

Engaging Learners in an Enterprise L&K System

Wesley M. Gifford, Ashish Jagmohan,
Yi-Min Chee, Anshul Sheopuri
IBM Research
Yorktown Heights, NY, USA

John Ambrose, Sue Rodeman,
Shota Aki
Skillsoft
Nashua, NH, USA

ABSTRACT

We describe a system being designed for a leading provider of enterprise learning solutions, to improve engagement among learners. The system consists of an engagement timing component which estimates a learner’s level of engagement and likely preferred interaction times, and a recommendation component which generates personalized content recommendations. We summarize early results from a recently initiated pilot deployment.

1. INTRODUCTION

The problem of interest is improving learner engagement in an enterprise learning system, by utilizing consumption data captured by the learning platform. The existing learning platform records each content launch, tracking user and content ID and launch time and duration, among other data. The platform also defines an expert-curated hierarchy of content, wherein assets are grouped into a forest of asset-folders on the basis of subject matter. We have developed an engagement system consisting of two major components: 1) an engagement timing component that is responsible for estimating both a learner’s level of engagement, and preference to interact at certain days and times; and 2) a recommendation component that generates personalized recommendations for each learner, based on historical learner activity. In an initial email-based pilot, these components have demonstrated significant improvements in user response compared to industry benchmarks.

2. ENGAGEMENT TIMING

The goal of the engagement system is to improve the level of engagement of its learners. The engagement timing component helps the system with proper timing of actions, based on each learner’s current level of engagement. We observed that learners often exhibit “bursty” or self-excitation behavior, where a learner’s interactions frequently occur in clusters. We model these interactions as arrivals from a stochastic process that captures the temporal dependencies in learner behavior; the typical homogeneous Poisson process is not capable of doing so.

For users with sufficient interaction histories, we consider a Hawkes’ process, which can be viewed as a counting process whose time-varying intensity function adheres to a spe-

cific structure that enables capturing of temporal dependencies. The original structure considered by Hawkes’ was [Hawkes 1971]: $\lambda(t) = \mu + \int_{-\infty}^t g(t-u; \theta) dN(u)$, where $N(u)$ is an appropriate point process. Hawkes’ specifically considered the case where $g(t) = \sum_{i=1}^P \alpha_i \exp\{-\beta_i t\}$, $t > 0$. This function states that the intensity at the current time consists of decayed contributions from prior events. If only a short period of time has elapsed since the learner’s last action, the intensity function is impacted by these recent events and hence captures the fact that the learner is more likely to reengage. If a long period has elapsed since the last action, the process behaves more like a homogeneous Poisson process with rate μ until the next action. Similar models have been used to model stock market trades and earthquake aftershocks.

The system estimates the level of engagement for each learner by considering their reengagement probability in a time window given their prior interaction history. For learners with sufficient history, the parameters of the model above can first be determined using maximum likelihood estimation. This was done using numerical maximization of the likelihood, whose expressions are available in [Ozaki 1979]¹. Then, the probability that a particular learner reengages in the next s days, given their prior interaction history, is equivalent to the event that there is at least one arrival in the time period of interest from the underlying stochastic process (details omitted for brevity).

In addition to knowing a learner’s engagement level, it is also important to know the best time of day and day of week to contact individual learners. This is derived from a learner’s prior interactions under the assumption that prior interaction times are indicative of preferred interaction times. Each day of the week is divided into n uniform duration bins, giving a multinomial distribution with a total of $7n$ categories. In many cases, estimation of the category probabilities suffers from sparsity due to limited interactions. This problem is solved by using Bayesian estimation with a Dirichlet prior that incorporates the aggregate preferences of the entire population. Results based on a test across multiple customers are shown in Figure 1. The preferences estimated in this test use an exponential weighting scheme (with parameter γ) to place more weight on recent activity. The plot indicates that the performance saturates for γ greater than 24 months. For this value of γ the estimated distribution significantly outperforms a naive model.

¹For learners with fewer prior interactions, one promising strategy is to aggregate their inter-arrival times and fit an aggregate model.

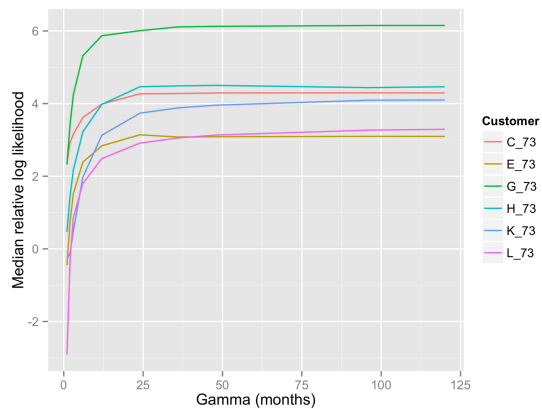


Figure 1: Performance of engagement time preference estimates relative to a naive estimate (uniform) for the case of 4 bins per day.

