Studying the Impact of Text Summarization on Contextual Advertising

G. Armano, A. Giuliani, and E. Vargiu
Intelligent Agents and Soft-Computing Group
Dept. of Electrical and Electronic Engineering
University of Cagliari, Italy
email: alessandro.giuliani@diee.unica.it
Outline

• Online Advertising and Contextual Advertising
• Text Summarization Methods
• Evaluation
• The Adopted Contextual Advertising System
• Experimental Results
• Conclusions
Online Advertising

- Form of promotion that uses the WWW for delivering marketing messages to attract customers
Contextual Advertising

Web Page

Ad

TIR 2011 – August 31, 2011 – Toulouse
Contextual Advertising

Online Carp Fishing Bait
Massive selection of carp fishing bait. Pellets, boilies, liquids.
www.bankside tackle.co.uk

Fishing
Fish in those hard to get to places No trailers - go fishing anywhere
www.porta-bote.com

Simply Fishing
Interested in Christian Videos? Watch Our Free Videos Online Today!
www.inspiration.net

We would like to be your online guide for all things having to do with fishing. Search our site and find websites specifically oriented to you and your needs. You can quickly access the most relevant categories by using the navigation on the left.

Cabela's - Fishing
Find rods, reels, jigs, bait more at Cabela's - Order now!

Canadian Fishing Trips
Take a Canadian Fly in Fishing Trip Fish big valleys northern's & trout

Carp Angling in Portugal
Company offering fishing holidays and Carp baits for Carp Angling.
www.fishingportugal.com

General Tips
Fishing is the activity of hunting for fish. It is an ancient and worldwide practice that dates back about 10,000 years with various techniques and traditions and it has been transformed by modern technological developments.

Fishing continued to be a favorite pastime in the United States, in 2001, 16% of the U.S. population 16 years old and older (16 million anglers) spent an average of 16 days fishing. Freshwater fishing was the most popular type of fishing with over 28 million anglers devoting nearly 467 million angler days to the sport.

Ads by Google

Ontario Flyin Fish Lodge
Only lodge on 3 lakes and a river. American Plan. Very nice Lodge!
www.latelleslake.com/

Quality Fishing Equipment
Lures, Tuna, Tackles all in 1 place Tested and Guaranteed
www.anglersmania.com

Fly Fishing Conditions
Up-to-Date Fly Fishing Conditions Reported by Local Outfitters
www.flyfishingconditions.com

AdSense ads
Practice Good Stewardship of our Waterways:
You can help to take care of our lakes, rivers, and other waterways so that others may enjoy these areas for years to come by practicing some of the following actions:

- Don't Litter... take along a trash bag or other receptacle for collecting your trash so that you can deposit it in the proper trash receptacle. Use proper dumping stations instead of tossing refuse into the water.
- Make sure that you use the correct type of bait and fishing gear permitted in that area. There may also be limits on the number, size, and kind of fish that you can keep. Check with your destination ahead of time to see what the local regulations allow. If you use a boat or watercraft when fishing, check to see what kinds of watercraft are allowed at the body of water you are going to fish.
- Pay attention to local procedures and cautions for cleaning your watercraft after you leave the water so that you don't encourage the spread of invasive species, such as the Toba Moth, to the rest of the world.
-
Contextual Advertising

- Task of placing ads within the content of a Web page
Classic approaches

- High computation in data, resources, and time

“Offline” advertising
Current approaches

- Need of reducing data → text summarization

→ highly dynamic! → Real time advertising!
**Focus: Text Summarization**

- Summaries:
  - obtained by single or multiple documents
  - preserve important information
  - short
Focus: TS in CA

- **Extract vs. abstract**
  - lists fragments of text vs. re-phrases content coherently in CA

- **Extractive techniques!!!**
  - Low computation
  - Reliance on single documents
  - Simple summaries (but effective)
  - Easier!
Kolcz's methods \( \rightarrow \) selection of meaningful paragraphs:

- Title
- First Paragraph
- First and Second Paragraphs
- First and Last Paragraphs
- Paragraph with most Keywords
- Paragraph with most Title Words
**TS in CA: Our Proposal**

- **Input:** HTML code → need of additional features
  - Adoption of *title* of the web page!

