Internal Promotion Optimization

Rupesh Gupta
LinkedIn Corporation
Mountain View, CA, USA
rugupta@linkedin.com

Guangde Chen
LinkedIn Corporation
Mountain View, CA, USA
guchen@linkedin.com

Shipeng Yu
LinkedIn Corporation
Mountain View, CA, USA
siyu@linkedin.com

ABSTRACT

Most large Internet companies run internal promotions to cross-promote their different products and/or to educate members on how to obtain additional value from the products that they already use. This in turn drives engagement and/or revenue for the company. However, since these internal promotions can distract a member away from the product or page where these are shown, there is a non-zero cannibalization loss incurred for showing these internal promotions. This loss has to be carefully weighed against the gain from showing internal promotions. This can be a complex problem if different internal promotions optimize for different objectives. In that case, it is difficult to compare not just the gain from a conversion through an internal promotion against the loss incurred for showing that internal promotion, but also the gains from conversions through different internal promotions. Hence, we need a principled approach for deciding which internal promotion (if any) to serve to a member in each opportunity to serve an internal promotion. This approach should optimize not just for the net gain to the company, but also for the member's experience. In this paper, we discuss our approach for optimization of internal promotions at LinkedIn. In particular, we present a cost-benefit analysis of showing internal promotions, our formulation of internal promotion optimization as a constrained optimization problem, the architecture of the system for solving the optimization problem and serving internal promotions in real-time, and experimental results from online A/B tests.

CCS CONCEPTS

• **Computing methodologies → Machine learning approaches**; • **Mathematics of computing → Mathematical optimization**; • **Applied computing**;

KEYWORDS

Optimization, internal promotion, internal cross-promotion, machine learning

ACM Reference Format:
Rupesh Gupta, Guangde Chen, and Shipeng Yu. 2019. Internal Promotion Optimization. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3292500.3330715

ACM ISBN 978-1-4503-6201-6/19/08...$15.00
https://doi.org/10.1145/3292500.3330715

1 INTRODUCTION

These days most large Internet companies offer a suite of products and services to their members. For example, Google offers Gmail for email, YouTube for video sharing and Google Maps for navigation. Similarly, LinkedIn offers a wide range of products to its members. Some examples are, the LinkedIn flagship product[1] that helps members build their professional identity and engage with their professional network, LinkedIn Jobs that helps job candidates find opportunities, LinkedIn Learning that helps members develop new skills through online courses and LinkedIn Sales Navigator that helps sales professionals find and target the right buyers. At times, a member using one product might also benefit from another product. For example, a member building her professional identity on the LinkedIn flagship product with an intent to find a job is likely to also benefit from the LinkedIn Jobs and LinkedIn Learning products. However, the member might not be aware of these products. One way to educate the member about the availability of these products is to show advertisements for these products on the LinkedIn flagship product. An example is shown in Figure 1. If the member converts through these advertisements then she starts obtaining additional value from the ecosystem of LinkedIn products, which in turn drives engagement and/or revenue for LinkedIn. Such in-house advertisements on in-house products are called internal promotions. Some internal promotions on YouTube are shown in Figure 2.

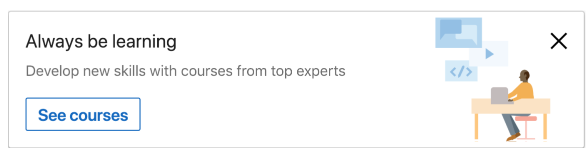
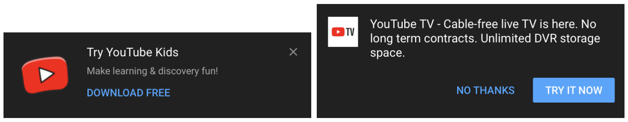


Figure 1: An internal promotion on LinkedIn

Internal promotions that are used for advertising products on other products are more specifically called internal *cross-promotions*. Internal cross-promotions are not the only type of internal promotions. Internal promotions are also used for educating members on how to obtain additional value from a product that they already use. This is achieved by recommending certain actions to members that are not directly in line with the value proposition of a page that they are viewing on a product. For example, an internal promotion on the feed page of the LinkedIn flagship product recommending a member to complete her profile (Figure 3). The feed page is designed to help a member engage in meaningful conversations with her professional network, and a recommendation to complete her profile is not directly in line with the value proposition of the feed

[1] https://www.linkedin.com



(a) (b)

Figure 2: Internal promotions on YouTube

page. The idea behind showing this internal promotion is that if the member converts through this internal promotion then she starts receiving additional value from the LinkedIn flagship product in the form of increased networking opportunities. This in turn drives engagement and/or revenue for the LinkedIn flagship product.

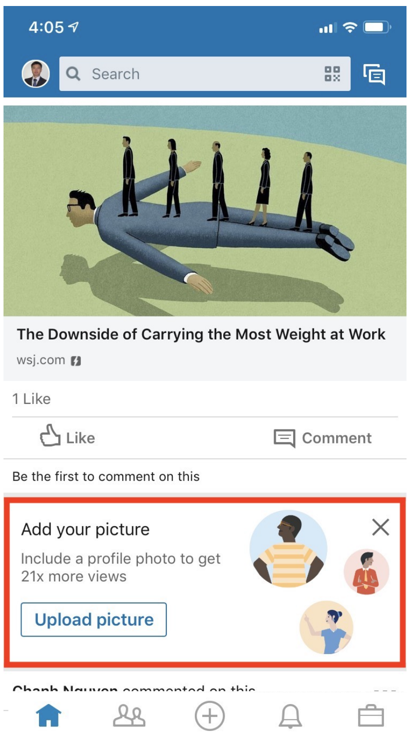


Figure 3: An internal promotion on the feed page of the LinkedIn flagship product

