This lecture

1. Logistics of the course
2. Introduction & Overview of Computational Advertising
   - Classic advertising
   - Web advertising
     - Graphical ads
     - Sponsored search
     - Content match
   - The marketplace size
   - The actors: Publishers, Advertisers, Users, & “Ad agency”
   - Advertising as information
   - Trends in Computational Advertising
     - Connection with recommender systems
     - Mobile advertising
     - Open markets
3. Summary
Instructor

- **Prof. Ram Akella**
  - Professor Information Systems and Technology Management at University Of California Santa Cruz
  - Ph.D., EECS, Indian Institute of Science, Bangalore, India, 1976-1982
  - B.S., EE, Indian Institute of Technology, Madras, India, 1971-1976
  - [http://www.soe.ucsc.edu/~akella/](http://www.soe.ucsc.edu/~akella/)
  - akella@soe.ucsc.edu
Instructor

- **Dr. Vanja Josifovski**
  - Principal Research Scientist at Yahoo! Research.
  - Research Area: Computational Advertising
  - Previously at IBM Research working on databases and enterprise search
  - M.Sc. from University of Florida, PhD from Linkopings University in Sweden.
  - vanjaj@yahoo-inc.com
  - http://research.yahoo.com/bouncer_user/88
Instructor

• Dr. Andrei Broder

• Fellow and Vice President for Search & Computational Advertising in Yahoo! Research.
• Chief Scientist of Yahoo’s Advertising Technology Group.
• Research interests: computational advertising, web search, context-driven information supply, and randomized algorithms.
• B. Sc. Summa cum Laude from Technion, M.Sc. and Ph.D. in Computer Science at Stanford University under Don Knuth.
• broder@yahoo-inc.com
• http://research.yahoo.com/bouncer_user/17
Contacts

- **Course Website:**
  - [www.soe.ucsc.edu/classes/ism293/Spring09/](http://www.soe.ucsc.edu/classes/ism293/Spring09/)

- **TAs:**
  - Bin An [ban@soe.ucsc.edu](mailto:ban@soe.ucsc.edu)
  - Chunye Wang [cwang@soe.ucsc.edu](mailto:cwang@soe.ucsc.edu)

- **Course Group Email:**
  - [Ism293-group@soe.ucsc.edu](mailto:ism293-group@soe.ucsc.edu)
  - Subscribing to ISM293-group email (enrolled students only)
    - [http://www.soe.ucsc.edu/mailman/listinfo/ism293-group](http://www.soe.ucsc.edu/mailman/listinfo/ism293-group)

- **Lecture:** 6-9:30 PM Wed.

- **Office Hours:**
  - 5~6 PM Wednesdays, by appointment only!
  - Room: TBD

- **For now, please use the course group email to reach the instructors and TAs**
Lecture Content

1. Introduction and Overview
2. Information Retrieval for Computational Advertising
3. Marketplace Design
5-6. Sponsored Search
7. Content Match
8. Graphical Ads and Guaranteed Delivery
9. Behavioral Targeting
10. Overflow lecture: catch up, or optional topics:
    - Emerging Advertising Media or
    - Evaluation of On-line Advertising system
General lecture structure

- Overview: 1~1.5 hours
- In depth: 1 hour
- Discussion: 20~30 Minutes
Boot camp

- Boot camp is twice per week, 5:00 to 6:00 PM Mon. and Wed. in order to help students fill in gaps, expand on some topics, & understand the course better. We initially want to cover Linear Algebra, the Lemur toolkit, and Elementary Statistics and Probability
- Generally, TAs will lead the boot camps.
- Prof. Ram Akella will lead some boot camps as well.
Homework

- One homework per week
- Based on reading research paper, typically answer a question in the following style:
  - Why does the algorithms in the paper work?
  - Try to extend the idea in the paper from a different view of angle?
  - How to modify it for a new scenario?
- Some homeworks will be conventional exercise style.
Project

- There will be one course project divided into several mini projects. Examples:
  - Crawling Ads, Indexing, and so on (TBD)
Exam

- Two quizzes or one midterm (TBD)
- One final exam
Grading

- Homework
  - 20%
- Exam
  - 40%
- Project
  - 40%
Questions?

