents and identifying and weighting the most distinguishing terms in documents in R. The updated query model is then used to retrieve the final ranked list of documents. We proceed in a similar fashion, but take into account the temporal distribution of documents in R. This leads to the key assumption of this paper: *documents in bursts are more likely to be relevant*.

Identification of bursts. More formally, each document D has an associated timestamp *time*(D). A term w in the document D is sampled with probability $P(w|D)$. We denote the set of bursts for R as *bursts*$(R) \subseteq \mathcal{P}(R)$, the calculation of which is based on the temporal distribution of documents in R. Let $\{$*counts*$(t, R) : t \in T\}$ be the (discrete and binned) distribution of timestamps of documents in R. The set of timepoints T_i for a burst $B_i \in$ *bursts*(R) is defined as $\{$*time*$(D) : D \in B_i\}$. We then define key properties of a burst B_i as follows:

- There exists $t \in T_i$, where *counts*(t, R) is more than two standard deviations away from the mean of all time points in R. Such t are called *peaks*.
- For all timepoints $t \in T_i$, *counts*(t, R) is more than one standard deviation away from the mean of all time points in R.
- For each timepoint $t \in T_i$, either $t - 1$, t, or $t + 1$ is a peak.

Term reweighting. For a burst B, we sample terms according to:

$$P(w|B) = \frac{1}{N_B} \sum_{D \in B} P(w|D)P(D|B), \tag{1}$$

where $P(D|B)$ is the probability of the document given the burst (defined below). We only use documents with the highest scores from the bursts, so D has to be in the top-N_B documents in the burst. The expansion terms are the M highest scored terms of each burst that occur in at least 10 documents without being stopwords (based on preliminary experiments). Their final probability is calculated as:

$$P(w|q) = \frac{1}{|\textit{bursts}(R)|} \sum_{B \in \textit{bursts}(R)} P(w|B). \tag{2}$$

The weight is normalized and set to zero for all non-expansion terms. For retrieval, documents are ranked using the divergence between the query model and the document model with the Kullback-Leibler (KL) divergence [8].

Decay functions. Adapted to bursts, $P_{exp}(D|B)$ decreases exponentially with its distance to the peak of the burst. Formally,

$$P_{exp}(D|B) = e^{\gamma(|\max(B) - \textit{time}(D)|)}, \tag{3}$$

where γ is the decay parameter and $\max(B)$ is the peak containing the most documents. The decay parameter γ is a free parameter, which needs to be trained or estimated; we propose a burst-adaptive variant for γ. Let $\sigma(B)$ be the standard deviation of *counts*(t, B) for all $t \in T$. Then, $\gamma_B = 1/(2\sigma(B)^2)$ is the decay parameter for exponential decay. This decay function is the best fitting gaussian. We call this an *adaptive* exponential decay, and $P_{exp\text{-}adapt}(D|B) = e^{1/(2\sigma(B)^2)\cdot(|\max(B) - time(D)|)}$.

4 Experimental Evaluation

We compare the effectiveness of our two models (trained γ and adaptive exponential γ) with the baseline used in [3] and the temporal prior [7]. Our test collection is Blogs06 [11], a collection of blog posts collected during a three month period (12/2005–02/2006) from a set of 100,000 blogs. After standard preprocessing we are left with just over 2.5 million blog posts and 150 topics for the blog collection (divided over three TREC Blog track years, 2006–2008). We only use the title field of the topics, that is, the keyword query.

We use two baselines, QL (query likelihood with Jelinek-Mercer smoothing) and EXP (due to [7]) which is like QL but with a non-uniform document prior $P(D) = \beta \cdot e^{-\beta(time(q) - time(D))}$. We follow [3] and set the smoothing parameter $\lambda = 0.4$ and $\beta = 0.015$. We compare the baselines against a temporal query model with a trained γ and with an adaptive γ; for the trained model we split the dataset in different ways: (i) leave-one-out cross validation (LV1), and (ii) three-fold cross-validation split by topic sets over the years (YSPLIT). The temporal granularity for burst estimation is days. Based on preliminary experiments, we fix the number of results returned to $N = 250$, the number of top documents per burst to $N_B = 25$, and a term should occur in at least 10 of those documents. We return $M = 5$ terms per burst.