Although we do expect internal promotions on every page to increase the overall engagement of members with a company's suite of products in the long term, they can distract members from carrying out the task for which they visited a page. For example, if the internal promotion in Figure 1 is shown to a member on the LinkedIn flagship product, then the member is taken to the LinkedIn Learning product from the LinkedIn flagship product if she clicks on this internal promotion. This might distract her from building her professional identity. Similarly, if a member clicks on the internal promotion in Figure 3 then she is taken to the profile page from the feed page of the LinkedIn flagship product. This might distract her from engaging with her professional network. Such a loss in engagement with a product or page is called cannibalization loss. To mitigate this loss, we avoid showing more than one internal promotion on each page. However, even with this restriction, we need to be very careful in deciding whether to show any internal promotion, and which internal promotion to show. The decision should be based on the cost and benefit of showing each internal promotion. The cost should include not just the cannibalization loss, but also the potential negative impact of an internal promotion on member experience. This is the problem of internal promotion optimization. Note that this is in sharp contrast with the standard advertisement optimization problem where the sole objective is to maximize revenue. It is assumed that showing an advertisement does not cost anything to the page on which it is shown, and most pages in fact have slots reserved exclusively for advertisements. So an advertisement optimization system is incentivized to always show an advertisement if one is available (unless the advertisement is very low quality). However, in an internal promotion optimization problem, there is more than one objective. The internal promotion optimization system needs to maximize the gains from showing internal promotions and minimize the losses incurred for showing these internal promotions while ensuring good member experience.

This paper covers the challenges, our approach and our experiences with regard to the problem of internal promotion optimization at LinkedIn. Our contributions are summarized below:

- We discuss the problem space, in terms of the types of internal promotions, in Section 3.
- We present a cost-benefit analysis of internal promotions in Section 4.
- We describe our formulation of internal promotion optimization as a constrained optimization problem in Section 5.
- We outline the architecture of our system supporting internal promotion optimization in Section 6.
- We report real world experimental results based on online A/B tests in Section 7.

2 RELATED WORK

Internal promotions and cross-promotions [4] have been used by marketers as low cost marketing communication channels for over several decades. With the rise of online marketing, there is advice available from marketers, such as in [12] and [9], on how to create effective online cross-promotions. There are even some tools available, such as [1] and [8], for running online internal promotions and cross-promotions.

Research in the area of cross-promotions has been mainly focused on strategies for selecting the right allies for running external cross-promotions [11]. There is little literature on running and optimizing internal cross-promotions.

Literature that is related to optimization of internal promotions is the literature on content recommendation with multiple objectives (such as [2]), and on advertisement optimization with inventory management (such as [6]). In [2], the authors describe a content recommendation approach that jointly optimizes for clicks and post-click downstream utilities (such as time spent on landing pages). In [6], the authors describe an advertisement optimization approach that maximizes clicks, under constraints on the minimum number of impressions of each advertisement. Although similar in spirit, these approaches are not directly applicable to optimization of internal promotions as they do not take into account either the cannibalization loss or member experience in their formulations.

The multi-objective optimization framework introduced in [2], along with its solution through Lagrangian duality introduced in [3], has been successfully leveraged in several optimization problems such as [7]. Our work also leverages this multi-objective optimization framework to balance the multiple objectives of maximizing gains from internal promotions, minimizing the cannibalization loss due to internal promotions, and minimizing the negative impact of internal promotions on member experience.

3 PROBLEM SPACE

We have several active internal promotions in our system at the time of this writing. Some examples are shown in Figures 1, 3 and 4. Each internal promotion is aimed at educating members about other LinkedIn products, or educating them on how to obtain additional value from a product that they already use. Each internal promotion has a call to action button such as the "Try Free Month" button on the internal promotion in Figure 4(a). An internal promotion is considered successful if it is able to convert members through its call to action. However, the way conversion is defined is different for each internal promotion. For example, for the internal promotion in Figure 1, a subscription to the LinkedIn Learning product is a conversion; for the internal promotion in Figure 3, a profile photo upload is a conversion; for the internal promotion in Figure 4(a), a trial of the LinkedIn Sales Navigator product is a conversion; and for the internal promotion in Figure 4(b), a review of the help page describing how to handle harassment or safety concerns is a conversion. We broadly categorize our internal promotions into 3 categories based on their objectives:

(1) Monetization: A monetization internal promotion aims to drive revenue. An example is the internal promotion in Figure 1.
(2) Engagement: An engagement internal promotion aims to drive some engagement metric. An example is the internal promotion in Figure 3.
(3) Branding: A branding internal promotion aims to drive the Net Promoter Score[2]. An example is the internal promotion in Figure 4(b).

Based on its objective, each internal promotion targets only a subset of the total member base of LinkedIn. For example, it makes no sense to show the internal promotion in Figure 1 to a member who is already subscribed to the LinkedIn Learning product. Similarly it would be unwise to show the internal promotion in Figure 4(a) to a member who is not a sales professional.

Each internal promotion also has a dismiss button, as highlighted in Figure 4(a). If a member dismisses an internal promotion then that internal promotion is not shown to this member again for a long period of time.

At this time, these internal promotions can only appear on the feed page of the LinkedIn flagship product (Figure 3). We do not allow more than one internal promotion on this page. This internal promotion can be served at a fixed position in the feed, a few positions below the position occupied by the first sponsored update. LinkedIn's more than 600 million members visit the feed page billions of times in a week, which means that we get billions of opportunities to serve internal promotions to our members in a week. Of these billions of opportunities, there are hundreds of millions of opportunities in which members scroll down to the feed position where an internal promotion can appear.

