How Much is Your Attention Worth? Analysis of Prices in LinkedIn Advertising Network

Chaolun Xia
Rutgers University
cx28@cs.rutgers.edu

Saikat Guha
Microsoft Research, India
saikat@microsoft.com

S. Muthukrishnan
Rutgers & MS, India
muthu@cs.rutgers.edu

ABSTRACT

In online advertising, advertisers bid on auctions to have their ads shown to targeted users. Online social networks like Facebook and LinkedIn offer very fine targeting controls over users, including their career, wealth and education information. They also provide Bid Suggestion, a function that maps targeting conditions to (suggested) bids needed to have their ads shown.

We address the question how much is the attention of a user worth by analyzing the bid suggestion from LinkedIn. (1) We build crawlers and run them for more than 100 days to harvest a unique dataset consisting of more than 100,000 suggested bids (for 188,260 distinct targeting conditions of 8 common user attributes) from LinkedIn. To analyze these bids, we also collect side data, e.g. GDP. (2) We explore these suggested bids, and discover many interesting and consistent results. For example, we find that suggested bids surprisingly have negative correlation with GDP and income, but the suggested bids are consistent with the demand and supply of the market.

1. INTRODUCTION

Online advertising is one of the pillars in the Internet industry. An online ad network allows advertisers to bid on reaching specific audience through its targeting language. Google AdWords\(^1\), the largest online ad network for instance, allows advertisers to target audiences based on search terms, the website (publisher) that the user browses, and simple user demographics (gender, age group, location). The price is set by a second-price auction \([4]\). Other online ad networks, specifically that run by Facebook, LinkedIn, Twitter and other Online Social Networks (OSNs), offer much finer targeting controls.

These OSNs contain detailed information shared directly by the user. This includes detailed educational records about the user, past and present employment experience, significant life events like changes in marital status or having a baby. This helps OSNs start to be in a position to offer advertisers significantly more control in precisely targeting their audience. LinkedIn, for instance, allows advertisers to target a software engineer in Microsoft, or a user who masters C++ but works in the medical industry.

Facing a variety of user segments, advertisers need guidance to compile and tune their ad campaigns. Fortunately, LinkedIn and Facebook satisfy the core of advertisers’ need by providing bid suggestion which is a function that, for each targeting condition, shows the suggested bid to win the auction and the number of users satisfying that condition. This is exciting because 1) suggested bids provide an economic view, i.e. the amount of money an advertiser has to pay to reach audience of their target, and 2) for the first time in the history of advertising, these prices are now transparent for very fine characteristics of users.

Motivated by this observation, we study the question how much is the attention of a user worth. We approach this problem by tapping into the bid suggestion function extensively. We present the first comprehensive analysis of suggested bids in OSNs with the following contributions. (1) We build a crawler and have run it for more than 100 days. As a result, we harvest a large dataset consisting of 188,260 suggested bids over 449 distinct targeting conditions of 8 attributes from LinkedIn. We will make the dataset available to the research community. (2) We present detailed analyses of suggested bids from LinkedIn. We analyze the temporal and spatial properties of the suggested bids, and investigate the suggested bid distributions over a variety of user attributes related to career. We discover many consistent results, including: (suggested) bids are generally stable over time; they negatively correlate with GDP and income; the bids for users from different industries vary a lot; the bids for users with a specific skill or job are correlated with the supply and demand of the labor markets; the user working for a larger company is set with a higher suggested bid.

2. BID SUGGESTION

In the advertising systems of LinkedIn, once an advertiser \(d\) creates an ad \(a\) with a targeting condition \(c\) at time \(t\), a suggested bid \(b\) and the number \(n\) of audience who qualifies the condition \(c\) are provided. The bid suggestion is formulated as the function \(BS\) in Eq (1).

\[
BS : (c, v, a, t, ...) \rightarrow (b, n)
\]  