• We welcome suggestions about all aspects of the course!
• E-mail to Ism293-group@soe.ucsc.edu
Introduction to Computational Advertising
Disclaimers

- This talk presents the opinions of the authors. It does not necessarily reflect the views of Yahoo! Inc or any other entity.
- Algorithms, techniques, features, etc mentioned here might or might not be in use by Yahoo! Or any other company.
Computational advertising – main challenge

Find the "best match" between a given user in a given context and a suitable advertisement.

- Examples
  - Context = Web search results → Sponsored search
  - Context = Publisher page → Content match, banners
  - Other contexts: mobile, video, newspapers, etc
What is “Computational Advertising”?

- New scientific sub-discipline, at the intersection of
  - Large scale search and text analysis
  - Information retrieval
  - Statistical modeling
  - Machine learning
  - Classification
  - Optimization
  - Microeconomics
  - Recommender systems
Establishing a new discipline…

Computational Advertising | Yahoo! Research
Computational advertising is a new scientific sub-discipline, at the ... Its central challenge is to find the best ad to present to a user engaged in a ...
research.yahoo.com/Computational_Advertising - Cached

ACL-08: HLT - Introduction to Computational Advertising
ACL-08: HLT - Annual Meeting of the Association for Computational Linguistics ... A new discipline - Computational Advertising - has recently emerged, which ...
research.yahoo.com/tutorials/acl08_compadv - Cached

Conceptual Trends and Current Topics
Trends in ideas, concepts, and gestalt ... Computational Advertising ... The central challenge of computational advertising is to find the “best match” ...
kk.org/ct2/2008/01/computational-advertising.php - Cached

Lorcan Dempsey's weblog: Computational advertising
A new discipline - Computational Advertising - has recently emerged, which ...
Computational advertising poses numerous challenges and open research problems ...
orweblog.oclc.org/archives/001633.html - Cached
Establishing a new discipline...
Key messages

1. **Computational advertising** = A principled way to find the "best match" between a given user in a given context and a suitable advertisement.

2. **The financial scale for computational advertising** is huge
   - Small constants matter
   - Expect plenty of further research

3. **Advertising is a form of information.**
   - Adding ads to a context is similar to the integration problem of other types of information
   - Finding the “best ad” is a type of information retrieval problem with multiple, possibly contradictory utility functions

4. **New application domains and new techniques** are emerging every day
   - Good area for research + new businesses
Classic Advertising
Brand advertising
Goal: create a distinct favorable image
Direct marketing
Advertising that involves a "direct response": buy, subscribe, vote, donate, etc, now or soon
Long history....

Japan, 1806

USA, 1890
Lots of computational this and that …

- Computational Biology
- Computational Chemistry
- Computational Finance
- Computational Geometry
- Computational Neuroscience
- Computational Physics
- Computational Mechanics
- Computational Economics
- …

All are about mixing an old science with large scale computing capabilities
What’s computational about it?

- **Classical:**
  - Relatively few venues – magazines, billboards, newspapers, handbills, TV, etc
  - High cost per venue ($3Mil for a Super Bowl TV ad)
  - No personalization possible
  - Targeting by the wisdom of ad-people
  - Hard to measure ROI

- **Computational** – almost the exact opposite:
  - Billions of opportunities
  - Billions of creatives
  - Totally personalizable
  - Tiny cost per opportunity
  - Much more quantifiable
Revenue flow basics

- What do advertisers pay?
  - **CPM** = cost per thousand impressions
    - Typically used for graphical/banner ads (brand advertising)
    - Could be paid in advance → “Guaranteed delivery”
  - **CPC** = cost per click
    - Typically used for textual ads
  - **CPT/CPA** = cost per transaction/action a.k.a. referral fees or affiliate fees
    - Typically used for shopping (“buy from our sponsors”), travel, etc.
    - … but now also used for textual ads (risk mitigation)

- What do publishers get?
  - Whatever advertisers pay minus **rev-share** (revenue-share) paid to intermediaries

- What do intermediaries get?
  - Whatever advertisers pay minus **TAC** (traffic acquisition costs) paid to publishers
US Online Advertising Spending

