Towards the compression based fully autonomous self-organizing recognizer - preliminary implementation of PRDC-CSOR

Toshinori Watanabe
Grad. School of Inf. Sys., Univ. of Electro-Communications,
Tokyo, Japan
komkumei@gmail.com

Reuse of the IGARSS 2012 presentation:
T. Watanabe, Compression based Self Organizing Recognizer PRDC-CSOR with preliminary application to EO-image analysis
Realizing an autonomous recognizer using data compression

T. Watanabe with a long komuso-shakuhachi

ESA-EUSC-JRC 2011, ISPRA, Varese, Italy, 2011.03.31

渡辺 俊典
Toshinori Watanabe
Professor, Dr. Eng
Grad. School of Inf. Sys.,
UEC,
Tokyo, Japan
komkumei@gmail.com
Contents

• Needs for an intelligent media data analyzer
• Technological problems to be solved
• Possible solution: PRDC-CSOR*
• Preliminary experimental results

* PRDC: Pattern Representation scheme using Data Compression, IEEE TPAMI, 2002
CSOR: Compression-based Self Organizing Recognizer
Contents

• Needs for an intelligent media data analyzer

• Technological problems to be solved

• Possible solution: PRDC-CSOR

• Experimental results
Needs for an intelligent media data analyzer

• Multimedia data explosion urges us to develop
  ◦ A highly autonomous data analysis scheme for classification and/or recognition of texts, sounds, images and videos
  ◦ Earth observation (EO) image data is a typical example

• Crucial problems to be solved
  ◦ Highly automated scheme to acquire the object recognition power out of incoming data only
  ◦ Unified scheme applicable to any media data containing variety of unknown objects in them
Contents

• Needs for an intelligent media data analyzer
• Technological problems to be solved
• Possible solution: PRDC-CSOR
• Experimental results
Example image object recognition task

- Can you find some objects in the following two images?
A possible answer

- Left image contains
  - A sea-like and an apartment-like objects
- Right image contains
  - A farm-like and an apartment-like objects
A possible object finding process

Signals → Low level labels → High level labels

- Sea
- Firm
- Turf
- Building
- Apartment
Technological problems to be solved

Low level mapping

(1) Find similar (color, shape) segments in the input
(2) Give a unique label to each of them

High level mapping

(1) Find repeated co-occurrences of low level label tuples
(2) Give a unique label to each of the tuples

Signals

Low level labels

High level labels
Contents

• Needs for an intelligent media data analyzer
• Technological problems to be solved
• Possible solution: PRDC-CSOR
• Experimental results
Evaluations of possible approaches

Object model based (statistical, etc.):
(X) applicable to limited and well-formed objects

Feature space based:
(O) many objects can be represented as multi-dim. feature vectors

Human tailored target data specific feature space:
(X) effective but laborious

Automatically generated general feature space:
(O) PRDC-CSOR
Overview of PRDC-CSOR

- **Low level mapping**
  - Textization of input images
  - Compressibility feature space construction by texts
  - Text segments mapping into the feature space
  - Feature vectors hashing into integer labels

- **High level mapping**
  - Hash-based detection and labeling of co-occurring low level labels
Low level mapping (1/4)
- Textization of input images

Image-MST

MST traverse

Output text

\[ T = \text{abbabbbbcdfcfeec} \]

Encoding table

<table>
<thead>
<tr>
<th>Color Direct.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>c</td>
</tr>
<tr>
<td>_</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>_</td>
<td>b</td>
<td>d</td>
</tr>
</tbody>
</table>

Color and color-contour (partial shape) are encoded at once
Segments tends to be mapped to connected substrings

Image-MST

Output text

\[ T = \text{abbbabbbbcdffeec} \]

River

Green

Ground
Low level mapping (2/4)  
- Feature space axis dictionary formation by LZW text compression

<table>
<thead>
<tr>
<th>Input text</th>
<th>Current dictionary</th>
<th>Lngst match</th>
<th>Out code</th>
<th>Rest text</th>
<th>New dic. Memb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.a.a.b</td>
<td>[a-1, b-2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a.a.a.b</td>
<td>[a-1, b-2]</td>
<td>a</td>
<td>1</td>
<td>a.a.b</td>
<td>a.a-3</td>
</tr>
<tr>
<td>a.a.b</td>
<td>[a-1, b-2, a.a-3]</td>
<td>a.a</td>
<td>3</td>
<td>b</td>
<td>a.a.b-4</td>
</tr>
<tr>
<td>b</td>
<td>[a-1, b-2, a.a-3, a.a.b-4]</td>
<td>b</td>
<td>2</td>
<td>nil</td>
<td>nil</td>
</tr>
<tr>
<td>nil</td>
<td>[a-1, b-2, a.a-3, a.a.b-4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dictionary used for feature space axis formation
Low level mapping (2/4)
- Process of text mapping into the feature space by the dictionary-conditioned compressibility

