Content-Targeted Advertising: Issues and First Solutions

Berthier Ribeiro-Neto
November 25, 2004
Fact: The Web is today a success of public, of media, and of business.

Question: Which is the fundamental characteristic of the Web that contributed most to this success?

Freedom of publishing. In the Web, anyone publishes whatever she wants, without having to pay for it or having to pass by the censorship of a board of editors. And, in the process, can reach millions of people.
According to IDC:

“More information has been generated and stored between 1999 and 2003 than all of the information created since the beginning of mankind !"
Fact: The Web has generated great interest in the corporate market.

Question: What is the basic characteristic of the Web that makes the Web so appealing to the corporate market?

**Low communication costs.** The Web provides the lowest cost communication medium between a corporation, institution, or government and its public. E-commerce, ISP services, home banking, airline reservation, digital call centers are all becoming increasingly popular.
Fact: Various businesses, of which online advertisement is a good example, have struggled to find the right business model for the Web.

Question: What makes the Web so different from the other communication medium?

Interactivity. In the Web the user can interrupt the communication link by simply hitting a button! That action may represent the difference between the user buying a product or service from you or from your fiercest competitor.
Content-Targeted Advertising

Introduction

1995
Web gains popularity

Large expenditures on online advertising

Business model: TV Ads

2000
Nasdaq crashes
Content-Targeted Advertising

Introduction

1995
Web gains popularity

2000
Nasdaq crashes

Reduction on Investments

New online marketing strategies required

- Targeted Marketing
- Privacy Issues

Recommended for Marco A P de Cristo

Your recommendations are based on 1 items you own and more view: All | New Releases | Coming Soon | Bargains

Browse top sellers!
Sorry, we have no recommendations just for you in this category below or check back again with us soon.

1. America The Daily Show with Jon Stewart Presents Democracy Inaction
   by the writers of The Daily Show, Jon Stewart
   Publication Date: September 21, 2004
   Our Price: $14.03, Used: $9.95
Content-Targeted Advertising

Introduction

1995
Web gains popularity

2000
Nasdaq crashes

Keyword Targeted Advertising (Marketing)
Content-Targeted Advertising

Introduction

1995
Web gains popularity

2000
Nasdaq crashes

Keyword Targeted Advertising

Announcer

User

Information Gatekeeper

products
services

exposure

interests
audience

$ (keywords, ads)

information
ads

$
Content-Targeted Advertising

Introduction

Advantages of Keyword Targeted Marketing

- **Associating the right ads to a user query provides benefits to all actors**
  - **User**
    - Information of good quality
    - Ads that have appeal
  - **Mediator**
    - Credibility
    - High clickthrough rates
    - Larger revenues from the ads
  - **Announcer**
    - Lower costs
    - Receptive user actitude
    - More sales

1995
- Web gains popularity

2000
- Nasdaq crashes
Content-Targeted Advertising

Introduction

1995
Web gains popularity

2000
Nasdaq crashes
Content-Targeted Advertising

The Problem

The problem of content-targeted advertising is to associate ads with Web pages in fully automatic fashion.

From a computer science perspective
Content-Targeted Advertising

The Problem

Advertisements (ads) are
- Composed of title, description, and URL
- Associated with keywords

**AD (Advertisement)**

- Title: FlowersWhisper.com
- Description: Free shipping for all online orders, Same day delivery. www.flowerswhisper.com
- URL

**KW (Keywords)**

- “Flowers”
- “Flowers whisper”

Announcer
Content-Targeted Advertising
Matching Strategies

- Match the ad and its keywords to the Web page
Content-Targeted Advertising
Matching Strategies

- Match the ad and its keywords to the Web page

Web page

“Sun Flowers - Vincent Van Gogh”

AD (Advertisement)

“Flowers”
“Flowers whisper”

KW (Keywords)

FlowersWhisper.com
Free shipping for all online orders
Same day delivery
www.flowerswhisper.com

