

Data Science Challenges for Online Advertising

*A Survey on Methods and Applications from a
Machine Learning Perspective*

Matina Thomaidou, PhD

IWD2016

Dublin, March 2016

■ Central problem of computational advertising

- Find the "best match" between a given user in a given context and a suitable advertisement
 - Context: e.g. a user entering a query in a search engine ("sponsored search"), a user reading a web page ("content match" and "display ads"), a user watching a movie on a portable device, etc.
 - Constraints: e.g. limited budget of the advertiser on a specific period
 - Advertising is a form of information – IR problem

■ Central Challenges

1. Design markets and exchanges that help in this task, and maximize value for users, advertisers (!), and publishers
2. Build the infrastructure to support this process

Participants of the Sponsored Search Advertising

Auctioneer (e.g., Google)

- The auctioneer organizes the ad auction process

Advertisers

- They want to promote a product or service
- Each advertiser wants to have better performance than the others



Publishers (in SSA case, equivalent with the **auctioneer**)

- They give some space for ads (pictures or text)
- In SSA, slots in the search engine results page (SERP)

Users

- The advertisers want to attract their attention
- The users may search on a search engine



Outcome of an Ad Auction

Query / Keywords:
initiating the ad auction

Google summer schools

Web Images More Search tools

About 733,000,000 results (0.29 seconds)

Ads related to summer schools

Summer School 2013 - sbcOxford.co.uk
www.sbcOxford.co.uk/SummerSchool2013
Oxford Summer School Focused on Fun & Learning. Book Now for 2013.
SBC Bournemouth 11 - 15 - Abberley Hall 8 - 12 - SBC Oxford 10 - 14

Summer School Oxford (UK) 0345 026 3133
www.oxfordimmerse.com/Summer-Schools
2 Week Academic Summer School. For 16 To 18 Year Olds. Only £3295!
Apply Now. Only £3295! - List Of Subjects That You Can Study

Summer Student Programs - SummerStudy.com
www.summerstudy.com/
Exciting Pre-College Enrichment College Credit, Enrichment Classes

Related searches: summer schools in europe university summer schools

Harvard Summer School | Summer Programs | College & High...
www.summer.harvard.edu/ - Cached
Includes an on-line course catalogue, downloadable application forms, and information about English language programs

Summer Schools in Europe - Directory of summer courses
www.summerschoolsineurope.eu/ - Cached
Summer Schools in Europe eu, directory of summer courses. Find a summer course in 3 steps ... Find summer courses by keyword ...
Courses - Summer Schools in Europe - Contact us - Submit courses

Summer school - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Summer_school - Cached

Algorithmic / Organic Results

Ads

International Summer Camp
www.villagecamps.com/
Switzerland, UK, France, Austria, Languages, Sports, Arts, Leadership

Summer English School UK
www.xkeys.co.uk/english-school/
Learn & improve your English
The best Summer school in England

Course
www.openuniversity.edu/
Visit the Open University™ Online & Register For 2013 Courses Today!

Summer School Brussels
www.ies.be/summerschool
European Policymaking
EU institutions, Brussels & Vienna

Military Summer Camp
www.militaryadventurecamp.com/
A fun and realistic Basic Training environment for teens 13-18.

Summer School 2013
www.fondazioneambiente.org/iss
Turin School of Local Regulation
Torino, 9-20 Sept. 2013. Apply now!

Sponsored Results

- Ads
- Textual Advertising: Ad-text (ad creative or ad copy)

Landing Page

- Click the ad
- Make an "action" (conversion) inside the page

Search Engine
Result Page
(SERP)

Business Metrics

■ Impression

- The appearance of an advertisement in a SERP after a user's query

■ Click-Through Rate (CTR)

- The percentage of people clicking on an advertisement when it appears in a SERP
- $CTR = \text{Clicks} / \text{Impressions}$

■ Conversion Rate (CR)

- The percentage of conversions against clicks
- $CR = \text{Conversions} / \text{Clicks}$

■ Bid (or maxCPC)

- The maximum amount of money that an advertiser is willing to pay for a click

■ Cost per click (CPC or avgCPC)

- The actual amount of money that an advertiser is being charged for a click on his advertisement
- Hybrid second-price auction

■ Quality Score

- Estimate of how relevant to the promoted product are ads, keywords, and landing page

Business Models

1. CPM (Cost Per Thousand Impressions)

- Advertisers pay for exposure of their message to a specific audience.

