March 2022

Generalized Mixture Overlap Model for Estimating Reach Surfaces

Matthew Clegg

Jiayu Peng

Generalized Mixture Overlap Model for Estimating Reach Surfaces

Authors: Matthew Clegg & Jiayu Peng

Introduction

Advertisers generally wish to know how many unique people can be reached by an advertisement, and, if additional money were spent to distribute an advertisement, how many additional unique people would be reached. Mathematically, such data can be represented by a reach curve, which estimates the number of unique people that see an advertisement as a function of the number of advertising impressions that are purchased on an advertising network.

To maximize a number of people reached by an advertisement, a company may wish to run the advertisement across multiple publishers. However, it can be challenging to estimate, based on individual reach curves corresponding to respective individual publishers, how many people will be reached by advertisements purchased on multiple publishers.

Accordingly, this paper presents a method for using a generalized mixture overlap model to construct a reach surface that enables reach estimates as a function of spend across more than one publisher.

Reach Surfaces Overview

A reach surface is a mathematical model that estimates the total number of unique people that will be reached by advertising across a collection of publishers. After construction, a reach surface model takes as input the amount spent with each publisher. As output, the reach surface model produces the total number of unique people that are reached. This paper proposes a method for constructing reach surface models based on reach curves of individual publishers along with a limited number of additional data points representing combined spend across publishers.

Using a reach surface, an advertiser can estimate the number of people that could potentially be reached under various different spending scenarios (e.g., for different variations of money distribution across the publishers). To construct a reach surface, the advertiser should receive a reach curve from each publisher with whom the advertiser contemplates advertising (or calculate the reach curve), where a reach curve is a function that estimates the number of unique people that may be reached on a particular publisher's platform alone. In addition, the advertiser should access a number of additional data points that measure the total reach across publishers for various different spend scenarios. The reach surface can be used to estimate the total number of unique people that would be reached for any arbitrary allocation of spend to different publishers.

Advantageously, if the reach surface is constructed using the generalized mixture overlap model of this paper, the reach curves submitted by the individual publishers can be arbitrary. Further, by increasing the number of additional data points given as input, the generalized mixture overlap model can fit increasingly complex reach surfaces.

<u>Reach Surface Model Properties</u>

An ideal reach surface model should be able to provide estimates of the total unique reach for any allocation of impressions to publishers. Thus, let $(i_1, i_2, \ldots, i_p)$ be a vector of impression allocations to p publishers. The reach evaluation function $r(i_1, i_2, \ldots, i_p)$ associates the total unique reach to this vector. This defines a surface in $(p+1)$-dimensional space. This paper proposes a method for approximating this surface.

A reach surface model is constructed in a two-step process. First, a single-publisher reach function is computed for each publisher (or can be provided by each publisher). Thus, let $r_j(i_j)$, $j = 1, \ldots, p$, be the number of unique people reached by publisher j when i_j impressions are purchased on that publisher's platform. Second, the single-publisher reach functions are combined with additional data to obtain the reach surface model $\hat{r}(i_1, i_2, \ldots, i_p)$. This paper focuses on this second step.

An ideal reach surface model should satisfy the following properties:

1. **Single-publisher agreement.** If $(i_1, i_2, \ldots, i_p)$ is an impression vector that is zero in all coordinates except one, say coordinate k, then $\hat{r}\left(i_1, i_2, \ldots, i_p\right) = r_k(i_k)$.
2. **Monotonic**. Let $(i_1, i_2, \ldots, i_p)$ and $(i'_1, i'_2, \ldots, i'_p)$ be impression vectors that satisfy $i_j \leq i'_j$ for all $j = 1, 2, \ldots, p$. Then, it should be the case that $\hat{r}\left(i_1, i_2, \ldots, i_p\right) \leq \hat{r}\left(i'_1, i'_2, \ldots, i'_p\right)$.
3. **Subadditive**. For any impression vector $\left(i_1, i_2, \ldots, i_p\right)$, it should be true that $\hat{r}\left(i_1, i_2, \ldots, i_p\right) \leq r_1(i_1) + r_2(i_2) + \cdots + r_p\left(i_p\right)$.
4. **Matches additional data.** The reach surface should closely agree with any additional data points that are provided as input.

