Inventory forecasting for ad campaigns

Introduction
GumGum is a start-up operating in the online ad business and located in Santa Monica. Our advertising platform brings together website owners who would like to display ads on their site (called “publishers”) and companies that would like to advertiser on those sites (called “advertisers”). We specialize in two kinds of ads: in-image ads (where an ad banner is displayed within an image) and in-screen ads (where an ad banner is projected onto the whole page).

When an advertiser approaches us with a new campaign we need to determine how often we can show ads for that campaign. This in turn depends on three variables: campaign requirements (for instance, minimum image size, day of week, browser type), inventory availability (i.e. how many and what kinds of images do we expect to have that can serve the ad), and competition (for example, if the new campaign is an “Oscar” campaign, how many other “Oscar” campaigns are we expected to run in the same period).

Our current forecasting method uses a brute-force approach: given the parameters of a new campaign it analyzes one day’s worth of activities on our publishers’ websites to find out how often the ad would have been displayed. That number is then used to estimate inventory availability for the whole campaign. Obviously this approach has a number of shortcomings. One of them is cost - a single forecast takes about an hour and requires a distributed MapReduce job using over 30 servers.

We would like to replace it with a better method for inventory forecasting.

Proposal
The proposed method for inventory forecasting should accomplish the following:

- be cost effective - a single forecast should ideally not take more than a few minutes
- take advantage of 12 months of historic data instead of a single 24 hour window
- factor in competition
- factor in all other filters that may be applied to an ad, such as frequency caps, browser type, day of week, user data such as geolocation, income, etc
- provide our ad managers with lower and upper bounds for a given confidence level (for example, “forecasting between 60K and 80K impressions with 90% confidence”
Data and resources provided

The team will have access to all of our historic data since January 1st, 2013. Our company is using Amazon’s cloud computing infrastructure for most of our operations, so ideally we would make the data available there. We will also provide the team with any computing infrastructure it might request.

References

Here are a few pointers to scientific publications addressing the issue of inventory forecasting:

<table>
<thead>
<tr>
<th>LINK TO RESOURCE</th>
<th>NOTE</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://maxchickering.com/publications/ec00.pdf">http://maxchickering.com/publications/ec00.pdf</a></td>
<td>general introduction to the topic of inventory forecasting</td>
</tr>
<tr>
<td><a href="http://etd.lsu.edu/docs/available/etd-01092008-113450/unrestricted/MRahman.pdf">http://etd.lsu.edu/docs/available/etd-01092008-113450/unrestricted/MRahman.pdf</a></td>
<td>stressing two important issues we face: seasonal variation and lack of data (for example, if we run a campaign with limited historical data available)</td>
</tr>
<tr>
<td><a href="http://www.sciencedirect.com/science/article/pii/0925527395001360">http://www.sciencedirect.com/science/article/pii/0925527395001360</a></td>
<td>contains good introduction to the general problem; one interesting aspect of their approach is the ability to easily add factors to the forecasting process</td>
</tr>
</tbody>
</table>