WEB Image Classification
Based on the Fusion of Image and Text Classifiers

Pedro R. Kalva, Fabricio Enembreck, Alessandro L. Koerich

9th International Conference on Document Analysis and Recognition
ICDAR 2007 Curitiba, Brazil September/2007
Outline

- Introduction
- Image Classification
- Text Classification
- Fusion
- Experimental Results
- Conclusion
Webpage

Image of Interest

Advertisement

Text
Conventional Approach

- All other information is discard
- Only the image is used in the classification process

![Diagram showing the conventional approach process: Segment the Image, Extract Features, Classifier, Image Class]
Motivation

• Why not using the text found in a webpage to assist the image classification?

• Is this text related to the image?

• Main assumption of this work: Yes, it is!!
Introduction

• Problem
  – Classify images in web pages

• Goal
  – To evaluate the impact of using contextual information to image classification considering more than two classes and a large database.
    – To propose an innovative approach which uses contextual information to classify the WEB images.
Overview
Filtering
Overview

Webpage with Images

Segmentation Image / Text

Images

Text Filtering

Naïve Bayes Classifier

Vocabulary

Probabilities

Feature Extraction

Neural Network Classifier

Fusion of Classifier Outputs

Image Class
Image Classification

- Zoning

- Three feature sets
  - color, shape, texture

- Feature vector
  - 120-dimensional

- MLP Neural Network
  - backpropagation
  - *a posteriori* probabilities at the output
Overview

Webpage with Images

Segmentation Image / Text

Image Filtering

Text Filtering

Feature Extraction

Naive Bayes Classifier

Vocabulary

Probabilities

Neural Network Classifier

Fusion of Classifier Outputs

Image Class
Text Classification

• The whole text from the webpage was used as “image context”

• Filtering process is also necessary because in HTML pages there are many formatting tags, scripts and other structural elements

• A training dataset composed by 3,104 labeled text documents (1,375,485 words)

• Vocabulary of 16,182 words

• Stopwords, symbols, digits and other special characters are eliminated.

• Naïve Bayes Classifier
Overview
## Fusion of Classifier Outputs

### Rule Based Approach

<table>
<thead>
<tr>
<th>Rule #</th>
<th>Rule Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NN TOP1 ≠ NB TOP1 &amp; NN TOP1 = NN*NB TOP1 &amp; NN TOP1 = NB TOP2 &amp;</td>
</tr>
<tr>
<td>2</td>
<td>NN TOP1 ≠ NB TOP1 &amp; NN TOP1 ≠ NB TOP2 &amp; NB TOP2 ≠ NN*NB TOP2 &amp;</td>
</tr>
<tr>
<td>3</td>
<td>NN TOP1 = NB TOP1 &amp;</td>
</tr>
<tr>
<td>4</td>
<td>NN TOP1 = Class 1</td>
</tr>
<tr>
<td>5</td>
<td>NB TOP1 = Class 1</td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>NB TOP1 = Class 2</td>
</tr>
<tr>
<td>8</td>
<td>NN*NB TOP2 = Class 1</td>
</tr>
<tr>
<td>9</td>
<td>NB TOP1 = Class 4</td>
</tr>
<tr>
<td>10</td>
<td>NN*NB TOP1 = Class 1</td>
</tr>
<tr>
<td>11</td>
<td>NN*NB TOP2 = Class 1</td>
</tr>
<tr>
<td>12</td>
<td>NN*NB TOP2 = Class 0</td>
</tr>
</tbody>
</table>

**Legend:**
- **NN:** neural network classifier
- **NB:** Naive Bayes classifier
- **NN*NB:** product of the outputs of the NN and NB classifiers
- **TOP1:** output with the highest a posteriori probability
- **TOP2:** output with the second highest a posteriori probability
Experimental Results

- 5,196 webpages

- Tool to capture and label images and text

- Five classes were chosen to facilitate the gathering of data

<table>
<thead>
<tr>
<th>Class</th>
<th># of Images</th>
<th># of Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle</td>
<td>1,087</td>
<td>1,029</td>
</tr>
<tr>
<td>People</td>
<td>880</td>
<td>1,010</td>
</tr>
<tr>
<td>Pet</td>
<td>1,166</td>
<td>1,069</td>
</tr>
<tr>
<td>Motorcycle</td>
<td>1,425</td>
<td>1,004</td>
</tr>
<tr>
<td>CD/DVD Cover</td>
<td>847</td>
<td>1,057</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,405</strong></td>
<td><strong>5,169</strong></td>
</tr>
</tbody>
</table>

- Data was split into training, validation and test set
Database
Experimental Results

Correct Classification Rate (%)

- Vehicles
- People
- Pets
- Motorcycles
- CD/DVD Covers
- Average

Classes

- Image
- Text
- Image+Text
Conclusions

• Contextual information is highly relevant to improve the image classification performance.

• The fusion NN+NB classifiers through heuristic rules has increased the correct classification rate from 68.06% to 85.02%.
Questions ?