Thus, the method presented herein constructs a reach surface model that satisfies the above four properties. As input, the reach surface model requires single publisher reach functions, each of which are assumed to be monotonic. In addition, the reach surface model requires as input a collection of additional data points describing points on the reach surface. For example, these points could be the total reach for various pairwise unions of the input individual publishing campaigns. Specifically, the generalized mixture overlap model presented here involves estimation of $p(c-1)+1$ parameters, where p is the number of publishers and $c \geq 1$ is a configurable integer parameter, as will be further described below.

Generalized Mixture Overlap Model for Estimating Reach Surfaces

Let the single-publisher reach functions be $r_1(x), r_2(x), \ldots, r_p(x)$. We assume that $r_j(0) = 0$ for each of these functions, and that each of these functions are monotonic, i.e., that $r_j(x) \leq r_j(x+c)$ for any $c \geq 0$.

For single-publisher reach function $r_j(x)$, let m_j be the maximum possible reach achievable with that function. This can either be given as an additional input or be computed by evaluating the reach function with a large input value.

The model computes a $p \times c$ matrix $A = \{a_{i,j}\}$ that has the following properties: (1) all entries in the matrix are non-negative, and (2) each row sums to 1. Further, one additional parameter N is computed, where N is a non-negative value that is at least as large as the maximum reach of any of the single-publisher reach functions, i.e., $N \geq m_j$ for $i = 1, 2, \ldots, p$.

Suppose we have such a matrix A. Then, we can define a reach surface as follows:

$$\hat{r}(x_1, x_2, \ldots, x_p) = N \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - \frac{a_{i,j}}{N} r_i(x_i) \right) \right] \tag{1}$$

While it may be tempting to replace both instances of N in equation (1) with $\frac{N}{c}$, in order to allow greater concentration of viewers reached by a particular publisher in a single component, such a formulation is problematic. If $\frac{c}{N} r_i(x_i) > 1$, then some terms in the product could be negative, which violates the probabilistic interpretation of equation (1) that is described in the next section.

Heuristic Justification of the Generalized Mixture Overlap Model

Before demonstrating that the generalized mixture overlap model given by equation (1) satisfies the first three properties above, we provide a justification for why equation (1) is an appropriate model for representing a reach surface.

Let U be the total universe of reachable viewers, and suppose that U has cardinality N. For publisher i, $i = 1, 2, \ldots, p$, let A_i be the set of people potentially reachable by that publisher. The cardinality of A_i is m_j and $U = A_1 \cup A_2 \cup \ldots \cup A_p$.

The reach function $r_i(x)$ can be thought of as representing the expected number of people in A_i that will be reached with a purchase of x impressions. Thus, we can define a function $f_i(x) = r_i(x)/m_i$, This function represents the probability Pr that a given viewer in A_i will be reached given that x impressions are shown by publisher i:

$$f_i(x) = \Pr(u \text{ reached by publisher } i |\, u \in A_i) = r_i(x)/m_i.$$

We now imagine that the universe U is partitioned into c disjoint sets $C_1, C_2, \ldots, C_c$ having the following conditional independence (CI) properties:

$$\Pr\left(u \text{ reached by publisher } i \text{ and } u \in C_j \middle| u \in A_i\right)$$

$$= \Pr(u \text{ reached by publisher } i | u \in A_i) \Pr\left(u \in C_j \middle| u \in A_i\right) \quad \text{(CI1)}$$

$$\Pr\left(u \in A_i \cap A_{i'} \middle| u \in C_j\right) = \Pr\left(u \in A_i \middle| u \in C_j\right) \Pr\left(u \in A_{i'} \middle| u \in C_j\right) \quad \text{(CI2)}$$

For a given publisher i and partition C_j, let $a_{i,j}$ be defined as $a_{i,j} = \Pr\left(u \in C_j \middle| u \in A_i\right)$. Since every user u belongs to exactly one C_j, it follows that $\sum_j a_{i,j} = 1$. Thus, for a randomly chosen user u:

$$\Pr(u \text{ reached})$$

$$= \sum_{j=1}^{c} \Pr(u \text{ reached and } u \in C_j)$$

$$= \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} (1 - \Pr(u \text{ reached by publisher } i \text{ and } u \in C_j))\right]$$

$$= \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} (1 - \Pr(u \text{ reached by publisher } i \text{ and } u \in C_j | u \in A_i) \Pr(u \in A_i))\right]$$

$$= \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} (1 - \Pr(u \text{ reached by publisher } i \text{ and } u \in A_i) \Pr(\in C_j | u \in A_i) \Pr(\in A_i))\right]$$

$$= \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - a_{i,j} f_i(x_i) \left(\frac{m_i}{N}\right)\right)\right]$$

$$= \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - a_{i,j} \frac{r_i(x_i)}{m_i} \left(\frac{m_i}{N}\right)\right)\right]$$