[2]https://www.netpromoter.com/know/

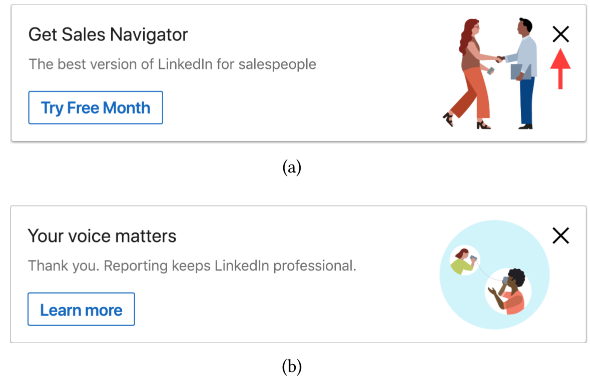


Figure 4: Internal promotions with very different calls to action. The red arrow in (a) points to the dismiss button

4 COST-BENEFIT ANALYSIS

In order to empirically examine the cost and benefit of internal promotions, we set up an online experiment as follows. We created a bucket of members by randomly selecting a small percentage of members from our member base. We then served internal promotions to members in this bucket through a *random* serving scheme. When a member in this bucket visited the feed page of the LinkedIn flagship product, we first fetched all internal promotions that targeted this member and then served a random internal promotion from this set of eligible internal promotions. To ensure that the member experience was not too poor, we did not show the same internal promotion to a member more than twice within a week. So, if a member was being targeted by only one internal promotion then that member saw at most two internal promotion impressions in a week. We denote this bucket of members as the *random* bucket. Members outside of this bucket did not see any internal promotion. We denote this other bucket of members as the *control* bucket.

We collected data from this experiment over a period of one week. Over this period, a large number of members in the *random* bucket saw several different types of internal promotions. We compared metrics from this *random* bucket with metrics from the *control* bucket.

We observed a reasonable click rate on internal promotions, and a corresponding increase in the various conversion metrics that the various internal promotions aimed at driving.

But we also saw a drop in various feed engagement metrics, as tabulated in Table 1. In this table, the drop in metrics was measured only over those members who visited the feed page. For example, members in *random* bucket who visited the feed page performed 1.2% less viral actions (like, comment, share) compared to the viral actions performed by such members in the *control* bucket. We also observed a statistically insignificant decrease in revenue from sponsored updates in the *random* bucket. This informs us that we need to weigh the gains and losses carefully when deciding whether to show any internal promotion.

Metric	Δ
Viral Actions	-1.2%
Likes	-1.23%
Video Plays	-0.76%
All Interactions	-0.55%

Table 1: Delta in feed page metrics, *random* vs. *control*

We also observed a large number of dismiss actions, of the order of the number of clicks, on internal promotions. This shows that members dislike irrelevant internal promotions as much as they like relevant internal promotions. This informs us that member experience needs to be factored into our decision of which internal promotion (if any) to show to a member.

5 PROBLEM FORMULATION

As described in Section 3, the conversion metrics for the various internal promotions can be very different. We also established in Section 4 that on the one hand internal promotions drive various conversion metrics, but on the other hand they can also cannibalize engagement and/or revenue from the page on which they are shown. If we had a way to compare a unit gain/loss in one metric against a unit gain/loss in another metric then we could easily compute the net gain as follows. We would map the gain in each conversion metric to a single metric and add them all to obtain the total gain. We would also map the cannibalization loss to that same metric to obtain the total loss. We would then subtract the total loss from the total gain to obtain the net gain in a single metric. In this case, the internal promotion optimization problem would be to simply maximize this net gain while ensuring good member experience.

However, a comparison across different metrics is very difficult. This is because there can be long delays before the impact of a unit gain/loss in one metric reflects on another metric. Let's say we want to map a unit gain in every conversion metric to revenue. Then, if a member converts on an internal promotion asking the member to try the Sales Navigator product (Figure 4(a)), it might take several days before the member becomes a paid subscriber of that product and moves the revenue metric. In another example, if a member converts on an internal promotion asking the member to review the help page describing how to handle harassment or safety concerns (Figure 4(b)), it might take months before the member builds trust in the LinkedIn flagship product, starts using the product more often, and then eventually clicks an advertisement which moves the revenue metric.

5.1 Budgeting

Inability to compare across different metrics means that we can neither decide what internal promotion to show to a member (as we cannot compare the conversion metric of one internal promotion with the conversion metric of another internal promotion), nor decide whether to show any internal promotion (as we cannot compare the conversion metrics of internal promotions with the cannibalization loss metric(s)). To overcome this problem we take the following approach.

- First, we pick a time duration over which we are going to make decisions on which internal promotion (if any) to show. We choose a time duration of one week.
- Then, the product owner of the page where internal promotions are going to be shown picks a metric of interest and sets a cap on the maximum acceptable loss in that metric for showing internal promotions. Let's say the owner of the feed page picks viral actions as the metric of interest and sets a cap of 0.6% loss in viral actions (which is 50% of the loss observed in the *random* bucket from Table 1). Then this implies that the feed page should not incur a loss of more than 0.6% viral actions for showing internal promotions. We denote this cap on the loss as L.
- Next, we compute an approximate maximum number of internal promotion impressions permissible within this cap on the loss. This computation is based on the data collected from the *random* bucket. In our example, since the cap was set at 50% of the metric loss observed in the *random* bucket, we approximate the maximum permissible internal promotion impressions as 50% × (number of internal promotion impressions in the *random* bucket) × (number of members to whom we want to show internal promotions)/(number of members in the *random* bucket). We denote this approximate maximum permissible internal promotion impressions as N.
- Next, a central authority guarantees a minimum number of impressions to some or all of the internal promotions. This guarantee is based on the central authority's own judgment of the relative importance of the conversion metrics of different internal promotions, as well as product priorities. We denote the minimum impressions guaranteed to internal promotion j as N_j (N_j may be 0). The sum total of these guaranteed impressions should be smaller than N.

