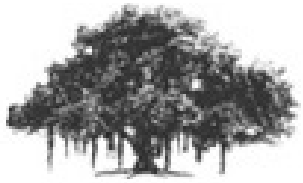


IIIT-H

Web Mining

Lecture 16: Introduction to Computational Advertising (Part 2): Contextual Advertising

Manish Gupta

21st Sep 2013

Slides borrowed (and modified) from
http://www.stanford.edu/class/msande239/

Recap of Lecture 15: Introduction to Computational Advertising (Part 1)

- Introduction to Computational Advertising
- Display Ads
- Textual Ads
- Auctions

Announcements

- Assignment 3 will be up soon.

Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- When to advertise
- Search-based ad selection for sponsored search
- Predicting clicks

Today's Agenda

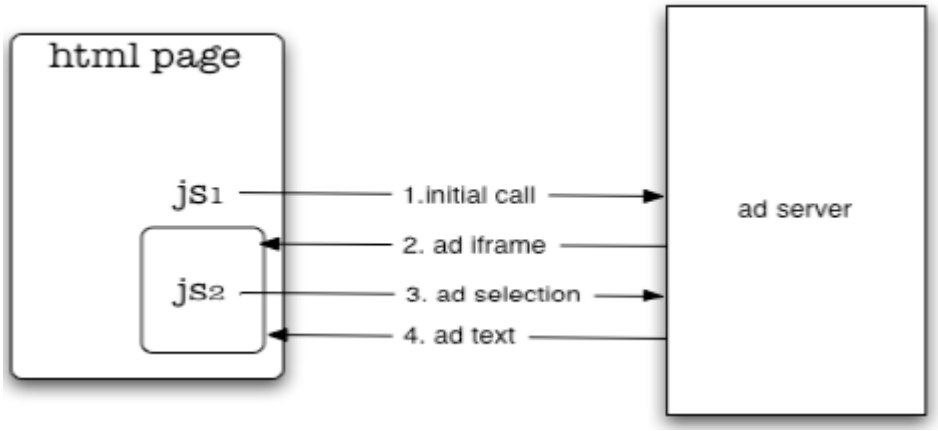
- **Contextual advertising basics**
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- When to advertise
- Search-based ad selection for sponsored search
- Predicting clicks

Contextual Advertising (Content Match)

- Textual advertising on third party web pages
- Complement the content of the web page with paid content
- Ubiquitous on the web
- Supports the diversity of the web
- Sites small and big rely on CM revenue to cover for the cost of existence
- Players
  - Google: Adsense
  - Microsoft: ContentAds

How does it all Work: The Front End

- Two main approaches
  - Page fully built by publisher using ads supplied by the ad network.
    - E.g.: XML feed (Usually done with large partners.)
  - Dynamic loading of ads



Relationship to Sponsored Search

- Main goal is to increase volume for textual campaigns in sponsored search
- Same type of ads
- Some differences with Sponsored Search
  - Lesser role of the ad network, increased role of the publisher
  - Ad Network: which ads
  - Publisher: how many/where/how
  - Ad selection using the content of a web page
  - Much more text
  - Less focused
  - Less intentional

Content Match: The Challenges

- Very thin margin business
- CTR very low – orders of magnitude, ranges in ranges 0.001-0.1%.
- Higher CTR variance
- Lower conversions – less of a clear intent
- High volume - many page views per day
- More difficult ad placement – not as intentional as search and more difficult for the advertisers to help
- Lower earnings: 1) lower bids 2) share revenue with the publisher
- Other benefits
  - User tracking

Today's Agenda

- Contextual advertising basics
- **Ad selection in contextual advertising**
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- When to advertise
- Search-based ad selection for sponsored search
- Predicting clicks