- **Our techniques:**
  - Title and First Paragraph
  - Title and First Two Paragraphs
  - Title, First and Last Paragraphs
  - Most Title Words and Keywords Paragraphs
Adding information about the title improves the performances

Each novel method have better performances than classic methods

The TFLP provides the best performance, as FLP does for the classic techniques
CA: The Adopted System

- Pre-processor
- Text summarizer
- Classifier

HTML document → Pre-processor → Text summarizer → Classifier

Filtered HTML document → Bag of words

Taxonomy → Classification features

Bag of words
CA: The Adopted System

Matcher

Classification features → Matcher

Bag of words → Matcher

Ads Repository

classification features

Bag of words

PAGE
• Text extraction, stop-words removing, stemming
Text Summarizer

- Summary of a web page
Text Summarizer

- Implemented for each analysed technique

Output
- Vector-based representation of BoW (TF-IDF)
Hierarchical classification
Classifier

- Semantic centroid-based classification

Output
- Vector-based representation of CF of page and ads
\[ \sigma(P, A) = \alpha \cdot \text{sim}_{BoW}(P, A) + \beta \cdot \text{sim}_{CF}(P, A) \]
The Impact of TS

- Adoption of each TS technique in the system and comparison
- Adopted Dataset (BankSearch)
### Experimental results

<table>
<thead>
<tr>
<th></th>
<th>p@1</th>
<th>p@2</th>
<th>p@3</th>
<th>p@4</th>
<th>p@5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T</strong></td>
<td>0.680</td>
<td>0.675</td>
<td>0.652</td>
<td>0.639</td>
<td>0.595</td>
</tr>
<tr>
<td><strong>FP</strong></td>
<td>0.488</td>
<td>0.476</td>
<td>0.448</td>
<td>0.421</td>
<td>0.391</td>
</tr>
<tr>
<td><strong>F2P</strong></td>
<td>0.613</td>
<td>0.609</td>
<td>0.588</td>
<td>0.558</td>
<td>0.514</td>
</tr>
<tr>
<td><strong>FLP</strong></td>
<td>0.674</td>
<td>0.653</td>
<td>0.617</td>
<td>0.582</td>
<td>0.546</td>
</tr>
<tr>
<td><strong>MK</strong></td>
<td>0.631</td>
<td>0.620</td>
<td>0.581</td>
<td>0.541</td>
<td>0.500</td>
</tr>
<tr>
<td><strong>MT</strong></td>
<td>0.640</td>
<td>0.617</td>
<td>0.610</td>
<td>0.586</td>
<td>0.547</td>
</tr>
<tr>
<td><strong>TFP</strong></td>
<td>0.744</td>
<td>0.707</td>
<td>0.691</td>
<td>0.669</td>
<td>0.637</td>
</tr>
<tr>
<td><strong>TF2P</strong></td>
<td>0.740</td>
<td>0.723</td>
<td>0.721</td>
<td>0.712</td>
<td>0.678</td>
</tr>
<tr>
<td><strong>TFLP</strong></td>
<td>0.768</td>
<td>0.750</td>
<td>0.729</td>
<td>0.701</td>
<td>0.663</td>
</tr>
<tr>
<td><strong>MTK</strong></td>
<td>0.711</td>
<td>0.698</td>
<td>0.685</td>
<td>0.663</td>
<td>0.608</td>
</tr>
</tbody>
</table>

- Title leads to an improvement of performances
Experimental results

![Graph showing experimental results](image)

- **TFLP**
- **Precision**
- **α**
- Curves for different values of p@1, p@2, p@3, p@4, and p@5
Conclusions

- We presented a comparative study on TS applied to CA.
- We propose several techniques that improve the classic methods.
- We evaluated the corresponding CA systems.
- Experimental results confirm the intuition that adopting TS techniques allows to improve performances in term of precision.
- As for future directions, further experiments are under way, with the adoption of larger dataset extracted by DMOZ.
Thanks for your attention!