This budgeting approach delegates the responsibility of comparison across different metrics to a central authority that manages all internal promotions. With this budgeting mechanism in place, there is a cap set on the cannibalization loss, and it can be assumed that a conversion from each internal promotion is equally valuable. Our goal then is to maximize the total number of conversions while ensuring good member experience. We use dismiss rate as a measure of member experience. So we formulate an optimization problem to maximize conversions and minimize the dismiss rate, under the budgeting constraints. The solution to the optimization problem tells us which internal promotion (if any) to serve when a member visits the feed page. The mathematical details are explained next.

5.2 Optimization

We would like to maximize the total number of conversions from all internal promotions and minimize the average dismiss rate across all internal promotions under the following constraints:

(1) Each internal promotion j should be impressed at least its guaranteed N_j number of times.
(2) The total loss incurred for showing internal promotions should not exceed L.

We make the following modifications to be able to solve the above optimization problem:

- It is not possible to simultaneously optimize for two objectives (conversions and dismiss rate). So we maximize conversions and set a constraint on dismiss rate. To set an appropriate value for this constraint, we once again use the data collected from the *random* bucket in Section 4. If the average dismiss rate in the *random* bucket is $dr(random)$ then we constrain the dismiss rate to be less than $f \times dr(random)$, where $f \in (0, 1)$. Let's denote this by r.
- A constraint on L will require us to compute the expected loss from serving each internal promotion in each opportunity to serve an internal promotion. Accurately predicting the loss is very difficult as the loss heavily depends on the context, such as the other items in the feed, the size of those items, etc. So we replace this constraint with a constraint on the approximate maximum number of internal promotion impressions N. Since this constraint on N is an approximation of the constraint on L, we put a loss monitoring mechanism in place to stop all internal promotions if the actual loss ever exceeds L. This mechanism is described in Section 6.3.

Now the optimization problem is to maximize the total number of conversions from all internal promotions under the following constraints:

(1) Each internal promotion j should be impressed at least its guaranteed N_j number of times.
(2) The total number of internal promotion impressions should not exceed N.
(3) The average dismiss rate across all internal promotions should not exceed r.

We now express the above problem mathematically as follows. As mentioned in the previous section, we would like to solve the optimization problem for a one week duration. Over this one week period, each anticipated view of the feed page by a member is an opportunity to serve an internal promotion. Let's denote an opportunity by i and the entire set of all the opportunities by I. As before, we denote an internal promotion by j and the entire set of internal promotions by J. Let's say that we have a function that gives us the probability c_{ij} of the member viewing the feed page in opportunity i to convert on an internal promotion j given that she views the internal promotion. Let's say that we also have a function that gives us the probability d_{ij} of the member viewing the feed page in opportunity i to dismiss an internal promotion j given that she views the internal promotion. Note that some of the $|J|$ internal promotions might not target the member viewing the page in opportunity i. For such an internal promotion j, we set c_{ij} and d_{ij} to large negative and large positive numbers respectively, such as, $c_{ij} = -10^4$ and $d_{ij} = 10^4$. This makes the subsequent formulation easy to read. Let x_{ij} be the probability with which we serve internal promotion j in opportunity i. This is the serving plan that we would like to obtain. Also, let s be the probability with which a member viewing the feed page scrolls down to the position where an internal promotion can be served. This probability would ideally depend on the member as well as the content of the page, but for simplicity we assume it to be a constant. Then $x_{ij} \times s$ becomes the probability with which internal promotion j is impressed on the member viewing the page in opportunity i. Then, we would like to obtain the optimal serving plan x_{ij} so as to:

$$\max_{x_{ij}} \quad \sum_i \sum_j c_{ij}\, x_{ij}\, s \tag{1}$$

s.t.

$$\sum_i x_{ij}\, s \geq N_j \quad \forall j \tag{2}$$

$$\sum_i \sum_j x_{ij}\, s \leq N \tag{3}$$

$$\frac{\sum_i \sum_j d_{ij}\, x_{ij}\, s}{\sum_i \sum_j x_{ij}\, s} \leq r \tag{4}$$

$$\sum_j x_{ij} \leq 1 \quad \forall i \tag{5}$$

$$0 \leq x_{ij} \leq 1 \quad \forall i, \forall j \tag{6}$$

In the above constrained optimization problem:

- $\sum_i \sum_j c_{ij}\, x_{ij}\, s$ in (1) is the expected total number of conversions from all internal promotions.
- $\sum_i x_{ij}\, s$ in (2) is the expected number of impressions of internal promotion j. Note that if we had the ability to map a unit gain in every conversion metric to a single metric (say the value v_j of a conversion from internal promotion j), then we would have multiplied v_j with c_{ij} in the objective (1) and removed the constraints in (2).
- $\sum_i \sum_j x_{ij}\, s$ in (3) is the expected total number of impressions of all internal promotions.
- $\sum_i \sum_j d_{ij} x_{ij}\, s$ in (4) is the expected total number of dismisses on all internal promotions. This divided by $\sum_i \sum_j x_{ij}\, s$ gives the expected dismiss rate across all internal promotions.
- $\sum_j x_{ij}$ in (5) is the expected number of internal promotions served in opportunity i. We want this number to be less than or equal to 1 to ensure that at most one internal promotion is selected for each opportunity. If this number is less than 1 for an opportunity i then there is a non-zero probability of serving no internal promotion in that opportunity.
- The last constraint (6) ensures that each x_{ij} is a probability.