<table>
<thead>
<tr>
<th>Input text</th>
<th>Frozen dictionary</th>
<th>Lngst match</th>
<th>Out code</th>
<th>Rest text</th>
<th>New dic. Memb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.a.b.b</td>
<td>[a-1, b-2, a.a-3, a.a.b-4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a.a.b.b</td>
<td>[a-1, b-2, a.a-3, a.a.b-4]</td>
<td>a.a.b</td>
<td>4</td>
<td>b</td>
<td>--</td>
</tr>
<tr>
<td>b</td>
<td>[a-1, b-2, a.a-3, a.a.b-4]</td>
<td>b</td>
<td>2</td>
<td>nil</td>
<td>--</td>
</tr>
<tr>
<td>nil</td>
<td>[a-1, b-2, a.a-3, a.a.b-4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Compression ratio = $\rho(x \mid \text{Dic})$
  = Output-length / Input-length
  = $|4.2| / |a.a.b.b| = 2 / 4 = 0.5$
  = Feature of the text from the viewpoint of the dictionary
Low level mapping (2/4)
- Original compressibility feature space of PRDC

Texts / Strings

\[ T2 = bbbbbbbbbbb \]
\[ T4 = bbbbabbbba \]
\[ T3 = aaaabaaaab \]
\[ T1 = abababaaa \]

Feature space

\[ \rho(x|\text{Dic}(T2)) \]
\[ \rho(x|\text{Dic}(T1)) \]

LZW
Frozen LZW

CV(T1) CV(T3) CV(T4) CV(T2)
Low level mapping (2/4)
- Similar dictionaries cause low resolution of the space

- We should use orthogonal dictionaries to span the space
Low level mapping (2/4) - Orthogonal feature space generation by PRDC-CSOR

Texts

\[ T_1 = \text{aabababa...} \]
\[ T_2 = \text{aaaaababa...} \]
\[ T_m = \text{bbbbbabbb...} \]

Segments of various lengths

Dic.A

Dic.B

Dic.C

LZW

Frozen LZW

Images

To cope with meaningful length uncertainty

\[ T_1 = \text{aaba, baba, aabababa} \]
\[ T_2 = \text{aaaa, baba, aaaababa} \]
\[ T_m = \text{bbbb, abbb, bbbbabbb} \]
Low level mapping (3/4) - Text segments mapping into the feature space

Texts

- T1 = aabababa...
- T2 = aaaaababa...
- Tm = bbbbabbb...

Segments of various lengths

To cope with the meaningful string uncertainty

Images

- I1
- I2
- Im

To cope with the meaningful string uncertainty

Dic.A

- aaba
- aaaa

Dic.B

- baba
- bbbb

Dic.C

- abbb

LZW

Frozen LZW

T1 = aaba, baba, aabababa
T2 = aaaa, baba, aaaaababa
Tm = bbbb, abbb, bbbbabbb

ESAEU-ESC-JRC-2012, Fri. 26 Oct., T. Watanabe, U.E.C Tokyo, Japan
Low level mapping (4/4)  
- Low level labeling by CV hashing

The integer value #4225039 becomes the low level object label of u
High level mapping
- Hashing based detection and labeling of co-occurring low level label tuples

<table>
<thead>
<tr>
<th>Low level label</th>
<th>Occurrences</th>
<th>Tuple hash value</th>
<th>High level label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>image1, image2</td>
<td>hash('image1.image2')</td>
<td>#15888430</td>
</tr>
<tr>
<td>2</td>
<td>image1, image2</td>
<td>hash('image1.image2')</td>
<td>#15888430</td>
</tr>
<tr>
<td>3</td>
<td>image2, image3</td>
<td>hash('image2.image3')</td>
<td>#2364674</td>
</tr>
<tr>
<td>4</td>
<td>Image2, image3</td>
<td>hash('image2.image3')</td>
<td>#2364674</td>
</tr>
<tr>
<td>5</td>
<td>image3</td>
<td>---</td>
<td>singleton</td>
</tr>
</tbody>
</table>
Contents

- Needs for an intelligent media data analyzer
- Technological problems to be solved
- Possible solution: PRDC-CSOR
- Experimental results
Experiment 1

- Two artificial apartment images (128 x 110 pixels)
  - Seaside apartment
  - Farm side apartment

- The compressibility feature space can separate two texts from the two images very well

![Graph showing compression ratio vs CV space axis number]

Good at Farm side

Good at Seaside

100-dim feature space
Discovered objects

Main parameters:
Min. string length = 16 chars
Hashing threshold $\theta$ for low level labeling = 0.6
Experiment 2

- Two airborne rural images

(1) Rural-1

(2) Rural-2

* Dominant 3 out of 11 low level labels are shown

#740 (rural-triple)

→ #740 (firm) + #3572 (bush?) + #3313 (houses)
CONCLUSION

• PRDC-CSOR is proposed
  ◦ Given a set of images, it automatically constructs the feature space spanned by a set of compression dictionaries
  ◦ Textized images segments are mapped into the space and given integer low level labels by hashing
  ◦ Co-occurring low level label tuples are found and labeled by hashing to discover compound objects
  ◦ Feasibility of PRDC-CSOR was illustrated by preliminary experiments

• Future tasks
  ◦ Refinements of labeling functions through realistic experiments
  ◦ Validation of the universal applicability of PRDC-CSOR to various media data
Challenging application
- commuting route recognizer

- **Input**
  - List of images along my commuting route
  - Taken on Monday, Tuesday, …, and so on

- **Output**
  - Meaningful objects appearing in these images
    - House, fence, duck, sky, hill, building, bicycle, rail, tree, door, book, etc.

- **Processing**
  - PRDC-CSOR scheme
My commuting route images
Komuso and children at dusk
By T. Watanabe

Thank you!