Ad Selection Methods: What Information is Provided from the Page

- Publisher can supply a variety of information to the ad network
- Page Content
  - Process the content of the page
  - Cannot be done on-line: crawl
  - Most flexible from the ad selection perspective
- Page Snippet
  - Part of the page
  - How much can we process online?
  - How much is enough?
- Custom Keywords
  - Sponsored Search – like mechanism
  - Least flexibility in ad selection
  - More control for the publisher

Two Main Implementation Strategies

- Phrase extraction (from the publisher page)
  - Map Content Match to Sponsored Search
  - Extract phrases from the page
  - Use these phrases to select ads (exact match or advanced match in Sponsored Search)
  - Ads selected on a single feature (phrase) from the page and the ad
- IR approach
  - Treat Content Match as a document similarity problem
  - Pages are compared to the ads in corpus in a common feature space
  - Bid phrase one of the features used in matching
  - Ads selected based on multiple (overlapping) features of the page and threads

Contextual Advertising Ad Selection: Case Studies

	Paper	Method
1.	Finding Advertising Keywords on Web Pages. Wen-tau Yih et al. In Proc. of WWW 2006	phrase extraction ad selection
2.	Impedance coupling in content-targeted advertising. Ribero-Neto et al. In Proc of SIGIR 2005	IR ad selection
3.	A Semantic Approach to Contextual Advertising..Broder et al, In Proc. of SIGIR 2007	IR ad selection
4.	Contextual Advertising by Combining Relevance with Click Feedback. D. Chakrabarti et al. In Proc of WWW 2008	IR ad selection using clicks
5.	To Swing or not to Swing: Predicting when (not) to Advertise. Broder et al, CIKM 2008	various

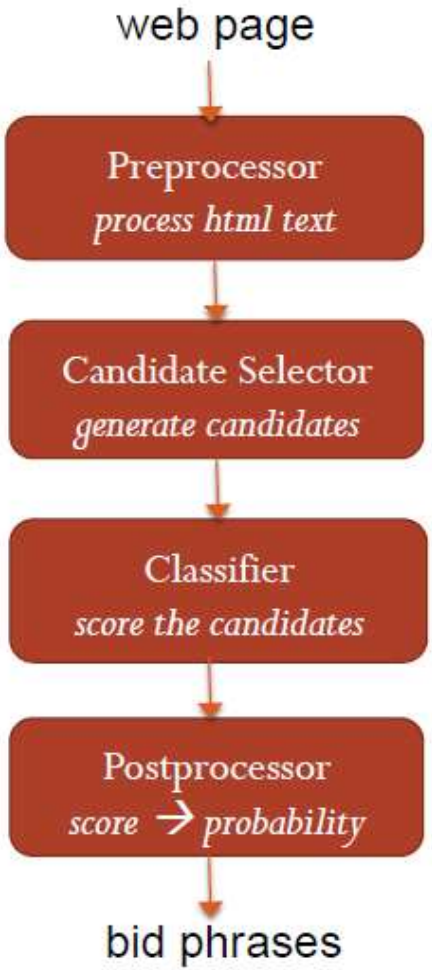
Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- **Phrase Extraction for Contextual Advertising**
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- When to advertise
- Search-based ad selection for sponsored search
- Predicting clicks

Finding Phrases for Contextual Ads

- Finding Advertising Keywords on Web Pages. Wen-tau Yih et al. In Proc. of WWW 2006
- Goal: Given a page find phrases that are good for placing ads
- Reverse search problem: Given a page, find the queries that would match (summarize) the content of this page
- Select ads based on a single selected keyword:
- Contextual Advertising translated into database approach of Sponsored Search
- Reuse of the Sponsored Search infrastructure – lower cost

System Architecture



Candidate Selection

- All phrases of length up to 5 (including single words)
  - Within a single page block (sentence)
- Two dimensions of candidate selection:
  - Individual occurrences extracted separately vs. combining all occurrences into entry per page (separate vs. combined)
  - Consider the phrase as a whole
  - Label individual words with their relationship with a phrase
    - Beginning of a phrase
    - Inside a phrase
    - Last word of a phrase