$= \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right)\right].$

By taking expectations over the set U, we then obtain:

$$E[\#\{u \text{ reached}\}] = N \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right)\right].$$

Note that the right-hand side of the above equation is identical to the reach surface model given in equation (1). Thus, the reach surface model given in equation (1) represents the expected number of people that would be reached if the universe could be divided into c disjoint sets $C_1, C_2, \ldots, C_c$ having the properties (CI1) and (CI2). From this perspective, the parameter N in the equation (1) is an estimate of the size of the universe, and the parameter $a_{i,j}$ is an estimate of $\Pr(C_j | A_i)$.

We now turn to proving that the generalized mixture overlap model given by equation (1) satisfies the first three consistency properties of an ideal reach surface, (1) single-publisher agreement, (2) monotonicity, and (3) subadditivity.

Single-Publisher Agreement

Let $(x_1, x_2, \ldots, x_p)$ be an impression vector where all coordinates are zero except for some coordinate i. We must prove that $\hat{r}(x_1, x_2, \ldots, x_p) = r_i(x_i)$. Since equation (1) is symmetric in each of the coordinates, it suffices to prove the assertion for $i = 1$.

For $i > 1$, $r_i(x) = 0$, so it follows that $1 - \frac{a_{i,j}}{N} r_i(x_i) = 1$. Therefore, the expression on the right-hand side of equation (1) can be rewritten as:

$$\hat{r}(x_1, x_2, \ldots, x_p) = N \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right)\right]$$

$$= N \sum_{j=1}^{c} \left[1 - \left(1 - \frac{a_{1,j}}{N} r_1(x_1)\right)\right]$$

$$= N \sum_{j=1}^{c} \frac{a_{1,j}}{N} r_1(x_1)$$

$$= r_1(x_1) \sum_{j=1}^{c} a_{1,j}$$

$$= r_1(x_1)$$

The last step in the above proof is justified because each column in A sums to 1.

Monotonicity

Let $(x_1, x_2, \ldots, x_p)$ and $(x'_1, x'_2, \ldots, x'_p)$ be impression vectors such that $x_i \le x'_i$ for $i = 1, 2, \ldots, p$. We must prove that $\hat{r}(x_1, x_2, \ldots, x_p) \le \hat{r}(x'_1, x'_2, \ldots, x'_p)$. Suppose that this statement is known to be true if the two impression vectors differ only in a single coordinate, then:

$$\hat{r}(x_1, x_2, \ldots, x_p) \le \hat{r}(x'_1, x_2, x_3 \ldots, x_p)$$

$$\le \hat{r}(x'_1, x'_2, x_3, \ldots, x_p)$$

$$\le \hat{r}(x'_1, x'_2, x'_3 \ldots, x_p)$$

$$\le \cdots$$

$$\le \hat{r}(x'_1, x'_2, x'_3 \ldots, x'_p)$$

Thus, if we can prove that the statement is true when the impression vectors differ in only a single coordinate, it follows that the statement is true more generally. Moreover, by symmetry, it suffices to prove the statement when the two impression vectors differ only in the first coordinate. Hence, we assume that $x'_i = x_i$ if $i > 1$. To simplify the proof, let

$$u_j = \prod_{i=2}^{p} \left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right)$$