Each targeting condition \(c\) defines a unique segment of audience (a.k.a. users), and \(b\) is the bid the OSN suggests the advertiser to place in auctions to win a unit of interactions (impression, click or conversion) from any audience satisfying the condition \(c\) \([1, 2]\). Each targeting condition is defined as a pair, the targeting attribute and the value of the targeting attribute. For example, in the condition \(\langle\text{Location} : \text{CA}\rangle\) which means that the advertiser wants to target users in California, Location is the attribute and CA is the value. In

\(^1\)https://adwords.google.com/
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Domain</th>
<th>Explanation or Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>51</td>
<td>50 states and D.C.</td>
</tr>
<tr>
<td>Industry</td>
<td>17</td>
<td>Agriculture, Medical</td>
</tr>
<tr>
<td>Skill</td>
<td>39</td>
<td>Programming, Cooking</td>
</tr>
<tr>
<td>Company Size</td>
<td>9</td>
<td>Number of employees</td>
</tr>
<tr>
<td>Company Name</td>
<td>171</td>
<td>Top US companies</td>
</tr>
<tr>
<td>Job Title</td>
<td>126</td>
<td>Sampled Job positions</td>
</tr>
<tr>
<td>Job Seniority</td>
<td>10</td>
<td>CXO, Director, Entry-level</td>
</tr>
<tr>
<td>Job Function</td>
<td>26</td>
<td>Sales, Support, Research</td>
</tr>
</tbody>
</table>

Table 1: Crawled LinkedIn User Attributes

practice, advertisers can target users by a set of conditions, however, in this paper we only consider one condition.

3. CRAWLERS AND DATASET

To collect the input and output of the mapping function $BS$ in Eq (1), we created a crawler interacting with LinkedIn\(^2\) advertising system as: the crawler chooses an advertiser account $d$ and selects a targeting condition $c$, and sends them to the advertising system which returns the query result $(b, n)$ to the advertiser. We ran the crawler with 449 targeting conditions for 104 days, from Nov 2014 to Feb 2015. Since the crawler does not create or run any real campaign, the advertising history of $d$ was unchanged during the entire crawling period. Besides, the crawling process neither spends money nor participates in real auctions, thus our crawler has no influence to the real market. During the crawling periods, we created and used new advertiser accounts without any advertising history or social interaction.

3.1 Crawling LinkedIn Bid Suggestion

Since LinkedIn uniquely has the accurate career information of users, we focus on targeting conditions related to users' career. In total, we consider 8 common attributes and 449 of their frequent values, in Table 1.

We run the crawler every 6 hours, and each time we send 449 queries. For each query, we select only one distinct condition to construct $c$, e.g. $c=(\text{Skill}: C++)$.

Figure 1: An example of the bid suggestion provided by LinkedIn. (a) suggested bid interval and (b) audience size.

Figure 2: CPM temporal distribution

4. CHARACTERISTICS OF SUGGESTED BIDS

In this section, we explore the temporal and spatial properties of suggested bids, and the distribution of suggested bids related to career attributes.

The type of the suggested bid we consider is CPM (Cost per mille, i.e. the unit price to show one ad to 1,000 specified users). Therefore, when we mention a suggested bid, it actually means the (suggested) CPM. For a LinkedIn CPM which is a pair ($\text{CPM}_{\text{min}}$, $\text{CPM}_{\text{max}}$), we use $\text{CPM}_{\text{min}}$ because it reflects the least price to target a user. Except the CPMs used for temporal analysis in Section 4.1, for the rest of this section, when we refer the CPM of a targeting condition, it means the average CPM of that condition over our entire crawling period. The price unit in this paper is U.S.D., and the width of the bars in Fig 4 and 5 indicate normalized audience size.

4.1 Temporal Distribution

First, we explore the temporal patterns of CPMs from LinkedIn which are shown in Fig 2. The CPM is for any user in the U.S.. We find a few interesting properties in the time series. First, we do not observe any periodical patterns, e.g. weekly or daily, which we expected. Although there is no document recording how LinkedIn generates the suggested bids, we conjecture that LinkedIn uses some models to produce them. One possible way is to smooth historical bids as suggested bids (which is similar to the way that Facebook generates suggested bids [5]).

