iPinYou DSP Bidding Log Format

— Where Computational Advertising Meets BigData

1. Data Set

iPinYou prepares a set of iPinYou DSP bidding, impression, click, and conversion logs from some advertising campaigns for the training and validation purpose. The training and validation set is released to participating teams.

For the log format and the sample log, see Section 2. Here is some description of the data in the log and the processing that iPinYou has done for a variety of purposes.

For each line of the bidding log, the first column is bid ID, which uniquely identifies one ad impression opportunity; in each line of the impression, click or conversion log, the bid ID is same as that in the corresponding bidding log. IPinYou applies a hash function to bid ID before the log is released.

1.1 Bidding Log

For each line of the bidding log, it contains information about the ad slot, the user, and bidding .

Timestamp column is when the bid request arrives at the DSP server.

Ipinyou ID is the user cookie set by iPinYou. iPinYou applies a hash function to the cookie ID before the log is released for user privacy protection. User Agent is the user agent of browser through which the user visit the publisher's website.

IP address is the first three bytes of user IP address. iPinYou removes the last byte for user privacy protection before the log is released. Region ID is the region (a.k.a., province or state) ID where the user visits. CityID is the city where the user visits.

Ad exchange column represents which ad exchange this ad impression comes from. Currently, the value can be adx (Google DoubleClick AdX), tanx (Alibaba), and tencent (Tencent).

Domain is the name of domain where the ad impression will show up. iPinYou applies a hash function to domain before the log is released. URL is where the ad impression will show up. The publisher may ask the ad exchange not to pass this information to DSPs. iPinYou applies a hash function to URL before the log is released.   Anonymous URL ID is the domain page ID set by the

ad exchange when URL is not available to DSPs, mentioned previously. Only one of URL and Anonymous URL ID is meaningful.

Ad Slot ID is the unique ID to show the location where ad impression will show up on the web page. One web page can have multiple ad slots. Ad slot width and Ad slot height show the width and height of the ad slot respectively. Both columns can be an array, which means the publisher allows multiple widths and heights of the ad creative.

Ad slot visibility column represent whether the ad slot is above the folder (1), below the folder (2), or unknown (0). Ad slot format is either fixed ad slot (1) or the popup window (2).

Ad slot floor price is the lowest price at which price the publisher allows DSPs to win the ad slot. If no DSP bid higher price than the floor price, this auction has no winning DSP. iPinYou do a linear scale of bidding price before the log is released. The price unit is RMB Yuan per CPM.

Creative ID is the ad creative of the advertiser for which the DSP bid on the exchange.

Bidding price is what the DSP bids on the exchange. iPinYou does a linear scale of bidding price before the log is released. The price unit is RMB Yuan per CPM.

1.2 Impression, Click, and Conversion Log

For the impression and click log, most columns are the same as those of bidding log. Here is the description of additional columns.

Timestamp is the time when the event (impression, click, or conversion) takes place.

Log type is 1 (impression), 2(click), or 3(conversion).

Paying price is what the winning DSP actually pays on the exchange. iPinYou does a linear scale of bidding price before the log is released. The price unit is RMB Yuan per CPM.

Key page URL is the URL the user visits if the user clicks the ad creative for the log type is 2, or the URL where the user makes the conversion for the log type is 3. iPinYou applies a hash function to the URL before the log is released. This column is not applicable for the impression log.

1.3 Region and City File

iPinYou also provides two separate text files. One is the mapping file from the region id to region name, and the other is the mapping file from the city id to the city name. All region and city values are currently geo locations from the mainland China. Each line of the region file is a pair of the form (region id, region name); each line of the city file is a pair of the form (city id, city

name).

2. Log Format and Sample Data

Columns with (*) means that data in the column is modified before the log is released.

2.1 Bidding log

Column	Example
Bid ID(*)	c0550000008e5a94ac18823d6f275121
Timestamp	20130218134701883
iPinyou ID(*)	dF_5qwD1UDI
User-Agent	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/535.11 (KHTML, like Gecko) Chrome/17.0.963.84 Safari/535.11 SE 2.X MetaSr 1.0
IP(*)	119.163.222.*
Region	146
City	147
Ad Exchange	Adx
Domain(*)	e80f4ec7f5bfbc9ca416a8c01cd1a049
URL(*)	hz55b000008e5a94ac18823d6f275121
Anonymous URL ID	Null
Ad slot ID	mm_10982364_973726_9023493
Ad slot width	300
Ad slot height	250
Ad slot visibility	1
Ad slot format	1
Ad slot floor price (RMB/CPM) (*)	0
Creative ID	f80f4ec7f5bfbc9ca416a8c01cd1a049
Bidding price (RMB/CPM) (*)	573

2.2 Impression, Click, and Conversion Log

Column	Example
Bid ID (*)	01530000008a77e7ac18823f5a4f5121
Timestamp	20130218001203638
Log type	1
iPinyou ID(*)	35605620124122340227135
User-Agent	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)
IP(*)	118.81.189.*
Region	15
City	16
Ad exchange	Adx
Domain(*)	e80f4ec7f5bfbc9ca416a8c01cd1a049
URL(*)	hz55b000008e5a94ac18823d6f275121
Anonymous URL ID	Null
Ad slot ID	mm_15191080_2147689_8764813
Ad slot width	300
Ad slot height	250
Ad slot visibility	2
Ad slot format	1
Ad slot floor price (RMB/CPM) (*)	0
Creative ID	e39e178ffdf366606f8cab791ee56bcd
Bidding price (RMB/CPM) (*)	753
Paying price (RMB/CPM) (*)	15
Key page URL (*)	a8be178ffdf366606f8cab791ee56bcd

If you have any question about the data, feel free to send it to dsp-competition@ipinyou.com