Note that because $r_i(x_i) \leq m_j \leq N$, each term in the above product is greater than or equal to zero, and therefore $u_j \geq 0$. The proof that $r\left(x'_1, x'_2, \dots, x'_p\right) - r\left(x_1, x_2, \dots, x_p\right) \geq 0$ proceeds as follows:

$$\hat{r}\left(x'_1, x'_2, \dots x'_p\right) - \hat{r}\left(x_1, x_2, \dots x_p\right)$$

$$= N\sum_{j=1}^{c}\left[1 - \prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x'_i)\right)\right] - N\sum_{j=1}^{c}\left[1 - \prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right)\right]$$

$$= N\sum_{j=1}^{c}\left\{\left[1 - \prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x'_i)\right)\right] - \left[1 - \prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right)\right]\right\}$$

$$= N\sum_{j=1}^{c}\left\{\prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right) - \prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x'_i)\right)\right\}$$

$$= N\sum_{j=1}^{c}\left\{\left(1 - \frac{a_{i,j}}{N} r_1(x_1)\right) u_j - \left(1 - \frac{a_{i,j}}{N} r_1(x'_1)\right) u_j\right\}$$

$$= N\sum_{j=1}^{c}\left\{\left(1 - \frac{a_{i,j}}{N} r_1(x_1)\right) - \left(1 - \frac{a_{i,j}}{N} r_1(x'_1)\right)\right\} u_j$$

$$= N\sum_{j=1}^{c}\left\{\frac{a_{i,j}}{N}\left(r_1(x'_1) - r_1(x_1)\right)\right\} u_j$$

$$\geq 0$$

The final inequality is justified because $r_1(x'_1) - r_1(x_1) \geq 0$ and the other quantities in the equation are also non-negative.

Subadditivity

To prove subadditivity, we have to show that $\hat{r}\left(x_1, x_2, \dots x_p\right) \leq r_1(x_1) + r_2(x) + \cdots + r_p\left(x_p\right)$. To prove this, we define $b_{i,j} = (a_{i,j}/N) r_i(x_i)$. Since $r_i(x_i) \leq m_j \leq N$ and since $0 \leq a_{i,j} \leq 1$, it follows that $0 \leq b_{i,j} \leq 1$. By induction on p, it follows that:

$$\prod_{i=1}^{p}\left(1 - \frac{a_{i,j}}{N} r_i(x_i)\right) = \prod_{i=1}^{p}\left(1 - b_{i,j}\right) \geq 1 - \sum_{i=1}^{p} b_{i,j}$$

Consequently, we have:

$$\hat{r}(x_1, x_2, \dots, x_p) = N \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} \left(1 - \frac{a_{i,j}}{N} r_i(x_i) \right) \right]$$

$$= N \sum_{j=1}^{c} \left[1 - \prod_{i=1}^{p} (1 - b_{i,j}) \right]$$

$$= N \sum_{j=1}^{c} \left[1 - \left(1 - \sum_{i=1}^{p} b_{i,j} \right) \right]$$

$$= N \sum_{i=1}^{p} \sum_{j=1}^{c} b_{i,j}$$

$$= N \sum_{i=1}^{p} \sum_{j=1}^{c} \frac{a_{i,j}}{N} r_i(x_i)$$

$$= \sum_{i=1}^{p} r_i(x_i) \sum_{j=1}^{c} a_{i,j}$$

$$= \sum_{i=1}^{p} r_i(x_i)$$

$$= r_1(x_1) + r_2(x_2) + \cdots + r_p(x_p)$$

Parameter Estimation

Given a list of reach functions $r_1(x)$, $r_2(x)$, ..., $r_p(x)$ and a collection of points $(x_{11}, x_{12}, \dots, x_{1p}, r_1)$, $(x_{21}, x_{22}, \dots, x_{2p}, r_2), \dots, (x_{n1}, x_{n2}, \dots, x_{np}, r_n)$, we would like to find estimates for the $a_{i,j}$ and N that most closely match the data. This can be accomplished using a multivariate minimizer such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

A complication is that the following constraints must be satisfied: (1) $a_{i,j} \geq 0$, (2) $\sum a_{i,j} = 1$, and (3) $N \geq \max_i \{m_i\}$. To satisfy these constraints, we optimize over a different set of parameters, $a'_{i,j}$ and N'. The $a'_{i,j}$ and N' are allowed to be arbitrary real numbers. These values are related to the original parameters via the following equations:

1. $s_i = \sum_j \exp(a'_{i,j})$
2. $a_{i,1} = 1/s_i$
3. $a_{i,j} = \frac{\exp(a'_{i,j})}{s_i}$ if $j > 1$
4. $N = \max_i\{m_i\} + \exp(N')$