Classifier

- Given a phrase predict if it is “keyword” (usable for selecting ads)
- Binary classifier
  - Logistic regression model

$$P(Y = 1 | x = \bar{x}) = \frac{1}{1 + e^{-xw}}$$

  - x is vector of features of a given phrase
  - w is a vector of importance weights learned from the training set

Features

- Linguistic features: is a noun; is a proper name; is a noun phrase; are all words in the phrase of the same type
- Capitalization: any/all/first word capitalization
- Section based features:
  - Hypertext – is the feature extracted from anchor text
  - Title
  - Meta tags
  - URL
- IR features: tf, idf, log(tf), log(idf), sentence length, phrase length, relative location in the document
- Query log features: log(phrase frequency), log(first/second/interior word frequency)

Measuring the Extraction Quality

- Editorial judgments
- Precision-recall – might be too difficult
  - Too long for the judges to find all the relevant phrases
  - Given a phrase – influence the judges
- A proxy for P-R
  - top-1 = top-1 results is in the list selected by the editor, count across the set of pages
  - top-10 = % of top-10 results in the editor set, averaged over the set of pages

Results

system	top-1	top-10
MoC (Monolithic, Combined), *-Lin*	30.06^{b}	46.97^{b}
MoC (Monolithic, Combined), *All*	29.94	46.45
MoS (Monolithic, Separate), *All*	27.95	$44.13^{\ddagger}$
DeS (Decomposed, Separate), *All*	$24.25^{\ddagger}$	$39.11^{\ddagger}$
KEA [7]	$23.57^{\ddagger}$	$38.21^{\ddagger}$
MoC (Monolithic, Combined), *IR*	$13.63^{\ddagger}$	$25.67^{\ddagger}$
MoC (Monolithic, Combined), *TFIDF*	$13.01^{\ddagger}$	$19.03^{\ddagger}$

features		top-1	top-10
A	all	29.94^{b}	46.45^{b}
-C	capitalization	30.11	46.27
-H	hypertext	30.79	$45.85^{\dagger}$
-IR	IR	$25.42^{\ddagger}$	$42.26^{\ddagger}$
-Len	length	30.49	$44.74^{\dagger}$
-Lin	linguistic	30.06	46.97
-Loc	location	29.52	$44.63^{\dagger}$
-M	meta	30.10	46.78
-Ms	meta section	29.33	46.33
-Q	query log	$24.82^{\dagger}$	$42.30^{\ddagger}$
-T	title	28.83	46.94
-U	URL	30.53	46.39

Summary of Phrase Extraction

- Mapping Contextual Advertising to Sponsored Search
- Extract phrases from the publisher's web page
- Select ads using exact or advanced match on this phrase
- Ad selection using a single feature
- Approach based on logistic regression trained on editorial judgments
- Editors extracting salient terms from pages
- Combining the information from multiple occurrences and treating the phrases as single units yields best results
- IR and query log features account for almost all of the signal
- Low precision – difficult problem

Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- **IR methods for content match ad retrieval**
- Holistic view at the page in Contextual Advertising
- When to advertise
- Search-based ad selection for sponsored search
- Predicting clicks

Using More Than One Feature in Ad Matching

- Impedance coupling in content-targeted advertising. Ribeiro-Neto et al. SIGIR 2005
- The phrase extraction approach uses one feature of the page (phrase) to select the ads
- Risk with ambiguous phrases: 'Tahoe' is a destination as well as a truck model
- Can we select ads based on multiple features from the page?
  - What are the features of the ad?
  - How to weight the features?
  - What metrics to use to relate the ads to the pages?