2. **CPC (Cost Per Click)** aka Pay per click (PPC)

- Advertisers pay every time a user clicks on their listing and is redirected to their website.

3. CPA (Cost Per Action) or (Cost Per Acquisition)

- The publisher takes all the risk of running the ad, and the advertiser pays only for the amount of users who complete a transaction, such as a purchase or sign-up.

Optimization problem from the auctioneer's perspective

■ **Balance Algorithm for the Generalized AdWords Problem** [Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. 2007. AdWords and generalized online matching. J. ACM 54, 5, Article 22]

- When bidders can make differentiating bids, have different budgets, and have different CTR for different queries, the Balance Algorithm awards an ad to the advertiser with the highest value of the function:

$$\Psi = x(1 - e^{-f})$$

- x : product of the bid and the CTR for that advertiser (...or/and Quality Score)
- f : fraction of the advertiser's budget that remains unspent

■ **But...: the online advertising process *is also a profit maximization problem for the advertiser***

- Bidding Strategies, ROI Maximization, Forecasting, Behavioral Targeting
 - Need to study it initially for the sake of simplicity as a “black box” returning clicks as a function of bid
 - Real advertising campaigns contain many different features. Even the goal of optimization has different aspects - Monetary profit or Clicks (Traffic) ?

Machine Learning Strategies

- Genetic Algorithms
- Artificial Neural Networks
- Particle Filters
- Mechanisms of Reinforcement Learning
 - Decision theory – Maximum Expected Utility
 - Find a proper utility function for the optimization
 - How an *agent* ought to take *actions* in an *environment* so as to maximize some notion of cumulative *reward*
 - Highly related to *dynamic programming* techniques

Budget optimization for multiple keywords

- Investment : Final item x which is a pair (k, b)
 - keyword: k
 - bid: b
- The advertiser has j options of (k, b) candidate pairs
 - Only *one pair per investment* for his final proposal
- Total number N of the final chosen investments = r available keywords of the campaign

Multiple-choice
Knapsack
Problem Formulation

$$\text{maximize } \sum_{i=1}^r \sum_{j \in N_i} v_{ij} x_{ij}$$

Profit from all selected pairs

$$\text{subject to } \sum_{i=1}^r \sum_{j \in N_i} w_{ij} x_{ij} \leq B$$

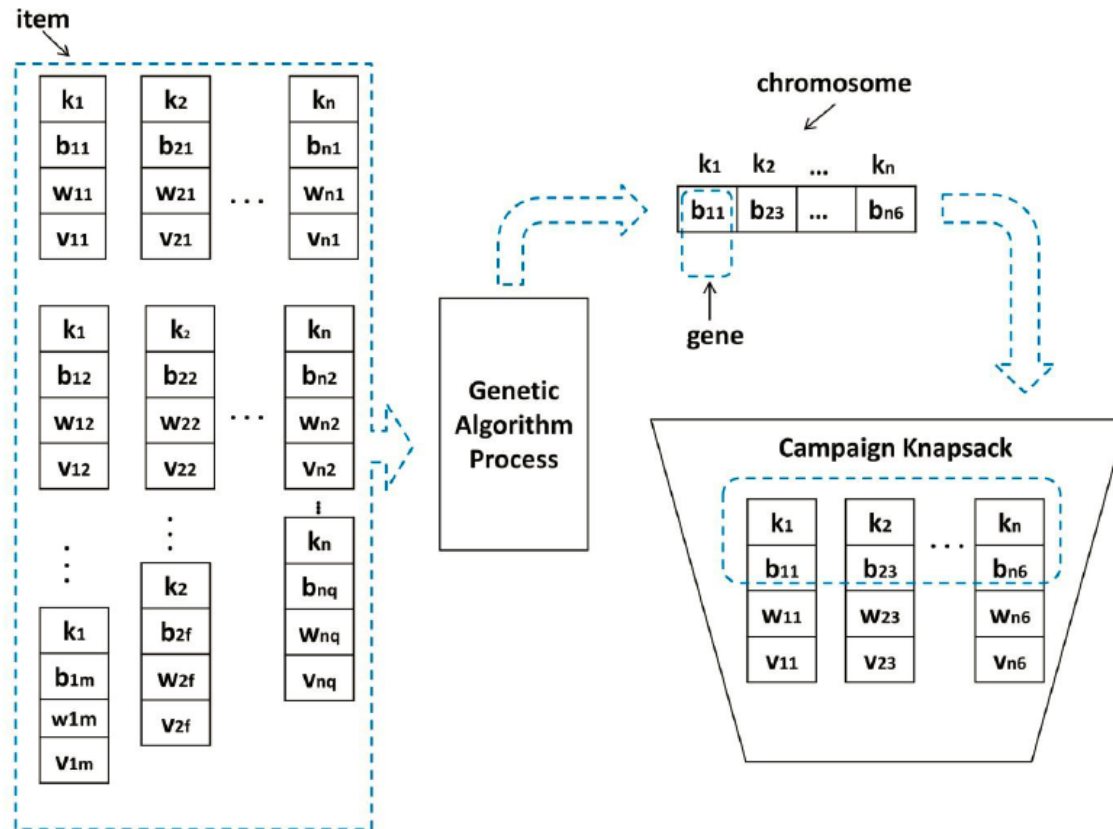
Cost of all selected pairs not exceeding budget constraint