Model Selection

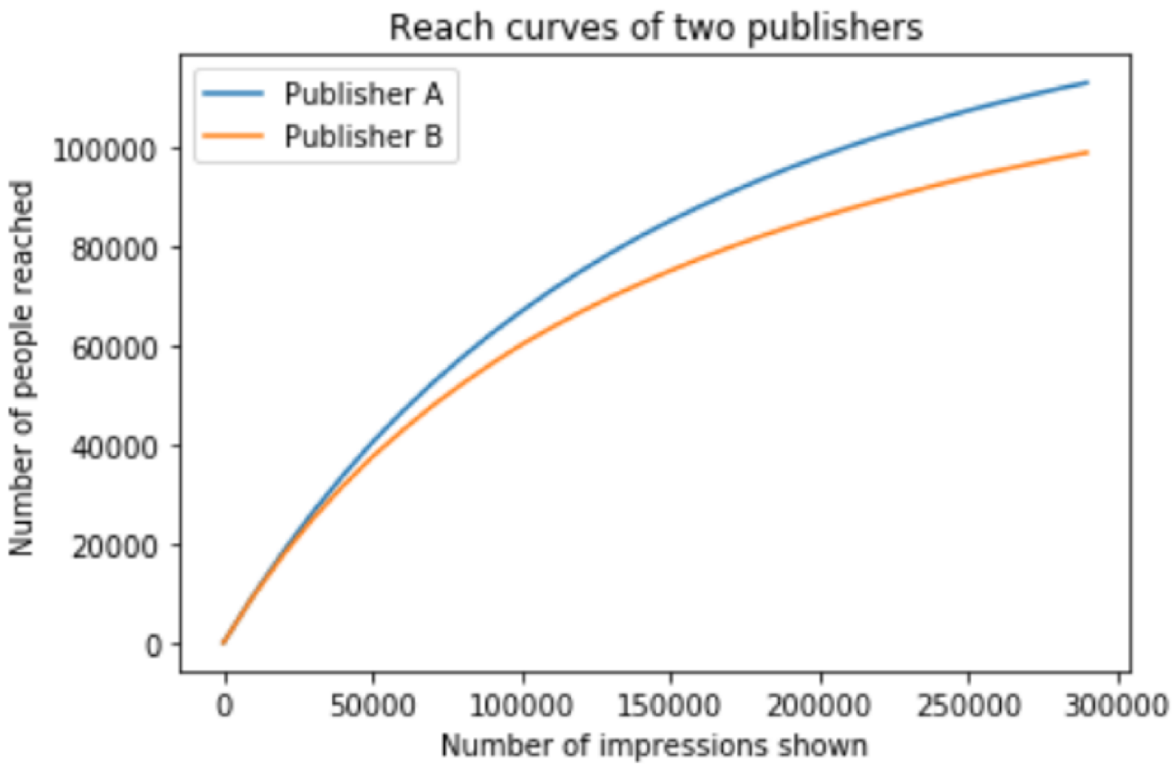
An important question is how best to select the number of partitions c. This is challenging because there is not a clear likelihood function. As a result, methods such as the likelihood ratio test, Akaike's Information Criterion (AIC) and Bayesian Information Criteria (BIC) do not appear to be usable in this scenario. In addition, data points are likely very expensive to obtain, making creating a holdback set difficult.

One option is to randomly create a holdback set of 20% of the data, and select the value of c that minimizes the squared error on the holdback set. Another option is to choose a large initial value of c (e.g., $c = p$), and use L_1 regularization to reduce the number of non-zero parameters to an acceptable value.

<u>Example Reach Surface</u>

This section discusses an example involving two publishers, publisher A and publisher B. Publisher A can reach 120,000 users and has a reach curve parameterized by a Gamma-Poisson distribution with shape 2 and scale 1. Publisher B can reach 100,000 users and has a reach curve parametrized by a Gamma-Poisson distribution with shape 1 and scale 2. The two publishers have 30,000 users in common, chosen uniformly at random. The reach curves of the two publishers is shown below in Figure 1.