Formalism for Comparing Ads and Pages: Vector Space Model

- Represent each ad a as a vector: a ={w_{1a}, w_{2a},…, w_{na}}
- In this study: a is the visible part of the ad (title and abstract)
- Represent the page p as a vector in the same space p ={w_{1p}, w_{2p},…, w_{np}}
- Weights using tf-idf method
- Use cosine of the angle between the vectors to rank the ads for a given page – denoted by sim()

Basic Set of Measures

- AD(p, a) = sim(p,a) – based on the visible parts of the ad
- KW(p,a) = sim(p, kw(a)) – based on the keyword of a
- AD_KW(p, a) = sim(p, a ∪ kw(a)) - using both the visible parts and the keyword
- Assuming that kw(a) summarizes well the essence of a, assure the presence of kw(a) in p

$$ANDKW(p,a) = \begin{cases} sim(p,a) & if\ kw(a) \subseteq p \\ 0 & otherwise \end{cases}$$

$$AAK(p,a) = \begin{cases} sim(p,a \cup kw(a)) & if\ kw(a) \subseteq p \\ 0 & otherwise \end{cases}$$

The Vocabulary Impedance Problem

- Language and the topic of the page and the ad can differ substantially
  - Publisher page belongs to a broader/narrower contextual scope
  - Ads concise in nature
  - 'Hidden topic' – not mentioned in the ad and/or the page
  - Intersection of the vocabularies of related pages and ads can be low: vocabulary impedance problem
- Solution
  - Expand vocabularies
    - Expand the page text using similar webpages
    - Expand the ad text using the text from the landing page

Summary of Context Match Ad Retrieval

- Using IR techniques to match ads and pages
- Both the ad and the page are mapped to a common vector space
- Cosine of the angle between the ad and the page as the basic similarity measure
- Bid phrase as a required feature – projection of the space
- Expanding pages using terms from similar pages improves results
- Landing page contains useful data for ad selection

Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- **Holistic view at the page in Contextual Advertising**
- When to advertise
- Search-based ad selection for sponsored search
- Predicting clicks

Motivation

- Semantic-Syntactic Approach to Contextual Advertising. AB, M. Fontoura, L. Riedel, VJ. SIGIR 2007
- Even with using multiple features there is still a risk that the subset used in matching does not represent the semantics of the page
- Can we somehow summarize the content of the whole page into a small number of features?
  - This work: supervised approach based on classification
- Use external knowledge: taxonomies
  - This work: a topical taxonomy
- What is a better signal: page class or page words? Or both?
- Both. Topical (semantic) similarity is a major component of the relevance score (~80%)

Semantic-Syntactic Match

- Figure out the topic of the page
  - Classification of the page into a commercial oriented taxonomy
- Pre-classify all the ads into the same taxonomy
- Restrict the matching to ads that are in related categories
- Use word similarity to improve the match

Page and Ad Classification

- Use a large scale classification to relate pages and ads
  - Need a taxonomy with sufficient resolution
- We used a taxonomy of 6,000+ nodes, primarily built for classifying commercial interest queries
  - Each node is a collection of query terms
- Rocchio-style nearest neighbor classifier
  - Meta-document produced of the queries at each taxonomy node
  - Classification is based on the cosine of the angle between the document and the centroid meta-documents
    - Each page or the ad document is assigned to the nearest matching taxonomy node

Taxonomy Requirements

- Enough resolution to be useful
- Not too specific to make maintenance too costly
  - Electronics - too broad
  - Electronics/Digital Camera/Canon - feasible
  - Electronics/Digital Camera/Canon/XT10i - hard to maintain

Scoring

- For a given page score every ad, select the top-k ads
- Linear combination of 2 scores
  - Taxonomy score (semantic distance)
  - Word and phrase score (syntactic distance)
- Allow generalization in the taxonomy
- $Score(p_i, a_i) = \alpha TaxScore\big(Tax(p_i), Tax(a_i)\big) + (1 - \alpha).KeywordScore(p_i, a_i)$