3. RECOMMENDATION ENGINE

The recommendation engine seeks to improve learner engagement by generating personalized recommendations for which the learner will likely have high preference and, hence, high consumption likelihood.

The recommendation engine utilizes a blended ensemble [Koren 2009], wherein several baseline recommenders are combined to yield a final set of recommendations. We use three groups of baseline recommenders, each group containing multiple individual recommenders. The first group consists of popularity-based recommenders which use several metrics to measure popularity, including temporal recency, and launch and duration information. The second group consists of content-based recommenders, wherein the learner’s historical consumption of certain asset-types, as determined by the expert-curated hierarchy, is leveraged to generate new recommendations. These include recommenders based on generative Bayesian models of the learner’s type-preferences, and based on tfidf type metrics over the content hierarchy. The third group consists of collaborative-filtering recommenders, which leverage the implicit feedback information [Hu et al. 2008] manifested in each user’s historical asset consumption activity; individual recommenders include some based on matrix factorization, and others based on separate user-user and asset-asset based filtering.

The recommender ensemble described above is combined to generate a final set of recommendations (typically 5-10) for each learner. Activity data was temporally split into training and validation data sets. We used several metrics to quantify recommender goodness, including metrics based on discounted cumulative gain and precision, and predictive metrics quantifying the number of trained recommendations which were consumed in the validation set. The metrics yielded largely consistent results. We tried multiple blending techniques including gradient-boosted decision trees and random-forests; random forests were found to yield best performance. Figure 2 shows, for one enterprise, a comparison of the blended recommender to the best popular recommender, as the number of recommendations varies. The performance is normalized to that of the best popular recommender. Note that this comparison is aggregated over all learners, including a significant number with no prior historical activity, for whom the popular recommender is best.

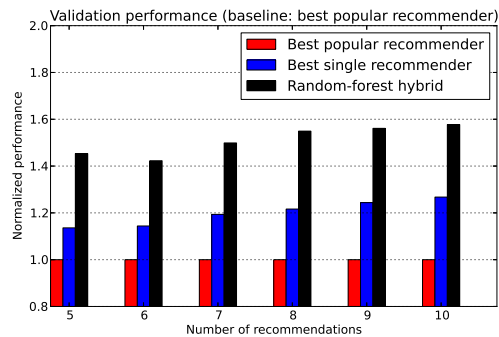


Figure 2: Performance comparison of recommenders. Metric: number of trained recommendations consumed in the validation set.

We created visualizations to help learners understand why they were receiving specific recommendations. One type of visualization shows the relative strength of each recommendation along each of the three broad recommender groups described above. Another set of visualizations compares the strength of the recommendations along a single dimension using their relative ranks from a specific recommender group. Anecdotal evidence indicates that users found these visualizations to be useful in helping to determine which of the recommendations might be of interest.

4. PRELIMINARY EVALUATION

Pilot deployments of the described engagement solution have been recently initiated. Learners receive emails at engagement times determined as in Section 2, containing personalized recommendations as described in Section 3. Some preliminary quantitative indications of the efficacy of the solution have been gleaned by examining initial email interaction metrics. After the first set of emails was sent to all participants, the click-through and click-to-open rates (ie. the fraction of participants who clicked upon one of the recommendations after opening the email) were tracked. The overall click-through rate was 5.6%, while the click-to-open rate was 31.6%. These metrics were compared to industry benchmarks for email campaigns in the education industry, as reported in [Silverpop 2014]. The comparison shows that the both rates are significantly higher than the industry median (2.8% and 14.3% respectively). These metrics give some preliminary confirmation of the promise of the proposed engagement approach.

5. REFERENCES

- [Hawkes 1971] Alan G Hawkes. 1971. Point spectra of some mutually exciting point processes. *JRSS. Series B* (1971), 438–443.
- [Hu et al. 2008] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *Proc. ICDM 2008*. 263–272.
- [Koren 2009] Yehuda Koren. 2009. The BellKor Solution to the Netflix Grand Prize. (2009).
- [Ozaki 1979] T Ozaki. 1979. Maximum likelihood estimation of Hawkes’ self-exciting point processes. *Ann. Inst. Stat. Math.* 31, 1 (1979), 145–155.
- [Silverpop 2014] Silverpop. 2014. 2014 Silverpop email marketing metrics benchmark study. (2014).