Figure 1: Individual Reach Curves of Publishers A and B



In this two-publisher example, we set $c = 1$. For this case, $a_{i,j} = 1$, and equation (1) reduces to:

$$r(x_1, x_2) = N\left[1 - \left(1 - \frac{r_1(x_1)}{N}\right)\left(1 - \frac{r_2(x_2)}{N}\right)\right]$$

The parameter N represents the value of the total universe for which the two publishers most closely approximate being independent. This occurs when:

$$N = \frac{r_1(x_1)r_2(x_2)}{r_1(x_1) + r_2(x_2) - r(x_1, x_2)}$$

In this example, x_1 and x_2 were chosen to be 120,000 and 100,000, respectively. In this particular simulation, the number of people reached by Publisher A was 66,785, while the number of people reached by Publisher B was 60,058. The combined reach was 116,709. Substituting these values into the above formula, the value of N was determined to be about 395,793.

Figures 2A and 2B illustrate the true reach versus the modeled reach in two scenarios. In Figure 2A, the total number of impressions was varied from 0 to 300,000. At each point along the x-axis, the total reach was calculated when 50% of the impressions were allocated to Publisher A and 50% were allocated to Publisher B. As illustrated in Figure 2A, the modeled and actual reach are identical. Other allocations between the publishers were also simulated, and the results were identical in all cases. In Figure 2B, the total number of impressions served was held constant at 100,000, but the allocation to publisher A varied from 0 to 100,000. The total reach was plotted for both the true reach surface and the modeled reach surface, and these curves were identical, as shown.

Figure 2A

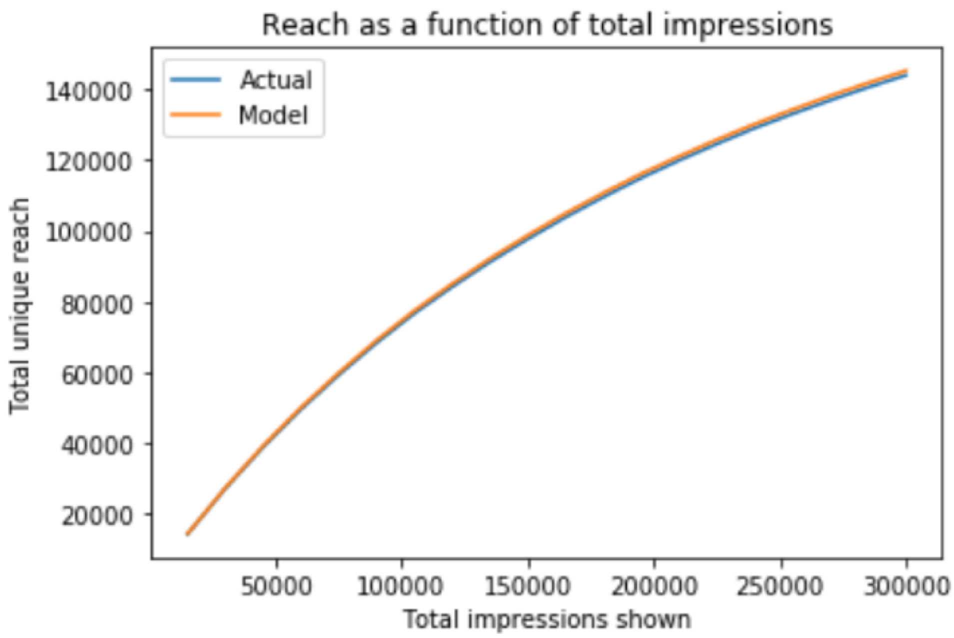
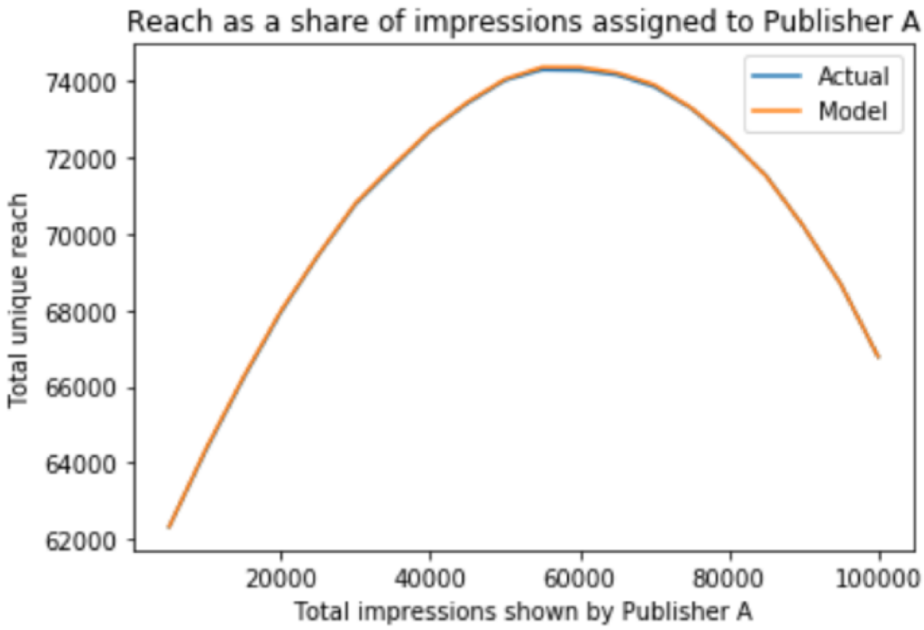