Semantic and Syntactic Scores

$$tWeight(kw^{si}) = weightSection(S_i) \cdot tf\_idf(kw)$$

$$Tax(p_i) = \{pc_{i1}, \ldots pc_{iu}\}$$

$$Tax(a_j) = \{ac_{j1} \ldots ac_{jv}\}$$

$$\sum_{c \in Tax(x_i)} cWeight(c) = 1$$

$$Score(p_i, a_i) = \alpha \cdot TaxScore(Tax(p_i), Tax(a_i))$$

$$+(1-\alpha) \cdot KeywordScore(p_i, a_i)$$

$$KeywordScore(p_i, a_i)$$

$$= \frac{\sum_{i \in |K|} tWeight(pw_i) \cdot tWeight(aw_i)}{\sqrt{\sum_{i \in |K|} (tWeight(pw_i))^2} \sqrt{\sum_{i \in |K|} (tWeight(aw_i))^2}}$$

where K is the set of all the keywords.

$$TaxScore(PC, AC) =$$

$$\sum_{pc \in PC} \sum_{ac \in AC} idist(LCA(pc, ac), ac) \cdot cWeight(pc) \cdot cWeight(ac)$$

$$idist(c, p) = \frac{n_c}{n_p}$$

where c represents the child node and p is the parent node.

Searching the Ad Space

- Ad search done in real time - how to make it fast enough?
- Index the ads using a inverted index
  - Use the page features as the query
- Find top-k ads with the highest score
- Monotonic scoring function that has the two sub-scores
- Evaluate the query using a variant of the WAND doc-at-a-time algorithm

Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- **When to advertise**
- Search-based ad selection for sponsored search
- Predicting clicks

Is Showing Ads "Good" in this Case?

- To Swing or not to Swing: Predicting when (not) to Advertise. Broder et al, CIKM 2008
- Repeatedly showing non-relevant ads can have detrimental long-term effects
- Want to be able to predict when (not) to show individual ads or a set of ads.
- Modeling actual short and long term costs of showing non-relevant ads is very difficult

Two Approaches

- Thresholding Approach
  - Rank the ads by score; cut-off at certain rank or score
  - Decision made on individual ads
  - Only based on ad scores
- Ad Set Machine Learning Approach
  - Decision made on sets of ads
  - Based on a variety of features
- Applies to both Sponsored Search and Contextual Advertising

Thresholding Approach

- Set a global score threshold
- Only retrieve ads with scores above threshold
- If none of the ad scores are above the threshold, then no ads are retrieved

Ad Set Approach

- Learn a binary prediction model for an entire set of ads
- If we decide to show ads, then all ads are retrieved
- If we decide not to show ads, then no ads are retrieved
- Must extract features defined over sets of ads, rather than individual ads
- Use support vector machines (SVMs)

Features

- Relevance features
  - Word overlap
  - Cosine similarity
- Vocabulary mismatch features
- Ad-based features
  - Bid price
  - Coefficient of variation of ad scores
- Result set cohesiveness features
  - Result set clarity: How much is the distribution of words different from noise (aggregate over all ads)
  - Entropy with respect to the Ad set language model
    - Language model: relative frequency of words in the ad conditioned on a given query

Summary: When (not) to Advertise

- Two approaches to determine when to show ads
  - Thresholding approach
    - Only shows ads above some global score threshold
    - Most effective for sponsored search
  - Machine learning approach
    - Predicts over entire set of ads
    - Semantic class features important for prediction
    - Effective for both sponsored search and content match
    - In practice we can combine both approaches

Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- When to advertise
- **Search-based ad selection for sponsored search**
- Predicting clicks

An Alternative View of Search Advertising

- Search Advertising Using Web Relevance Feedback: AB, P. Ciccolo, M. Fontoura, E. Gabrilovich, VJ, L. Riedel. ACM CIKM 2008

- View Search Advertising as CA on the web search result page
- More general: use the web search results as a basis for ad selection
- What are the benefits?
  - Uniform look of the result page – improved user experience
  - Re-use of the web search technology
  - Circumstantial evidence for Search Advertising
- The approach
  - Web search results as (pseudo) feedback for the web search query
  - Expanded web search query used as a long ad query
  - Evaluate the ad query to select the ads

Features?