US Online Advertising Spending, 2001-2011 (billions)

2001 $7.1
2002 $6.0
2003 $7.3
2004 $9.6
2005 $12.5
2006 $16.9
2007 $21.4
2008 $27.5
2009 $32.5
2010 $37.5
2011 $42.0

Note: eMarketer benchmarks its US online advertising spending projections against the Interactive Advertising Bureau (IAB)/PricewaterhouseCoopers (PwC) data, for which the last full year measured was 2006; online ad data includes categories as defined by IAB/PwC benchmark—display ads (such as banners), paid search ads (including contextual text links), rich media (including video), classified ads, sponsorships, referrals (lead generation) and e-mail (embedded ads only); excludes mobile ad spending.

Source: eMarketer, October 2007
Still growing (15% FH 2008 vs. FH 2007)

<table>
<thead>
<tr>
<th></th>
<th>FH 2007</th>
<th>FH 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>41% ($4,097)</td>
<td>44% ($5,064)</td>
</tr>
<tr>
<td>Display Related:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Banner Ads</td>
<td>21% ($2,099)</td>
<td>21% ($2,418)</td>
</tr>
<tr>
<td>-Rich Media</td>
<td>7% ($699)</td>
<td>7% ($806)</td>
</tr>
<tr>
<td>-Digital Video</td>
<td>1% ($100)</td>
<td>3% ($345)</td>
</tr>
<tr>
<td>-Sponsorship</td>
<td>3% ($300)</td>
<td>2% ($230)</td>
</tr>
<tr>
<td>Classifieds</td>
<td>17% ($1,699)</td>
<td>14% ($1,611)</td>
</tr>
<tr>
<td>Referrals/Lead Generation</td>
<td>8% ($799)</td>
<td>7% ($806)</td>
</tr>
<tr>
<td>E-mail</td>
<td>2% ($200)</td>
<td>2% ($230)</td>
</tr>
</tbody>
</table>

Source: IAB
## US Online vs. Offline advertising spend

<table>
<thead>
<tr>
<th>Year</th>
<th>Internet (billions)</th>
<th>Total media (billions)</th>
<th>Internet % of total media</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>$16.9</td>
<td>$281.6</td>
<td>6.0%</td>
</tr>
<tr>
<td>2007</td>
<td>$21.4</td>
<td>$287.5</td>
<td>7.4%</td>
</tr>
<tr>
<td>2008</td>
<td>$27.5</td>
<td>$295.5</td>
<td>9.3%</td>
</tr>
<tr>
<td>2009</td>
<td>$32.5</td>
<td>$301.5</td>
<td>10.8%</td>
</tr>
<tr>
<td>2010</td>
<td>$37.5</td>
<td>$309.0</td>
<td>12.1%</td>
</tr>
<tr>
<td>2011</td>
<td>$42.0</td>
<td>$316.0</td>
<td>13.3%</td>
</tr>
</tbody>
</table>

Note: eMarketer benchmarks its US online advertising spending projections against the Interactive Advertising Bureau (IAB)/PricewaterhouseCoopers (PwC) data, for which the last full year measured was 2006; online ad data includes categories as defined by IAB/PwC benchmark—display ads (such as banners), paid search ads (including contextual text links), rich media (including video), classified ads, sponsorships, referrals (lead generation) and e-mail (embedded ads only); excludes mobile ad spending; eMarketer benchmarks its US total media advertising spending projections against the Universal McCann data, for which the last full year measured was 2006; includes television (broadcast and cable), radio, newspapers, magazines, Internet (excludes mobile), outdoor, direct mail, yellow pages and other.