Since s is a constant, we can slightly simplify the above problem to the following:

$$\begin{aligned}
\max_{x_{ij}} \quad & \sum_i \sum_j c_{ij}\, x_{ij} \\
\text{s.t.} \quad & \\
& \sum_i x_{ij} \geq n_j \quad \forall j \\
& \sum_i \sum_j x_{ij} \leq n \\
& \sum_i \sum_j d_{ij}\, x_{ij} \leq r \sum_i \sum_j x_{ij} \\
& \sum_j x_{ij} \leq 1 \quad \forall i \\
& 0 \leq x_{ij} \leq 1 \quad \forall i, \forall j
\end{aligned} \tag{7}$$

where $n_j = N_j / s$ and $n = N/s$.

Now it may seem that we can follow this simple approach for serving internal promotions. At the beginning of a week, forecast all the opportunities I for serving internal promotions in the upcoming week, solve the above linear program (LP), and use the optimal serving plan x_{ij} obtained to serve internal promotions. However, this is unrealistic in practice due to the following challenge: we do not have a way to accurately forecast the opportunities I that we will get in the upcoming week.

Fortunately, we observe that the distribution of member visits to the feed page does not change significantly week over week. So we use the opportunities received in the past week as a forecast for I. However, similarity in distribution does not provide us the solution $\mathbf{x}_i = [x_{i1}, ..., x_{iJ}]^T$ for every opportunity i that we will *actually* get in the upcoming week. To that end, we make use of a primal-dual trick that allows us to obtain the primal solution vector without using any part of the dual solution vector that depends on i. This trick is inspired by [3]. For this, we first add a quadratic regularization term to the objective of (7) to make the problem strongly convex.

$$
\begin{aligned}
\max_{x_{ij}} \quad & \sum_i \sum_j \left[c_{ij} x_{ij} - \frac{\lambda}{2}(x_{ij} - q)^2 \right] && (8) \\
\text{s.t.} \quad & \\
& \sum_i x_{ij} \geq n_j \quad \forall j && (9) \\
& \sum_i \sum_j x_{ij} \leq n && (10) \\
& \sum_i \sum_j d_{ij}\, x_{ij} \leq r \sum_i \sum_j x_{ij} && (11) \\
& \sum_j x_{ij} \leq 1 \quad \forall i && (12) \\
& 0 \leq x_{ij} \leq 1 \quad \forall i, \forall j && (13)
\end{aligned}
$$

where $q \in [0, 1]$ is some prior on the serving probability, and $\lambda > 0$ is a regularization parameter. This converts our LP to a quadratic program (QP). We now solve the dual of the primal problem above using the opportunities received in the past week as a forecast for I. Standard QP solvers are unable to handle a problem of our scale due to the large number of variables in the dual problem corresponding to the constraints (12) and (13). We employ an in-house large scale implementation of the operator splitting algorithm [10] to solve this QP. Let $\alpha_j \forall j$, β and γ be the solutions of the dual problem corresponding to the per internal promotion impressions constraints (9), the total impressions constraint (10) and the dismiss rate constraint (11) respectively.

Now for each actual opportunity i, we use the following approach to obtain $\mathbf{x}_i$. Note that we can rewrite the QP above as:

$$
\begin{aligned}
\max_{x_{ij}} \quad & \sum_i \sum_j \left[c_{ij} x_{ij} - \frac{\lambda}{2}(x_{ij} - q)^2 \right] \\
\text{s.t.} \quad & \\
& \sum_i x_{ij} \geq n_j \quad \forall j \\
& \sum_i \sum_j x_{ij} \leq n \\
& \sum_i \sum_j d_{ij} x_{ij} \leq r \sum_i \sum_j x_{ij} \\
& \mathbf{x}_i \in K_i \quad \forall i
\end{aligned}
$$

where K_i is the convex region defined by (12) and (13) for a given i. We can also write the above problem as:

$$
\begin{aligned}
\max_{x_{ij}} \quad & \sum_i \sum_j \left[c_{ij} x_{ij} - \frac{\lambda}{2}(x_{ij} - q)^2 \right] - \sum_i \mathbb{I}_{K_i}(\mathbf{x}_i) \\
\text{s.t.} \quad & \\
& \sum_i x_{ij} \geq n_j \quad \forall j \\
& \sum_i \sum_j x_{ij} \leq n \\
& \sum_i \sum_j d_{ij} x_{ij} \leq r \sum_i \sum_j x_{ij}
\end{aligned}
$$

where $\mathbb{I}_{K_i}(\mathbf{x}_i) = 0$ if $\mathbf{x}_i \in K_i$ and infinity otherwise. It can be shown that at the point of optimality:

$$
\mathbf{x}_i = \Pi_{K_i}\left(\frac{\mathbf{c}_i + \boldsymbol{\alpha} - \beta - \gamma \mathbf{d}_i + \gamma r}{\lambda} + q \right) \tag{14}
$$

where

$$
\begin{aligned}
\mathbf{c}_i &= [c_{i1}, ..., c_{iJ}]^T, \\
\boldsymbol{\alpha} &= [\alpha_1, ..., \alpha_J]^T, \\
\mathbf{d}_i &= [d_{i1}, ..., d_{iJ}]^T,
\end{aligned}
$$

and $\Pi_{K_i}(.)$ is a projection on to K_i. If we let $\mathbf{z}_i = \frac{\mathbf{c}_i + \boldsymbol{\alpha} - \beta - \gamma \mathbf{d}_i + \gamma r}{\lambda} + q$, then the projection can be written as a small QP in $|J|$ variables as follows.