URL

Web page title

description
Content-Targeted Advertising
Matching Strategies

- **AD**: match the ads to the Web page
Content-Targeted Advertising
Matching Strategies

- **KW**: match the keywords to the Web page
Content-Targeted Advertising
Matching Strategies

- AD_KW: match AD+KW to the Web page
Content-Targeted Advertising
Matching Strategies

- **ANDKW**: match the keywords and force their appearance

### Web page

![Web page screenshot]

### Collection of Ad Keywords

![Collection of Ad Keywords]

**Flowers**
Content-Targeted Advertising Matching Strategies

- **AD_ANDKW (AAK):** match AD+KW and force KW appearance

Web page

Collection of Ads

*Fresh Flowers*

- **Send Fresh Flowers**
  - Welcome to VirtualFlowers.com™
  - In a hurry? Try our Quick-Pick shopping.
  - Now Featuring...
  - Flowers
  - Send Autumn Flowers
  - Did you know?
  - Upcoming Holidays
  - Thanksgiving Oct 24
  - Christmas Dec 25

*Collection of Ads*

- **FlowersWhisper.com**
  - Free shipping for all online orders
  - Same day delivery
  - www.flowerswhisper.com

*Flowers*
Content-Targeted Advertising
Matching Strategies for Placing Ads

- We distinguish 5 direct ad matching strategies

1. AD: direct match of ad
2. KW: match the ad keywords
3. AD_KW: match the ad and its keywords
4. ANDKW: match the ad keywords and force their appearence
5. AD_ANDKW (AAK): match the ad, its keywords and force their appearence
Content-Targeted Advertising
Experimental Evaluation

- Ad Sample Collection
  - 1,744 announcers
  - 93,972 ads in 2,029 campaigns
  - 68,238 keywords

- Test Collection
  - 100 pages of a Brazilian newspaper
  - Topics include economy, sports, culture, and politics

- Matching Function
  - Vector based with tf-idf weights
Content-Targeted Advertising
Tuning Idf Factors

Recall

Precision

Campaign Idf
Announcer Idf
Ad Idf
Content-Targeted Advertising
Comparison among Matching Strategies

AAK is best, 60% better than a trivial AD strategy!
Content-Targeted Advertising

Vocabulary Impedance Problem

- Web page belongs to larger contextual scope than an ad
  - Pages are about any subject
    - Many specific terms
  - Ads are concise and more topic restricted
    - More general terms

- Target page is diffuse
  - Multiple topics
  - Unimportant topic can offer good opportunity for advertising

- Vocabulary impedance problem
  - Vocabulary in the Web page is distinct from vocabulary used in the ad

- Solution: expand the vocabulary of the Web page!
Content-Targeted Advertising
Impedance Coupling

Web Page

Vocabulary Impedance

Advertisement

Expansion Terms

keywords
Content-Targeted Advertising
Impedance Coupling Model

\[ P(T_i \mid P) = \zeta \sum_{\forall j \mid t_i \in d_j} \text{sim}(p, d_j)w_{ij} \]
## Content-Targeted Advertising

### Web page terms versus expansion terms

<table>
<thead>
<tr>
<th>Rank</th>
<th>TF-IDF</th>
<th>Impedance Coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>term</td>
<td>score</td>
</tr>
<tr>
<td>1</td>
<td>argentina</td>
<td>0.090</td>
</tr>
<tr>
<td>2</td>
<td>obtained</td>
<td>0.047</td>
</tr>
<tr>
<td>3</td>
<td>class</td>
<td>0.036</td>
</tr>
<tr>
<td>4</td>
<td>whites</td>
<td>0.035</td>
</tr>
<tr>
<td>5</td>
<td>french</td>
<td>0.031</td>
</tr>
<tr>
<td>6</td>
<td>origin</td>
<td>0.029</td>
</tr>
<tr>
<td>7</td>
<td>france</td>
<td>0.029</td>
</tr>
<tr>
<td>8</td>
<td>grape</td>
<td>0.017</td>
</tr>
<tr>
<td>9</td>
<td>sweet</td>
<td>0.016</td>
</tr>
<tr>
<td>10</td>
<td>country</td>
<td>0.013</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>35</td>
<td>wines</td>
<td>0.010</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Terms in red are distinct
Content-Targeted Advertising