Source: eMarketer, October 2007
The advertising $$ budget vs. the human time budget

<table>
<thead>
<tr>
<th>Media Type</th>
<th>Time Spent (%)</th>
<th>Spending (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>37%</td>
<td>32%</td>
</tr>
<tr>
<td>Internet (personal and work)</td>
<td>29%</td>
<td>8%</td>
</tr>
<tr>
<td>Radio</td>
<td>19%</td>
<td>9%</td>
</tr>
<tr>
<td>Newspapers</td>
<td>20%</td>
<td>8%</td>
</tr>
<tr>
<td>Magazines</td>
<td>6%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Note: *consumer media time does not include time spent using a cellphone, watching DVDs or playing video games*

Computational Advertising Landscape
Graphical ads

SRK honoured

SRK who is to be honoured with a Malaysian title said he would not be able make it for the ceremony due to his schedule. ▶ More

• "SRK's 'Dhoom' will boost tourism in Malacca" ▶ More about Shah Rukh Khan

SRK humbled with Malaysian knighthood

I feel I'm underutilized as an actor, Suri

▶ More Movies | Movie stills | Interviews | Videos

Sports betting suffers setback after Mumbai carnage

Betting activities on international cricket matches have come to a halt after last week's terror attacks in Mumbai. Bookmakers in Lahore are not issuing rates of any cricket matches reports said. ▶ More

Find Prospects from your Community Anywhere! For FREE Registration Click Here
More graphical ads

The New York Times

Health

WORLD  U.S.  N.Y. REGION  BUSINESS  TECHNOLOGY  SCIENCE  HEALTH  SPORTS  OPINION  ARTS  STYLE  TRAVEL  JOBS  REAL ESTATE  AUTOS

FITNESS & NUTRITION  HEALTH CARE POLICY  MENTAL HEALTH & BEHAVIOR

PERSONAL HEALTH
Growing Older, and Adjusting to the Dark
By JANE E. BRODY
Published: March 13, 2007

How well do you see at night? If you’re over 50, probably not as well as you think, no matter how many carrots you eat. The typical 50-year-old driver needs twice as much light to see as well after dark as a 30-year-old. Yet few of us compensate adequately for the reduction in nighttime acuity that occurs in the aging eye.
Textual ads

1. Ads driven by search keywords – “sponsored search” (a.k.a. “keyword driven ads”, “paid search”, etc)

2. Ads directly driven by the content of a web page – “context match” (a.k.a. “context driven ads”, “contextual ads”, etc)

Textual ads are heavily related to Search and IR
Sponsored search:
Text-based ads driven by a keyword search

Also try: the leela bangalore, hotel leela bangalore, More...

Leela Palace Kempinski
Hotels in Bangalore. Up to 70% off, Book Online or Call.
www.Hotels-And-Discounts.com

Leela Bangalore India Hotels & Resorts, India Luxury, Leisure Hotel of...
Luxury hotel located near Bangalore airport. Offers accommodation, restaurants, spa treatments, business centers, and shopping mall.
www.theleela.com/hotel bangalore.html - Cached

40% Room Increased by The Leela Palaces & Resorts Luxury
The Leela Place Kempinski, Bangalore gets grander with a 40% increase in room ... Fifty percent of the Leela’s guests in Bangalore are American. ...

Unbiased Travel Reviews: Leela Palace Bangalore - TripAdvisor
Leela Palace Bangalore, Bangalore, Karnataka: Leela Palace - lives up to what the name
More sponsored search:

Text-based ads driven by a keyword search
Beastie Boys recording "political" album

By John Benson – Fri Oct 24, 4:13 am ET

CLEVELAND (Billboard) – After their Get Out and Vote tour, the group is recording the 2007 instrumental album “This Section is Open — If You’re Ready”

"We’re actually in the middle of recording the album," says Adam “Ad-Rock” Horovitz to Billboard online. "It’s called ‘This Section is Open — If You’re Ready’ and it’s political, depending on how you look at it. It’s toilet talk and fart jokes are still a part of it, though, yeah, very."

Any chance of new material from the Beastie Boys on the 2008 tour? "I don’t think so," Ad-Rock says. "I don’t think you play the new songs that much because when you’re playing the old songs, it always seems like..."
The actors: Publishers, Advertisers, Users, & “Ad agency”
Dual roles

- Sponsored search:
  - Pub = AA (Yahoo!, Google)
- Content match:
  - Pub = AA (Yahoo! content)
  - Pub = Adv (“House Ads”)
Advertising as information