$$
\begin{aligned}
\min_{\mathbf{x}_i} \quad & \|\mathbf{x}_i - \mathbf{z}_i\|_2^2 \\
\text{s.t.} \quad & \\
& \sum_j x_{ij} \leq 1 \\
& 0 \leq x_{ij} \leq 1 \quad \forall j
\end{aligned} \tag{15}
$$

Note that for an opportunity i, $\mathbf{z}_i$ can be computed easily using the dual solutions ($\alpha_j \forall j$, β and γ), the predicted probability of conversion for each internal promotion ($c_{ij} \forall j$) and the predicted probability of dismiss for each internal promotion ($d_{ij} \forall j$). After computing $\mathbf{z}_i$, the small QP in (15) can be solved using any standard QP solver to obtain $\mathbf{x}_i$. Then the probability of selection of internal promotion j for opportunity i is x_{ij}. Since $\sum_j x_{ij} \leq 1$, no internal promotion may be selected for opportunity i with a probability ≥ 0.

It is worth pointing out that (15) allows us to make a decision, on which internal promotion (if any) to serve, *independently* for each opportunity to serve an internal promotion. All that we need to maintain is a set of $|J| + 2$ coefficients: $\alpha_j \forall j$, β and γ. We call these *optimization coefficients*.

5.3 Response Prediction

The optimization problem above assumes availability of the probability c_{ij} of the member viewing the feed page in opportunity i to convert on an internal promotion j given that she views the internal promotion, and the probability d_{ij} of the member viewing the feed page in opportunity i to dismiss an internal promotion j given that she views the internal promotion. We obtain these probabilities from models trained on unbiased data collected from the *random* bucket described in Section 4. Each training example comprises of an internal promotion j that was impressed on a member m, and the response of m on j. We train two gradient boosted decision trees models, one for each of the two responses, conversion and dismiss. We choose tree models due to their ability to naturally incorporate interactions between features. We include 4 broad categories of features in our models, viz.,

(1) Member m's profile features such as industry, locale etc.
(2) Member m's activity features such as time since last visit, number of connection requests sent in the last 7 days, number of jobs viewed in the last 7 days, etc. These features help us capture the intent of m.
(3) Member m's past experience with internal promotions such as the last internal promotion viewed, the time since the last internal promotion was viewed, number of internal promotions viewed, clicked and dismissed in the last 7 days, etc.
(4) ID of the internal promotion j. We are able to use the ID of the internal promotion as a feature because the cardinality of this feature is small.

We train these models using XGBoost [5] in Spark [13].

We compute 2 metrics to measure the performance of our response prediction models:

(1) Area under the receiver operating characteristic curve (AUC): This is to verify whether a model is directionally correct, i.e., larger predictions for positive examples.
(2) Observed to expected ratio (O/E ratio): This is to verify the scale of the predicted probabilities from a model. This metric is computed as the number of positive test examples divided by the sum of predicted probabilities for all test examples. An O/E ratio close to 1 is desired. It is important to verify the scale of the predicted probabilities because the sum of predicted probabilities is used to approximate the actual number of conversions and dismisses in (1) and (4) respectively.

The AUC and O/E ratio metrics for our response prediction models are tabulated in Table 2.

Response	AUC	O/E ratio
Conversion	0.87	0.96
Dismiss	0.75	0.97

Table 2: Validation metrics for response prediction models

6 SYSTEM ARCHITECTURE

In this section, we outline the core components of our system that supports optimization of internal promotions. This system is illustrated through a block diagram in Figure 5. We can divide our system into three subsystems, viz., an online serving subsystem, an offline training subsystem, and a loss monitoring subsystem.

6.1 Online Serving Subsystem

As emphasized in Section 5.2, we are able to make a decision, on which internal promotion (if any) to serve, *independently* for each opportunity to serve an internal promotion. This makes the online serving subsystem simple, and enables scaling of the subsystem to billions of requests per week for serving internal promotions.

When a member visits the feed page of the LinkedIn flagship product, the Internal Promotions Fetcher fetches a list of internal promotions that target this member from the Internal Promotions Database. This list is passed down to the Response Prediction Engine. The Response Prediction Engine fetches features of the visiting member from the Tracking Data Store to create a partial feature vector. It then goes through each internal promotion in the list and appends the ID of the internal promotion to be scored to the partial feature vector to prepare the full feature vector. Finally, it uses the full feature vector of each internal promotion in the response prediction models supplied by the Response Prediction Model Trainer to predict the probability of conversion c_{ij} and dismiss d_{ij} for each internal promotion j in the list. These predictions are passed to the Optimization Decision Engine. The Optimization Decision Engine solves (15) with these predictions and the optimization coefficients ($\alpha_j \forall j,\ \beta,\ \gamma$) supplied by the Optimization Solver to obtain the solution $\mathbf{x}_i$. It then uses this $\mathbf{x}_i$ to serve one or no internal promotion. The member's interactions with the impressed internal promotion, along with her profile and activity data are recorded in the Tracking Data Store.

6.2 Offline Training Subsystem

Snapshots of the Tracking Data Store are loaded into HDFS. The snapshot data is used in the Response Prediction Model Trainer for training conversion and dismiss prediction models in Spark. These models are fed into the Response Prediction Engine as well as the Optimization Solver. The Optimization Solver employs these models on the opportunity data from the past to week to predict responses $c_{ij},\ d_{ij}$. It then uses these predicted responses and the supplied constraints N_j, N, f and solves the optimization problem in (8)-(13). The optimization coefficients ($\alpha_j \forall j,\ \beta,\ \gamma$) obtained are fed into the Optimization Decision Engine.