We observe intensive local oscillations in 3 short periods, e.g. around 2015-01-22. Interestingly, during each of the 3 periods, the CPM switches repeatedly between only two numbers. Finally, around holidays, e.g. Thanksgiving Day and the Christmas, the CPMs are higher than other periods (before the Thanksgiving and after New Year’s Eve) as we expected. We conjecture that during those holidays, the advertisers may run more campaigns than usual, thus CPMs are increased. Overall, we find that the CPMs are generally stable.

\(^2\)https://www.linkedin.com/ads/

\(^3\)http://www.bls.gov/
4.2 Spatial Distribution

In this part, we investigate how the user attribute, location, influences users’ CPMs. We present the CPMs of 50 states in the U.S. as a heat map in Fig 3 where a darker color indicates a higher CPM. Through hypothesis testing with 8 common distributions, e.g. Lognormal [3], Uniform and Poisson, we find that the suggested CPMs of the 50 states fit a Gaussian distribution $N(3.45, 0.47^2)$ with p-value as 0.46. We also compare the CPM with GDP per Capita data released by BEA for each state, and get a negative correlation as $-0.31$ (p-value as 0.02). This finding does not follow our expectation, therefore we re-examine the correlation between the GDP per Capita and the suggested CPMs of the 50 states fit a Gaussian distribution and Poisson, we find that the suggested CPMs of the 50 states fit a Gaussian distribution $N(3.45, 0.47^2)$ with p-value as 0.46. We also compare the CPM with GDP per Capita data released by BEA for each state, and get a negative correlation as $-0.31$ (p-value as 0.02). This finding does not follow our expectation, therefore we re-examine the correlation between the GDP per Capita and the suggested CPMs of the 50 states.

![Figure 3: CPM spatial distribution](image)

4.3 Career

We next explore CPMs of user career attributes from the three categories, job, skill and company.

4.3.1 Job

LinkedIn provides targeting options on user job attributes, including seniority, industry and specific title.

**Seniority.** First we analyze the distribution of CPMs over different levels of job seniority. We collect all the 10 levels of job seniority from LinkedIn targeting language, and the results are shown in Fig 4 where data points are sorted by their CPMs and the width indicates the normalized audience size. We identify that the CPMs of management levels (CXO, Manager, Director and VP) are all higher than the CPMs of non-management levels (Senior, Entry and Training). Moreover, among non-management levels, their CPMs directly correlate with the seniority as Senior($3.16) >$ Entry($3.07) >$ Training($2.91). However, we also observe an outlier, Unpaid, whose CPM is the highest. One potential reason is that unpaid jobs are usually temporal and transitional, e.g. most internships and voluntary jobs, thus the users with unpaid jobs may actively look for other jobs. As we observe that a large proportion of ads in LinkedIn are about job positions, advertisers may compete intensively for such users, which leads to a high CPM.

**Industry.** LinkedIn provides a list of 17 high-level industries, and one can target users according to the industry in which they work. We present the CPM distribution over these 17 industries in Fig 5. We identify that the Medical industry has the highest CPM while the High Tech and Corporate industries have the lowest CPMs. For each industry we compare its CPM with GDP per Capita from BEA. Since the taxonomy of industries in BEA is not the exactly same as that in LinkedIn, we match them with our best knowledge.

![Figure 4: CPM distribution over job seniority levels](image)

![Figure 5: CPM distribution over industries](image)

To avoid ambiguity, we exclude two industries, Service and Non-profit. As a result, for the other 15 industries, we observe a negative correlation as $-0.43$ (p-value 0.109)\(^6\), which is consistent with what we observed in Section 4.2, i.e. the negative correlation between the GDP per Capita and the CPMs of the 50 states.