Figure 2B



Thus, at least in this example, the generalized mixture overlap model agrees with the true reach surface. Future work will explore the accuracy of the model for more than two publishers, and explore optimal methods for choosing the hyperparameter c.

Differential Privacy

Here, we explore whether the generalized mixture overlap model can be used to compute reach surfaces that utilize only locally differentially private methods, thereby avoiding the use of a secure multi-party computation (SMPC). Differential privacy refers to a mathematically-provable guarantee that the information of individual users is protected.

We assume that we have a per-publisher privacy budget of ϵ. Each publisher allocates half of its privacy budget to the calculation of a differentially private single-publisher reach curve, and allocates the other half to the calculation of a differentially private Vector of Counts.

To model a single-publisher reach curve, one approach is to specify the reach curve using three parameters: the total number of reachable users N, and Gamma distribution parameters α and β specifying the distribution of impressions to users via a Gamma-Poisson mixture. These three parameters can be fit using an exponential mechanism approach, which is described in another paper by this author, "Using the Exponential Mechanism to Compute Differentially Private Reach Curves." (Pending publication).

The Vector of Counts is a differentially private data structure that can be used to compute intersections, and is discussed in a paper by Peng, et al.[1]

A central server can collect the differentially private reach curve and Vector of Counts from each publisher. The central server can use the Vectors of Counts to compute the sizes of the pairwise unions of the publisher campaigns. The pairwise unions and the single publisher reach curves are then used as input to the generalized mixture overlap model, from which a reach surface for the publishers is computed. The reach surface can then be used to estimate the total combined reach of all publishers and, accordingly, to determine optimal spend allocations for various spend values across the publishers.

Example Computing System

Figure 3 (see below) illustrates an example computing system 100 which can implement the techniques of this paper. It should be understood that the computing system 100 is an example computing system, and that other systems capable of performing the techniques of this paper may be envisioned.

The example computing system 100 includes a client computing device 102 (also referred to herein as "client device 102") coupled to a network 120. The client device 102 may be a computing device such as a smartphone, laptop computer, or desktop computer. The client device 102 may include a memory 106, one or more processors (CPUs) 104, a network interface 114, a user interface 116, and an input/output (I/O) interface 118. The

[1] Peng, Jiayu, et al. "Privacy-centric Cross-publisher Reach and Frequency Estimation Via Vector of Counts." (2020). Last accessed February 28, 2022 at https://research.google/pubs/pub50153.

client device 102 may also include components not shown in Fig. 3 such as a graphics processing unit (GPU).