6.3 Loss Monitoring Subsystem

We always maintain a small *control* bucket of members to whom we do not serve any internal promotion. We periodically compare metrics in this *control* bucket against metrics in the *treatment* buckets where internal promotions are served. This helps us monitor both the gains and losses from showing internal promotions in each *treatment* bucket.

In Section 5.2, we had replaced the constraint on the maximum loss in the metric of interest L with a constraint on the *approximate* maximum number of internal promotion impressions N. To ensure

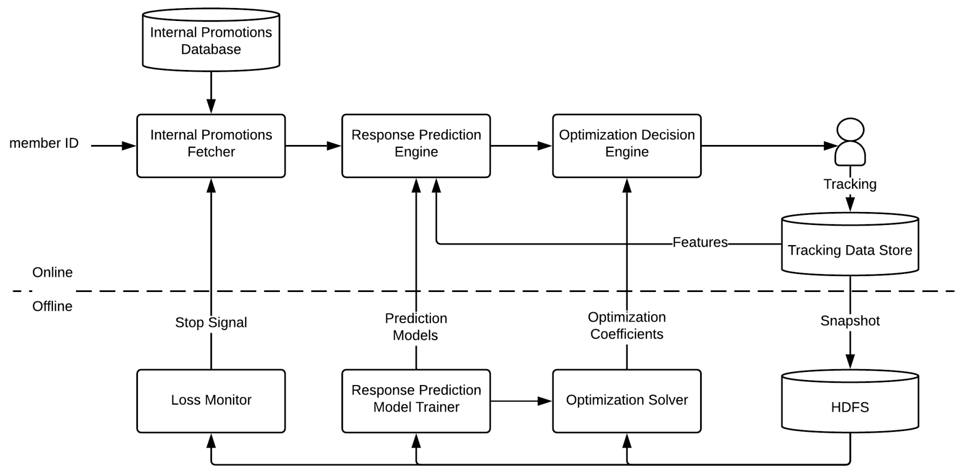


Figure 5: Simplified architecture of the system supporting optimization of internal promotions

that the actual loss never exceeds L, we need to stop showing internal promotions in a *treatment* bucket if the measured loss in the metric of interest in this bucket ever exceeds L. This is accomplished through the Loss Monitor in Figure 5. For example, let's say that the metric of interest is number of viral actions and $L = 0.6\%$. Then the Loss Monitor monitors the total number of viral actions in every member bucket. At the end of each day, if the difference in the number of viral actions in the *control* bucket and the number of viral actions in a *treatment* bucket exceeds 0.6% then the Loss Monitor sends a stop signal to the Internal Promotions Fetcher to stop serving internal promotions in that *treatment* bucket.

7 EXPERIMENTS & RESULTS

7.1 Experimental Setup

We set up an online A/B test experiment to evaluate our internal promotion optimization approach. We created three member buckets. Members in these three buckets were served internal promotions using different models. Members in the first bucket were served by a *random* model, same as the one used in Section 4. Members in the other two buckets were served by models obtained from solving constrained optimization problems in (1)-(6) with slightly different constraints. The constraints on per internal promotion impressions N_j were set by a central authority and hence identical for the two models. The owner of the feed page chose viral actions as the metric of interest and set a cap of $L = 1.2\%$ loss in this metric for showing internal promotions. The equivalent constraint on maximum internal promotion impressions N was also identical for the two models. The constraint on the dismiss rate (4) was however different for the two models.

(1) For the first model we set $r = 1$ which made the dismiss rate constraint inactive as the dismiss rate would always be less than or equal to 1. We call this the *optConversion* model as it optimizes only for conversions.

(2) For the second model we set $r = 0.95 \times dr(random)$ where $dr(random)$ was the average dismiss rate observed in the *random* bucket in Section 4. We call this the *optConversionDismiss* model as it optimizes for both conversions and dismiss rate. Note that setting r to a very small number and $N_j \forall j$ to large numbers can make the constrained optimization problem in (1)-(6) infeasible.

We chose a small value of the regularization hyper-parameter $\lambda = 10^{-4}$ in (8) as we did not want the prior on serving probability q to significantly affect the decisions. We set $q = N/(|I| \times |J|)$.

7.2 Results And Analysis

Since we solve the optimization problem for a one week duration, we measured metrics for all our models over a one week period. Our observations are summarized below.

- Both optimization models *optConversion* and *optConversionDismiss* satisfied the constraints on the minimum number of impressions of each internal promotion. These constraints were neither enforced, nor satisfied by the *random* model.
- Both optimization models satisfied the constraint on the maximum number of impressions of all internal promotions. This constraint was not enforced but satisfied by the *random* model as L was set equal to the loss observed in the *random* bucket in Section 4.
- The loss monitoring subsystem (from Section 6.3) did not send a stop signal for any of the models. This implies that the constraint on N was able to enforce a constraint on L.
- The *optConversionDismiss* model satisfied its constraint on dismiss rate.
- A comparison of the performance of the two optimization models against the *random* model is presented in Table 3. Model *optConversion*, that optimizes just for conversions, increased the number of conversions but also increased the number of dismisses. This indicates that the probability to convert is positively correlated with the probability to dismiss. If these two were negatively correlated then we would

not have the need for an explicit constraint on dismiss rate. Model *optConversionDismiss*, that optimizes for both conversions and dismiss rate, increased conversions by a smaller percentage compared to *optConversion* but reduced dismisses by a significant amount. It also reduced the number of impressions as it did not serve any internal promotion in opportunities where all internal promotions were very likely to be dismissed and unlikely to convert.

metric	*optConversion*	*optConversionDismiss*
conversions	+16.8%	+14.3%
dismisses	+6.9%	-7.2%
impressions	0%	-1.7%
dismiss rate	+6.9%	-5.5%

Table 3: A/B test results for two optimization models compared against the *random* model

Based on these observations, *optConversionDismiss* model was deployed for majority of members.