- “I do not regard advertising as entertainment or an art form, but as a medium of information....” [David Ogilvy, 1985]
- “Advertising as Information” [Nelson, 1974]
- Irrelevant ads are annoying; relevant ads are interesting
  - Vogue, Skiing, etc are mostly ads and advertorials
Finding the “best ad” as an Information Retrieval (IR) problem

1. Analyze the “query” and extract query-features
   - Query = full context (content, user, environment, etc)

2. Analyze the documents (= ads) and extract doc-features

3. Devise a scoring function = predicates on q-features and d-features + weights

4. Build a search engine that produces quickly the ads that maximize the scoring function
Behind the scenes…

Setting the ad retrieval problem:

**Ads corpus** =
- Bid phrase(s) + Title + Creative + URL + Landing Page + …

**Query features** =
- Search Keywords + Outside Knowledge Expansion + Context features

**Context features (for sponsored search)** =
- Location + User data + Previous searches + …

**Context features (for context match)** =
- Location + User data + Page topic + Page keywords …

Search problem similar to web search, but

- Ad database is smaller
- Ad database entries are “small pages” [+ URL]
- Ranking depends also on bids
- Ranking depends also on click-through-rate
What is the best match?

- An ad has different utilities for publishers, advertisers, users
- Quality (utility) Factor (QF) is different
  - A-QF, U-QF, P-QF
- The ad agency has its own economic interest
- Might have different types of ads that are not easily compared
- Might have economics/contractual obligations that need to be fulfilled.
A bit deeper: how does the matching happen?
A Semantic Approach to Contextual Advertising
[SIGIR 2007]

[AB, M. Fontoura, V. Josifovski, & L. Riedel]

- What is more important: the words or the context?
- Contextual ad matching based on a combination of semantic and syntactic features.
- Classify both ads and pages into a 6000 nodes commercial taxonomy
- The class information captures the “about-ness” of pages and ads
JOE’S SKIING BLOG

.... My Atomic skis floated on the fresh powder....

(Imperfect) match

Atomic boards 75% off!!
Semantic and syntactic score

Semantic component – weighted taxonomy distance

\[ Tax(d_j) = \{d_{j1} \ldots d_{jv}\} \quad \sum_{d \in Tax(x_i)} cWeight(d) = 1 \quad idist(c, p) = \frac{n_c}{n_p} \]

\[ TaxScore(PC, AC) = \sum_{pc \in PC} \sum_{ac \in AC} idist(LCA(pc, ac), ac) \cdot cWeight(pc) \cdot cWeight(ac) \]

Syntactic component - term vector cosine

\[ tWeight(kw^{s_i}) = weightSection(S_i) \cdot tf_{-}idf(kw) \]

\[ KeywordScore(p_i, a_i) = \frac{\sum_{i \in |K|} tWeight(pw_i) \cdot tWeight(kw_i)}{\sqrt{\sum_{i \in |K|} (tWeight(pw_i))^2} \sqrt{\sum_{i \in |K|} (tWeight(aw_i))^2}} \]
Final score

\[ \text{Score}(p_i, a_i) = \alpha \cdot \text{TaxScore}(\text{Tax}(p_i), \text{Tax}(a_i)) \]

\[ + (1 - \alpha) \cdot \text{KeywordScore}(p_i, a_i) \]
Results

![Graph showing precision and recall with different parameters](graph.png)
The crystal ball chapter: Trends in Computational Advertising

A very biased sample
The Recommender Systems
Connection
The ad matching problem as a recommendation challenge

- The traditional IR approach is based on a fixed query ↔ results correspondence
- For ads we
  - Need CTR probability or user utility rather than top-K results
  - Have a continuous click-through feedback
- Challenge: incorporate the feedback
  1. **Long term loop**: improve the ranking function
     - ML based ranking
  2. **Short term loop**: use the statistics we have for a particular (query, ads) pair
     - Closest to recommender systems
A generic view – dyadic interaction systems

- D. Agarwal, B.C. Chen “Feature based factorization model for dyadic data” [In preparation]
- Dyadic Interaction data pervasive
  - Recommendation systems (user-movie, user-music, user-book)
  - Web advertising (match ads to webpage/query)
  - Content Optimization (match articles to users)
- Unit of measurement: dyad \( (i, j) \)
  - \( i = \) user, webpage,..; \( j = \) movies, ads,…
  - Measure some response: ratings, click-rates,…
  - Often have meta-data on dyadic elements
    - Demographics, genres,….
  - Goal: predict response for unknown dyads
    - Better match-making, prediction
Challenge: clicks are very rare