Table 1. MAP scores on TREC Blog track data sets

			trained γ		
Year	QL	EXP	(YSPLIT)	(LV1)	adaptive γ
2006	0.2571	0.2573	0.2783▲	**0.3875**▲	0.2776▲
2007	0.3742	0.3758	0.3798	**0.4220**▲	0.3838
2008	0.3088	0.3084	0.3165	**0.3423**▲	0.3118
all	0.3134	0.3138	0.3249▲	**0.3702**▲	0.3244▲

As to the results, in general, we can see in Table 1 that for the complete set of queries ("all") the MAP scores are significantly higher than the baselines for all approaches. The performance using a γ trained with LV1 is significantly better than using data set split over years. Thus, optimizing γ pays off: leave-one-out cross validation results in a different value from using split training sets (-0.8 vs. -0.9) and a significant improvement over using split training data. Queries modeled with LV1 seem to be more focused.

Table 1 provides an overview of the results split by year, and indeed, the significance of the results decreases, as over the years, the assessors gained distance from the data and specific events. Thus, later queries are not as temporally focused but more general. This is reflected in the number of temporal queries per collection: temporal queries are much more common in the 2006 collection (64%) than in 2008 (40%).

5 Conclusion

We have introduced two temporal query models, a trained model and an adaptive exponential model. We have analyzed their effectiveness on news-related data and found that they consistently improve MAP without decreasing precision on the TREC Blog track collection. Our models are situation dependent: given enough training data, the optimal parameter settings can be found using a grid search to tune the adaptive model, while the alternative model can be used in situations with insufficient training data.

Acknowledgments. This research was 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, 380-70-011, 727.011.005, the Center for Creation, Content and Technology (CCCT), the Hyperlocal Service Platform project funded by the Service Innovation & ICT program, the WAHSP project funded by the CLARIN-nl program, under COMMIT project Infiniti and by the ESF Research Network Program ELIAS.

References

[1] Balog, K., Bron, M., de Rijke, M.: Category-Based Query Modeling for Entity Search. In: Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., van Rijsbergen, K. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 319–331. Springer, Heidelberg (2010)
[2] Dakka, W., Gravano, L., Ipeirotis, P.G.: Answering General Time Sensitive Queries. In: CIKM 2008, pp. 1437–1438 (2008)
[3] Efron, M., Golovchinsky, G.: Estimation Methods for Ranking Recent Information. In: SIGIR 2011, pp. 495–504 (2011)
[4] Jones, R., Diaz, F.: Temporal Profiles of Queries. ACM Trans. Inf. Syst. 25(3), Article 14 (2007)
[5] Keikha, M., Gerani, S., Crestani, F.: TEMPER: A Temporal Relevance Feedback Method. In: Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V. (eds.) ECIR 2011. LNCS, vol. 6611, pp. 436–447. Springer, Heidelberg (2011)
[6] Lavrenko, V., Croft, W.B.: Relevance-Based Language Models. In: SIGIR 2001, pp. 120–127 (2001)
[7] Li, X., Croft, W.B.: Time-Based Language Models. In: CIKM 2003, pp. 469–475 (2003)
[8] Manning, C., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press (2008)
[9] Massoudi, K., Tsagkias, M., de Rijke, M., Weerkamp, W.: Incorporating Query Expansion and Quality Indicators in Searching Microblog Posts. In: Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V. (eds.) ECIR 2011. LNCS, vol. 6611, pp. 362–367. Springer, Heidelberg (2011)
[10] Meij, E., de Rijke, M.: Supervised Query Modeling Using Wikipedia. In: SIGIR 2010, pp. 875–876 (2010)
[11] Ounis, I., de Rijke, M., Macdonald, C., Mishne, G., Soboroff, I.: Overview of the TREC-2006 Blog Track. In: TREC 2006, Gaithersburg, USA (2006)
[12] Weerkamp, W., de Rijke, M.: Credibility Improves Topical Blog Post Retrieval. In: Proceedings of ACL 2008: HLT, pp. 923–931 (2008)