**Specific Job.** We sample 130 jobs from BLS, and manually match them to the job title targeting options in LinkedIn. As a result, we get 126 matched job titles, and their CPMs. Due to the limited space, we sort all 126 job titles in the non-ascending order of CPMs, and present the top 10, bottom 10 and 24 in-between (randomly chosen) job titles in Fig 6. As expected, we observe high consistency between the CPMs of industries in Fig 5 and the CPMs of job titles. For example, the industry with the highest CPM is Medical in Fig 5, and the top 10 job titles with the highest CPMs are all related to Medical. Similarly, the industry with the lowest CPM is Corporate, and the related job, Recruiter, lies on the bottom in Fig 6 too (with the lowest CPM). To statistically explore whether the CPMs of industry are consistent with the CPMs of jobs, we match all the 126 jobs, according to the taxonomy provided by BLS, to the 17 industries with our best knowledge, and find a considerable correlation $0.555$ (p-value $< 10^{-5}$) between the CPM of a job and the CPM of the industry to which the job belongs. We test whether the demand/supply ratio of the job in labor market correlates with the CPM of the job. For each job title, we use its audience size to approximate its supply in the job market, and crawl the number of its openings from LinkedIn job search\(^7\) as the approximation of its demand. We define the demand/supply ratio of a job as the demand divided by the supply. By calculating the Pearson correlation coefficient between the ratios and the CPMs of all the jobs, we observe a weak correlation $0.2$ (p-value 0.049).

\(^6\)One major reason why the p-value is larger than 0.05 is that we only have 15 data points for the two lists.

\(^7\)https://www.linkedin.com/vsearch/j
4.3.2 Skill

LinkedIn allows a user to add skills, e.g., C++, to her profile, and advertisers can target users by their skills. We explore whether a subset of skills are favored by advertisers, i.e., with higher CPMs, which might indicate these skills are more needed in the job marketing. Skills are in the form of free text created by users, we manually compose a list of 39 common skills related to computer science, because we want to limit the impact caused by other factors, such as industry or job seniority, on CPMs. The results are shown in Fig 7. We also find a moderate negative correlation -0.37 (p-value < 0.03) between the CPM and the audience size of skills, which supports the conjecture that a skill acquired by less people has a higher CPM.

4.3.3 Company

We examine the CPM distributions over the companies in which users work, including two attributes, Company Size and Company Name.

Company Name. Considering the large volume of existing companies, we present the analysis on representative companies only. Among the top 500 companies with the largest market capitalization in the world, we get all 183 companies based in the US. Then we match them to the companies listed in LinkedIn. As a result, we get 171 exactly matched companies and their CPMs. Due to the space limitation, we select 20 representative companies, including 8 famous IT companies and 12 others with the largest market capitalization, and present their CPMs in Fig 8. We find that the top 3 companies with the largest CPMs, Walmart, Costco, and Lowe’s, are all giant retailers. By contrast, all the 8 IT companies have lower CPMs than the other 12. We also observe a positive correlation, 0.34 (p-value < 10^{-5}) between CPM and audience size for all these companies, which is consistent with what we observed in the analysis of company size (larger companies have larger CPMs). To explore whether the CPMs of industry are consistent with the CPMs of companies, we map these companies to the 17 industries in Fig 5. To balance the number of companies for each industry, we finally randomly choose 112 companies. The correlation between the CPM of a company and the CPM of the industry to which the company belongs is 0.314 (p-value < 10^{-3}). This finding is consistent with what we observed in the analysis of specific job titles, and again shows the considerable impact of industries on CPMs.

5. DISCUSSION

We also crawled suggested bids from Facebook, and find that the suggested bids from these two OSNs have a moderate positive correlation. Besides, from Facebook suggested bids, we observed that users with high or low income have higher CPMs than users with median income. To find out the reason of the bias of CPMs, it is interesting to study this open question: what ads are shown to OSN users with what attributes?

Assuming that the suggested bid is the actual cost to target a qualifying user, we study how advertisers can use the bids strategically. As a future work, we formulate the fractional cheap targeting problem. We show through data analysis that targeting subsets of users is a viable approach, and then we propose a greedy algorithm to help advertisers reach (up to) 40% more desired audience.

The cheap targeting strategy takes advantage of the arbitrage among the costs to target different user segments. Although it benefits some advertisers, it hurts the OSN’s revenue. Therefore, our another future work includes devising a pricing which eliminates any potential arbitrage for the OSN.

6. REFERENCES