- Need to aggregate clicks/recommendations
  - Pages belong to site section to sites: \( \text{Page} \in \text{CNN/sports} \in \text{CNN} \)
  - Ads belong to Campaign to Advertisers: \( \text{Ad} \in \text{Ford Focus} \in \text{Ford} \)
  - Similar to preference estimation for: \( \text{Annie Hall} \in \text{Woody Allen Comedies} \in \text{Comedies} \)
Challenge: how to find new ads?

- Similar to music recommender systems: need to explore new songs but need to keep some similarity
- Can take advantage of semantic taxonomy
Mobile advertising
How popular is mobile advertising?
(Limbo’s Mobile Advertising Report, US Users, 4th quarter 2007)

Have you seen advertising in any of the following places (% of all mobile users)

- **ANY**: 31%
- **SMS**: 17%
- **MMS**: 9%
- **WAP**: 8%
- Mobile TV or Video: 5%
- Mobile Radio: 3%
Specific features (and targeting dimensions) for mobile advertising

- High precision demographics
- Location
  - Geo (GPS or tower triangulation)
  - Location functionality: Shopping mall, at home, airport, etc
  - Personal context: working, on vacation, etc
- Social context
  - Alone, with colleagues, with family, with friends, travelling, …
- Language
- Handset capabilities
- Bandwidth
- Operator
- Time of day
- Speed (walking vs. driving)
- Recent history
- Etc
What can change the game?

- Instant couponing
  - One time use coupons sent via SMS/MMS to handset
- Handset as contactless credit/cash card
  - Interaction with point-of-sale via optical readers (one way), Bluetooth (two way), or wireless network
  - Direct purchase (tickets, take-out food, etc)
- Prevalent high precision GPS
- I-Phone, G1-Phone, …
  - High res touch screen, closer to web experience
- Lots of things we haven’t thought about …
What can change the game?
Inference from noisy data

- What can you infer from lots of people reporting their position every few seconds?
  - Real time traffic …
  - Opening hours …
  - Night clubs going out of fashion …
  - Very accurate life-style predictions
  - Interesting gossip 😊
Marketplace
Advertiser agencies: The Open Exchange

- AdSense
- Ad.com

- Bids $0.60
- Bids $0.50
- Bids $0.75 via Network...

- Has ad impression to sell -- AUCTIONS
- ... which becomes $0.45 bid

- Bids $0.65—WINS!

Transparency and value
So many topics, so little time…

- **Publishers & media**
  - Internet Protocol Television (IPTV)
  - Variable-data printing (VDP) newspapers
  - …

- **Users**
  - Fluid legal landscape wrt regulation, privacy, etc.
  - As usual willing to trade privacy for functionality
  - OpenSocial, Universal Profile, etc
  - …

- **Advertisers**
  - Move towards performance advertising
  - Fine granularity ROI computation, analytics
  - See R. Lewis & D. Reiley “Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!”
  - …

- **Advertising agencies**
  - Market becoming more efficient (arbitrage, exchange)
  - Less is more – manage user attention
  - Single user multi-media campaigns
  - …
Summary
Key messages

1. Computational advertising = A principled way to find the "best match" between a given user in a given context and a suitable advertisement.
2. The financial scale for computational advertising is huge
   ⇒ Small constants matter
   ⇒ Expect plenty of further research
3. Advertising is a form of information.
   ⇒ Adding ads to a context is similar to the integration problem of other types of information
   ⇒ Finding the “best ad” is a type of information retrieval problem with multiple, possibly contradictory utility functions
4. New application domains and new techniques are emerging every day
   ⇒ Good area for research + new businesses
Thank you!

broder@yahoo-inc.com
vanjaj@yahoo-inc.com

http://research.yahoo.com
This talk is Copyright 2009. Authors retain all rights, including copyrights and distribution rights. No publication or further distribution in full or in part permitted without explicit written permission.