Tech Review: Beta-Gamma Thresholds

Recommendation systems fulfill users' long-term information needs, such as a desire to be sent documents relevant to a topic of their interest. Examples of recommendation systems include notifications for new product recommendations, movie reviews, news articles, and sports scores. While a recommendation system is different from a search engine, where users explicitly ask for information using a query, both types of systems need to be able to rank documents, based on some kind of scoring system, in order to return the most relevant documents to the user. Where these systems differ, is that a search engine can return all of the available documents, ranked, while a recommendation system must only send relevant documents to the user.

In order to determine which documents should be sent to the user, and which ones should not, a recommendation system must employ a threshold to their ranking system. If a document's rank lies above the threshold, it is sent to the user, while if it falls below the threshold, it is not. In addition, as new documents are added to the recommendation system, this threshold-based filtering must adapt to the relevance of these new documents. A document that would have been sent to the user when the corpus of topical documents was small, may lose its relevance as the corpus grows over time. As a result of this need, an area of research known as Adaptive Information Filtering was born.

Content-Based Filtering is one aspect of Adaptive Information Filtering, and focuses on determining how relevant a document is for a particular topic. For example, the original implementation of the recommendation system leveraging beta-gamma thresholds used topic profiles consisting of a term vector, inverse-document frequency (IDF) statistics, and a score threshold. A per-topic score for each document was computed from each profile's term vector and IDF statistics, while the threshold was used to determine if the document should be recommended to the user. The recommendation system updated the threshold as new documents were added to the recommendation system, using a technique referred to as "beta-gamma adaptive threshold regulation."

Beta-Gamma Adaptive Threshold Regulation calculates a threshold, θ, based on an interpolation between the optimal-utility set of results (θ_{opt}) and zero-utility set of results (θ_{zero}). The interpolation coefficient, α, was modeled as a function of β and γ, where β is a parameter chosen between 0 and 1, inclusive, to account for any bias from the training set and γ is a parameter representing the level of confidence in the optimal threshold based on the number of training documents in the corpus, *N*. The formulas used to compute θ from these inputs are:

$$\theta = \alpha * \theta_{zero} + (1-\alpha) * \theta_{opt}$$

$$\alpha = \beta + (1-\beta) * e^{-\gamma N}$$

Beta-gamma threshold regulation was shown to be more effective for returning documents using F1 and F3 utility functions than a constant α against multiple document corpuses.

Later work found that the term vector and threshold should be updated after a minimum number of documents have been added, or after a maximum delay where no new documents have been added. This research found that higher frequency of threshold updates consistently improved the performance of the recommendation system, after enough samples have been added. Frequent updates early on, when the corpus is small, tied the recommendation system's performance to the initial training samples. If the initial training samples were of poor quality, the profile's long-term performance would suffer severely.

It is worth noting that beta-gamma thresholding is a learning algorithm, and likewise it cannot be used to set an initial threshold for any particular topic profile. Gamma-beta thresholding is usually used in combination with a "Delivery Ratio" threshold. This initial threshold is calculated based on an initial estimation of how many documents to send to the user, as a percentage of the total number of documents available for that topic. Since each document is ranked, the most useful documents would be initially returned, until enough feedback has been received for the beta-gamma threshold regulation to take over.

Further work found a heuristic of $\gamma = 0.1$ to be effective when β was set to 0, even when the initial delivery ratio returned results of a negative utility. While the subsequent recommendations, made after the beta-gamma thresholds were tuned, never reached maximum utility, they consistently overcame the negative utility of the initial delivery ratio threshold, and recommended a set of positive-utility documents. The $\gamma = 0.1$ setting also outperformed other parameters of $\gamma = 0.000125, 0.00025, 0.005, 0.001$, and 0.002. In the original paper, the authors found $\gamma = 0.05$ to be the most effective for their corpus.

Beta-gamma thresholding was also proven useful in research with support vector machines (SVMs) applied to text classification. While the term-vector model views each document as a vector in a multidimensional space, the support vector machine model classifies documents into a positive-result class and negative-result class on a hyperplane. While SVMs provide a strong ranking system, the beta-gamma threshold can be used to improve the recall of such a system without severely impacting precision.

SVMs can also be used to compute an initial beta-gamma threshold if given a utility function to start. This is valuable because the model can provide more accurate results earlier on as documents are added. Together, it was shown that computing an initial beta-gamma threshold and adapting it over time, led to much better results in the recommendation system built with support vector machines.

In conclusion, beta-gamma thresholding is a powerful mechanism to allow recommendation systems to improve its recommendations over time. It has proven valuable across multiple document corpora, either as a system based on term vectors or support vector machines. Finally, the optimal value for γ was found to be 0.1 when tested against multiple corpora.

References

Chengxiang Zhai, Peter Jansen, Emilia Stoica, Norbert Grot and David A. Evans,"Threshold Calibration in CLARIT Adaptive Filtering", in: Voorhees, E.M. & D.K. Harman (Editors), The Seventh Text REtrieval Conference (TREC-7). NIST Special Publication 500- 242. Washington,DC: U.S. GovernmentPrinting Office, 1999, 149-156.

Chengxiang Zhai, Peter Jansen, Norbert Roma, Emilia Stoica, and David A. Evans, "Optimization in CLARIT Adaptive Filtering", in: Voorhees, E.M. & D.K. Harman (Editors), TheEighth TextREtrieval Conference (TREC- 8).

Zhai, C., Jansen, P., & Evans, D. A. (2000). “Exploration of a heuristic approach to threshold learning in adaptive filtering.” *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, 360-362.

Shanahan, James & Roma, Norbert. (2003). “Improving SVM Text Classification Performance through Threshold Adjustment.” 2837. 361-372. 10.1007/978-3-540-39857-8_33.