$$\text{with } \sum_{j \in N_i} x_{ij} = 1, \text{ for all } 1 \leq i \leq r$$

and $x_{ij} \in \{0, 1\}$, for all $1 \leq i \leq r$ and all $j \in N_i$

Only one assigned bid for a specific keyword

Translate it to a Genetic Algorithm: Mapping of Campaign System to the MCKP



- Items: options of keyword-bid pairs along with their profit v and cost w
- Chromosome \equiv Set of selected items

Performance Evaluation of Advertiser's Strategies

- Evaluation Metrics? → ROI (Traffic/Clicks/CTR/ ... or... Monetary Profit/Conversions/CR?)
 - Depends on the business goals – how do you define profit?
 - Measure lift
 - Uplift modelling – A/B Tests / Control – “treated” groups

Predicting Clicks, CTR, or other metrics

- **KDD CUP 2012 - Track 2 on Kaggle** - User Click Modeling based on Search Engine Log Data: CTR Prediction Task (for auctioneers / apply similar techniques for advertisers if they have LOT of data 😊)

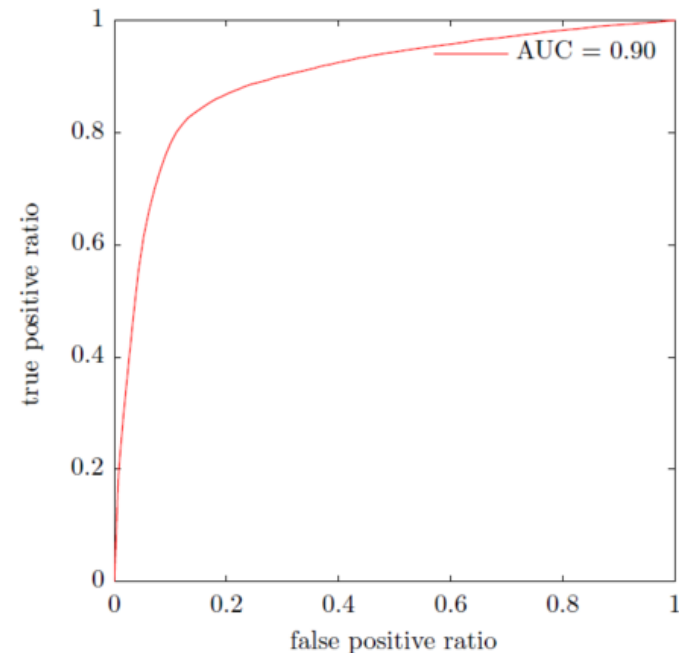
- **Boosted Regression Trees**

- Click Modeling Illustrative Features

- Demographic/firmographic /user features
- Tf-idf, BM25 scores
- Ad Quality
- CurrentPosition