- Derive features from snippets or full pages?
- Number of search results to obtain?
- Number of features per search result?
- How to aggregate the signals across multiple search results?

Today's Agenda

- Contextual advertising basics
- Ad selection in contextual advertising
- Phrase Extraction for Contextual Advertising
- IR methods for content match ad retrieval
- Holistic view at the page in Contextual Advertising
- When to advertise
- Search-based ad selection for sponsored search
- **Predicting clicks**

Interpreting Clicks: Positional Bias

- Ads shown on position 1 are more likely to get clicks even if they are less relevant
- How does this impact the training in our click-based weighting system?
- If the clicks of an ad are all at position 1
- Are those clicks because the ad was relevant?
- Or are those clicks caused by the inherent bias of the user to click the top ad?
- A study has shown that even if you swap the ads on position 1 and 2, position 1 still gets more clicks

De-biasing Click Data - Click Models

- To deal with this bias we need a model of user behavior
- Model #1: p(click)=p(seen)p(relevant)
  - Ads at position 1 are more likely to be seen than other positions
  - Ads at position 1 are more likely to be relevant: ranked retrieval
- We need to separate the positional and relevancy effect
- Use normalized CTR by the expected CTR at a position:
  - "The ad a is twice more likely to be clicked than an average ad at the same position"
- Count an impression only if the ad has been seen – if there is a click on a lower position – "Cascade model"
  - [Craswell et al, WSDM 2008]

Predicting CTR

- Predicting Clicks: Estimating the Clickthrough Rate for New Ads: M. Richardson, E. Dominowska, R. Ragno, WWW2008
- Features
  - Predict CTR considering ad bid phrase as query and using query CTR
  - Predict using ad features
    - Ads with same bid phrase
    - Ads with similar bid phrases
  - Ad quality features
    - Appearance: number of words in each part; word length, capitals, punctuation
    - Attention capture: action words (“buy” , “join”,…), numbers, prices, discounts
    - **Landing page:** complexity of the HTML
    - **Relevance**: bid term in the title, body; subset of the term
    - **Reputation** of the advertiser domain
  - Ad group specificity
    - Entropy of the results of the bid phrase classification
    - Number of bid phrases in the ad group
  - Web search features
    - Query frequency
    - Web page frequency

Take-away Messages

- Basic Working of Contextual Advertising Systems
- Two Ad Selection Mechanisms: Phrase extraction (from the publisher page) and the IR approach
- Using topic information for matching content and ads
- Factors to consider when deciding whether to advertise or not
- Using content match for sponsored search
- Predicting CTR

Further Reading

- Finding Advertising Keywords on Web Pages. Wen-tau Yih et al. In Proc. of WWW 2006
- Impedance coupling in content-targeted advertising. Ribeiro-Neto et al. SIGIR 2005
- Semantic-Syntactic Approach to Contextual Advertising. A. Broder, M. Fontoura, L. Riedel, VJ. SIGIR 2007
- To Swing or not to Swing: Predicting when (not) to Advertise. Broder et al, CIKM 2008
- Search Advertising Using Web Relevance Feedback: A. Broder, P. Ciccolo, M. Fontoura, E. Gabrilovich, VJ, L. Riedel. ACM CIKM 2008
- Predicting Clicks: Estimating the Clickthrough Rate for New Ads: M. Richardson, E. Dominowska, R. Ragno, WWW2008

Preview of Lecture 17: Mining Structured Information from the Web (Part 1)

- Extracting lists from web pages
- Annotating and searching web tables
- Answering table augmentation queries from unstructured lists on the Web

Disclaimers

- This course represents opinions of the instructor only. It does not reflect views of Microsoft or any other entity (except of authors from whom the slides have been borrowed).
- Algorithms, techniques, features, etc mentioned here might or might not be in use by Microsoft or any other company.
- Lot of material covered in this course is borrowed from slides across many universities and conference tutorials. These are gratefully acknowledged.

Thanks!