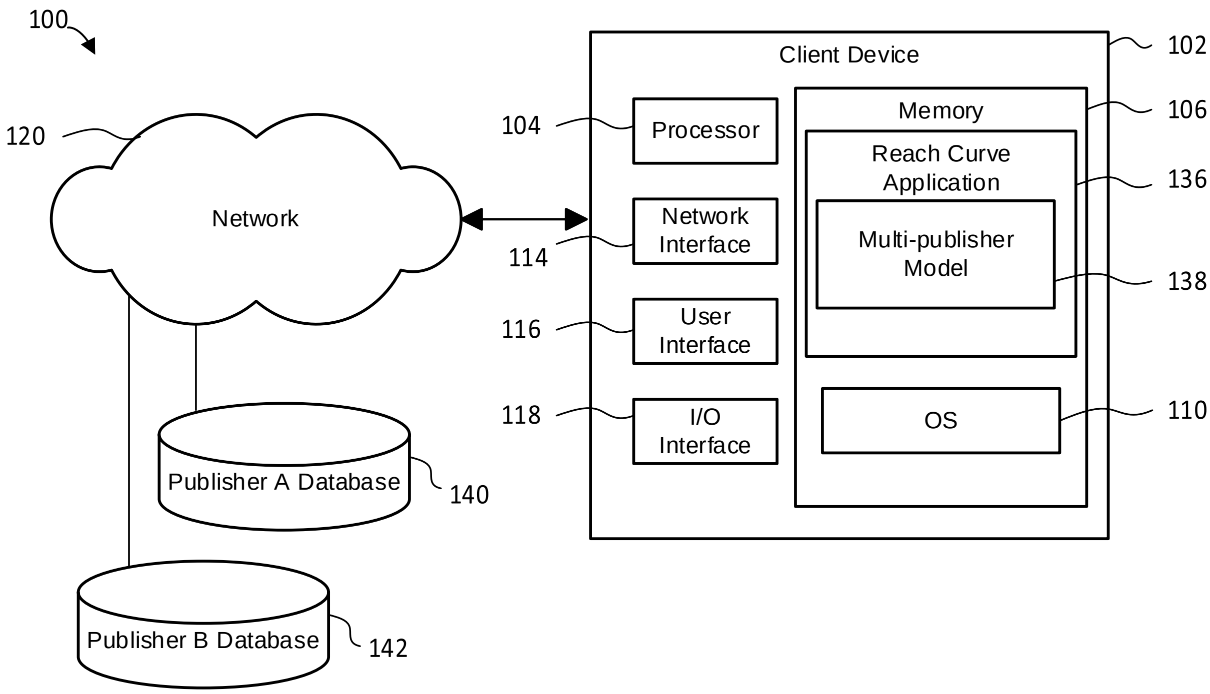
The network interface 114 may include one or more communication interfaces such as for enabling communications via the network 120. The user interface 116 may be configured to provide information, such as reach curves and/or reach surfaces, to the user. The I/O interface 118 may include various I/O components (e.g., ports, capacitive or resistive touch sensitive input panels, keys, buttons, lights, LEDs). For example, the I/O interface 118 may be a touch screen, or may include input and output devices such as a keyboard, mouse, and display screen. A user can generate, modify, and/or implement models for generating reach curves and/or reach surfaces by interacting with the client device 102 via the I/O interface 118.

The memory 106 may store machine-readable instructions executable on the one or more processors 104 and/or special processing units of the client device 102. The memory 106 also stores an operating system (OS) 110, which can be any suitable mobile or general-purpose OS. In addition, the memory can store one or more applications that communicate data via the network 120, including a reach curve application 136.

Depending on the implementation, the reach curve application 136 may display reach curves and/or reach surfaces, data from publishers such as reach, numbers of impressions purchased, amount of spend, request and receive generated reach curves and/or reach surfaces, display instructions (e.g., code) for implementing a reach curve generation model or reach surface generation model, provide user-controls for interacting with reach curve-related (and/or reach surface-related) data and/or models, etc. In particular, the reach curve application 136 may use a multi-publisher model 138, which may be the generalized mixture overlap model described above. Although Fig. 3 illustrates the reach curve application 136 as a standalone application, the functionality of the reach curve application 136 also can be provided in the form of an online service accessible via a web browser executing on the client device 102, as a plug-in or extension for another software application executing on the client device 102, etc. Further, the reach curve application 136 and/or the multi-publisher model 138 may be stored on a different computing device, such as a server, that the client device 102 can communicate with via the network 120.

The client device 102 may be able to access publisher data via the network (e.g., Publisher A data stored in a Publisher A database 140, Publisher B data stored in a Publisher B database 142, etc.) and/or may store publisher data on the memory 106. For example, publisher data may indicate, for different numbers of purchased impressions, how many unique people were reached, and may include reach curves for the individual publisher..

Figure 3: Example Computing System



ABSTRACT

A reach curve estimates the number of unique people that see an advertisement as a function of the number of advertising impressions that are purchased on an advertising network. A generalized mixture overlap model can be used to compute a reach surface based on the reach curves of individual publishers. Such a reach surface can then be used to estimate the total number of unique people that can be reached for any arbitrary allocation of spend across multiple publishers.