- **Evaluation**

- **ROC & Area Under Curve (AUC)**



The role of text

- Keyword & Ad snippet generation
- Social Network Campaigns
 - Same basic structure for Facebook advertising campaigns
 - Generated text as a promotion tweet for Twitter
- Corporate Reputation Mining for the advertised products/services/brand names
 - Opinion Mining from web pages with *reviews*

Automated Keyword & Ad-text Generators

- Automatic setting of the proper *matching option* for each generated keyword - *Clustering* of keywords
 - Broad Match
 - Phrase Match
 - Exact Match
 - Negative Keyword
- N-gram generation
- Text summarization: Sentence extraction and compression
- **Sentiment Analysis: Keep the positive 😊**

Sentiment Analysis Filtering

- Inside the landing page might exist also negative reviews or comments that can distort our ads
 - Filter out negative snippets
- Amazon (reviews) Sentiment Dataset Snapshot
 - Balanced dataset
 - Does not contain any neutral reviews (i.e., rated with 3 stars)
 - Each line in the positive and negative set corresponds to a single snippet (usually containing roughly one single sentence)
- Start simple: Train a Naïve Bayes Classifier
 - Bag-of-words
 - Train on about 260,000 instances, test on 87,000 instances
 - Accuracy: 0.841

Summary

- There is space for further research from a machine learning approach due to system complexity
- Discover the proper features and exploit them to adjust the bid value
- Need for a good, organized dataset for our purposes
- Simulations as an initial evaluation of a bidding strategy or prediction task and then apply it to real world conditions and environment
- Online learning problems – Difficult to find a precise evaluation function for continuous involving/dynamic models

Selected References

- S. Thomaidou, M. Vazirgiannis, K. Liakopoulos. **Toward an Integrated Framework for Automated Development and Optimization of Online Advertising Campaigns**. *Intelligent Data Analysis Journal*. 2015.
- S. Thomaidou, I. Lourentzou, P. Katsivelis-Perakis, M. Vazirgiannis. **Automated Snippet Generation for Online Advertising**. ACM International Conference on Information and Knowledge Management (CIKM'13).
- S. Thomaidou, K. Leymonis, M. Vazirgiannis. **GrammAds: Keyword and Ad Creative Generator for Online Advertising Campaigns**. Digital Enterprise Design & Management Conference (DED&M'13).
- S. Thomaidou, K. Leymonis, K. Liakopoulos, M. Vazirgiannis. **AD-MAD: Integrated System for Automated Development and Optimization of Online Advertising Campaigns**. IEEE International Conference on Data Mining Workshop (ICDMW'12).
- K. Liakopoulos, S. Thomaidou, M. Vazirgiannis. **The Adomaton Prototype: Automated Online Advertising Campaign Monitoring and Optimization**. Ad Auctions Workshop, ACM Conference on Electronic Commerce (AAW'12-EC'12).
- S. Thomaidou, M. Vazirgiannis. **Multiword Keyword Recommendation System for Online Advertising**. IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM'11).
- Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz: *"Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords"*. American Economic Review 97(1), 2007 pp 242-259
- P. Maille, E. Markakis, M. Naldi, G. D. Stamoulis, B. Tuffin. *Sponsored Search Auctions: An Overview of Research with Emphasis on Game Theoretic Aspects*. To appear in the Electronic Commerce Research journal (ECR).
- Andrei Broder, Vanja Josifovski. [Introduction to Computational Advertising Course](#), Stanford University, California
- Anand Rajaraman and Jeffrey D. Ullman. Mining of massive datasets. Cambridge University Press, 2012, *Chapter 8 – Advertising on the Web*
- James Shanahan. *Digital Advertising and Marketing: A review of three generations*. Tutorial on WWW 2012
- Google AdWords Help <http://support.google.com/adwords/?hl=en>
- IAB's Internet Advertising Revenue Report <http://www.iab.net/AdRevenueReport>

Thank you! 😊

Stamatina (Matina) Thomaidou, Ph.D.

Senior Data Scientist, Data Science & Optimization

IBM Digital Sales

Find me on LinkedIn: <https://ie.linkedin.com/in/matinathomaidou>