8 CONCLUSION AND FUTURE WORK

To the best of our knowledge, this is the first work describing a principled approach for optimization of online internal promotions, that takes into account the cannibalization loss incurred from showing internal promotions and the impact of internal promotions on member experience, along with the gains from conversions through internal promotions. We have introduced the challenging nature of the problem due to the difficulty of comparison across the different conversion metrics of different internal promotions and the cannibalization loss metric. It has been found that internal promotions not only cause a cannibalization loss but can also potentially degrade member experience on the product where they are shown. We have presented a budgeting mechanism for handling comparison across the different conversion metrics and the cannibalization loss metric. This budgeting mechanism guarantees a minimum number of impressions to some or all internal promotions, while capping the total number of impressions of all internal promotions. We have formulated the problem of which internal promotion (if any) to serve, for each opportunity to serve an internal promotion, as a constrained optimization problem. The objective of this constrained optimization problem is to maximize conversions and minimize the dismiss rate across all internal promotions, while respecting the budget constraints. We have discussed the challenges faced in solving this constrained optimization problem, as well as using the solution of the optimization problem for serving internal promotions. The approach presented for overcoming these challenges allows us to make a decision, on which internal promotion (if any) to serve, *independently* for each opportunity. This makes the design of the online serving subsystem very simple. The design of the other core components of the internal promotion optimization system, namely, offline training subsystem and loss monitoring subsystem has been outlined. Finally, the effectiveness of our internal promotion optimization approach has been demonstrated through results of online A/B test experiments run on the LinkedIn flagship product.

Interesting problems for future work include the following.

- Placement selection, i.e., incorporating the ability to answer the question of where to place an internal promotion from a given set of available placements on a product. This problem is complex because accurately predicting the cannibalization loss from different placements of an internal promotion is a hard problem. Accurately predicting the cannibalization loss is hard as it heavily depends on the context in which the internal promotion is going to be shown.
- Handling time-sensitive internal promotions. Some internal promotions may have a very short life span, such as, an internal cross-promotion for discounted subscription to a product on Thanksgiving Day. Such an internal promotion may need to be shown a guaranteed minimum number of times on a specific day, and cannot be shown on any other day. Although not impossible, it is somewhat difficult to handle such time-sensitive internal promotions with our current approach.

9 ACKNOWLEDGMENTS

We would like to thank our team at LinkedIn comprising of Curtis Wang, Tyler Elliott, Kiryl Yesipau, Yida Yao, Grant Hoffecker, Amitha Nazeer, Shruthi Ramesh, Sam Sedighian, Swapnil Ghike, Souvik Ghosh, Deepak Agarwal, Kinjal Basu and many others.

REFERENCES

[1] Adzerk. [n. d.]. Internal Promotions Platform | Adzerk. ([n. d.]). https://adzerk.com/solutions/internal-promotions/.
[2] Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, and Xuanhui Wang. 2011. Click Shaping to Optimize Multiple Objectives. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. ACM, New York, NY, USA, 132–140. https://doi.org/10.1145/2020408.2020435
[3] Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, and Xuanhui Wang. 2012. Personalized Click Shaping Through Lagrangian Duality for Online Recommendation. In *ACM SIGIR*. https://doi.org/10.1145/2348283.2348350
[4] Sam Ashe-Edmunds. [n. d.]. What Is a Corporate Cross Promotion? ([n. d.]). https://yourbusiness.azcentral.com/corporate-cross-promotion-2182.html.
[5] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785
[6] David Maxwell Chickering and David Heckerman. 2003. Targeted advertising on the web with inventory management. *Interfaces* 33, 5 (2003), 71–77.
[7] Rupesh Gupta, Guanfeng Liang, Hsiao-Ping Tseng, Ravi Kiran Holur Vijay, Xiaoyu Chen, and Romer Rosales. 2016. Email Volume Optimization at LinkedIn. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 97–106. https://doi.org/10.1145/2939672.2939692
[8] Abhishek Khurana. 2014. Cross Promotion Through House Ads With InMobi Analytics. (2014). https://www.inmobi.com/blog/2014/07/10/cross-promotion-through-house-ads-with-inmobi-analytics/.
[9] Sohan Maheshwar. 2014. 5 Things Developers Need to Know About Cross-Promoting Their Gaming App. (2014). https://www.inmobi.com/blog/2014/07/11/5-things-developers-need-to-know-about-cross-promoting-their-gaming-app.
[10] Brendan O'donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. 2016. Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding. *J. Optim. Theory Appl.* 169, 3 (June 2016), 1042–1068. https://doi.org/10.1007/s10957-016-0892-3
[11] TingTing Song and Qian Tang. 2015. Cross-Promotion in Social Media: Choosing the Right Allies. (2015).
[12] David Vaghari. 2017. The 6 Keys To Flawless Internal Cross Promotions. (2017). http://blog.upsight.com/blog/the-6-keys-to-flawless-internal-cross-promotions.
[13] Matei Zaharia, Reynold S. Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J. Franklin, Ali Ghodsi, Joseph Gonzalez, Scott Shenker, and Ion Stoica. 2016. Apache Spark: A Unified Engine for Big Data Processing. *Commun. ACM* 59, 11 (Oct. 2016), 56–65. https://doi.org/10.1145/2934664