Strategies for placing ads

- Impedance Coupling Strategies
  - Strategies for reducing vocabulary impedance
    - AAK_T: match the ad and keywords to the set of expansion terms
    - AAK_EXP: match the ad and keywords to the expanded target page
Content-Targeted Advertising
Experimental Evaluation

- **Ad Collection**
  - 1,744 announcers
  - 93,972 ads in 2,029 campaigns
  - 68,238 keywords

- **Test Collection**
  - 100 pages of a Brazilian newspaper
  - Topics include economy, sports, culture, and politics

- **Web Collection for Term Expansion**
  - TodoBR collection
  - 5,939,061 pages of “br” domain
Content-Targeted Advertising
Impact of Impedance Coupling

![Graph showing performance metrics for different models: AAK_EXP, AAK_T, and AD_ANDKW (AAK). The graph plots Precision against Recall. Each model's performance curve is represented by a line of a different color: AAK_EXP (blue), AAK_T (magenta), and AD_ANDKW (AAK) (red). The y-axis represents Precision ranging from 0 to 0.7, while the x-axis represents Recall ranging from 0 to 1. The curves illustrate how each model's Precision decreases as Recall increases.](image-url)
Content-Targeted Advertising
Comparison among All Methods

![Graph showing comparison of different methods in content-targeted advertising. The x-axis represents recall, ranging from 0 to 1, and the y-axis represents precision, also ranging from 0 to 1. The graph compares various methods such as ANDKW, AD_ANDKW (AAK), KW, AD_KW, AD, AAK_EXP, and AAK_T.](image-url)
## Content-Targeted Advertising
### Comparison among All Methods

<table>
<thead>
<tr>
<th>Methods</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>total</th>
<th>score</th>
<th>gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>41</td>
<td>32</td>
<td>13</td>
<td>86</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td>AD_KW</td>
<td>51</td>
<td>28</td>
<td>17</td>
<td>96</td>
<td>0.106</td>
<td>+ 1.9</td>
</tr>
<tr>
<td>KW</td>
<td>46</td>
<td>34</td>
<td>28</td>
<td>108</td>
<td>0.125</td>
<td>+ 20.2</td>
</tr>
<tr>
<td>ANDKW</td>
<td>49</td>
<td>37</td>
<td>35</td>
<td>121</td>
<td>0.153</td>
<td>+ 47.1</td>
</tr>
<tr>
<td>AD_ANDKW</td>
<td>51</td>
<td>48</td>
<td>39</td>
<td>138</td>
<td>0.168</td>
<td>+ 61.5</td>
</tr>
<tr>
<td>AAK_T</td>
<td>65</td>
<td>49</td>
<td>43</td>
<td>157</td>
<td>0.226</td>
<td>+ 117.3</td>
</tr>
<tr>
<td>AAK_EXP</td>
<td>70</td>
<td>52</td>
<td>53</td>
<td>175</td>
<td>0.242</td>
<td>+ 143.3</td>
</tr>
</tbody>
</table>
Content-Targeted Advertising

What if we consider the page pointed by the ad?
Content-Targeted Advertising
Impact of using Page Pointed by the Ad

Graph showing recall and precision with different metrics for AAK_EXP_H, AAK_H, AD_ANDKW(AAK), and H.
Content-Targeted Advertising
Comparison among All Methods
Content-Targeted Advertising

Conclusions

- We have used only fully automatic methods
- Basic matching strategies can provide nice improvements in precision if engineered carefully
- More advanced impedance coupling strategies can yield large gains in average precision

- High quality content-targeted advertising is feasible and practical