Classification of Images Using Their Neighboring Sentences

Kazutaka SHIMADA[†] **Tetsuro ITO** [†] **Tsutomu ENDO**[††]

[†]Department of Computer Science and Intelligent Systems
Oita University
700 Dannoharu Oita, 870-1192 Japan
E-mail:{shimada, ito}@csis.oita-u.ac.jp

[††]Department of Artificial Intelligence
Faculty of Computer Science and Systems Engineering
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN
E-mail: endo@pluto.ai.kyutech.ac.jp

Abstract

There are many images on the WWW. Traditional image analysis, however, can not estimate the contents of images because they lack information such as colors, resolutions and so on. On the other hand a page on the WWW consists of sentences and images, and the images are closely related to their neighboring sentences. In this paper, we propose a method for classification of images using surface expressions in their neighboring sentences. First, the method extracts images and relational sentences from a page on the WWW. It then classifies the contents of images using weighted keywords. We evaluated the performance for three keyword sets, which were constructed by handwork, a TF*IDF method and a Bayes theorem based method. The highest accuracy attained to 99.5% by the Bayes theorem based method for the training data, and 79.0% by the TF*IDF method for the test data. The keywords by the TF*IDF method and the Bayes theorem based method outperformed those by the handwork for both the training and test data. They were suited to this classification.

Keywords Surface expression, Neighboring sentences, Classification, WWW

1. Introduction

There is a vast amount of information on the WWW. Information on the WWW is not only text but also images and tables. Traditional image analysis, however, can not identify the contents of images because they lack information such as colors, resolutions and so on. On the other hand a page on the WWW consists of sentences and images, and the images are closely related to their neighboring sentences. Their neighboring sentences are appropriate for classifying images of the WWW because classification using them can handle low-resolution images. There are several approaches to identify the relation between text and images [1][2][3][4][11][13]. However, [1],[11] and [13] only align a person's name and picture. Our proposed method can classify images of corporate advertising about computer systems. [2]-[4] require some information such as colors and text in the images. To use information from images on the WWW is unsuitable because of the low-resolution.

(C) 2001 The Pacific Association for Computational Linguistics

We are developing a multi-specifications summarization system using extracted important data from specifications[5][6] and integrating the summary and images for multimedia summarization. The outline of the system is shown in Figure 1. A multimedia summarization system is required to identify and classify the content of images to integrate sentences and images. Since we deal with images on the WWW, an image analysis method not fully dependent on the quality of images is necessary. A home page is exemplified in Figure 2. We proposed a method for identification and classification of images using their neighboring sentences in [7]. The keywords for classification were constructed by handwork, but they were subjective. Experimental results were insufficient. In this paper, we evaluate the performance for three keyword sets, which are constructed by handwork, a TF*IDF method and a Bayes theorem based method.

2. Image and Neighboring Sentences Extraction

We deal with pages of corporate advertising about computer systems. Our system extracts images and their neighboring sentences from the HTML source.

Figure 2. A home page

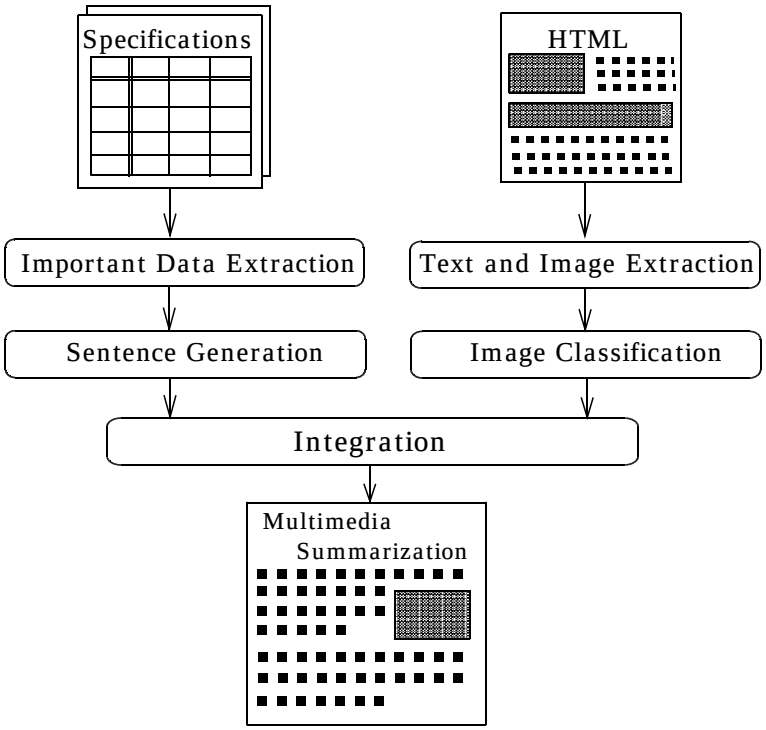


Figure 1. Outline of our system

(C) 2001 The Pacific Association for Computational Linguistics

2.1 Image Extraction

There are two kinds of images in a page of corporate advertising about computer systems.

Type1: Letter images, button images and so on.

Type2: Images about the products.

These image types are exemplified in Figure 3. We deal with **Type2** images for classification. Our system extracts **Type2** images using the size of an image. Conditions for extraction are constructed from 337 images, and are given as follows:

1. The size of an image is more than 55 * 55(pixels).
2. The ratio of width to height is less than 3.8.

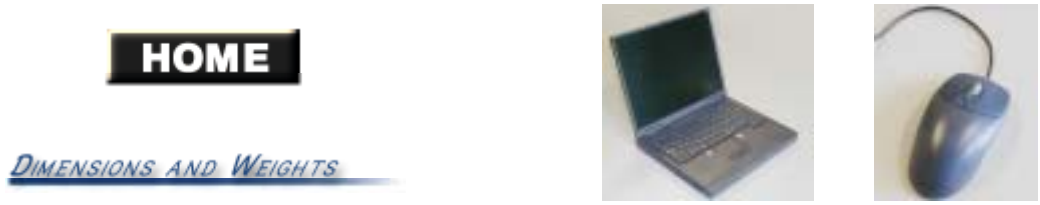
If the height of bordering images is the same, we handle them as one image because a large image is often divided into several small ones for the WWW.

2.2 Neighboring Sentences Extraction

The system extracts neighboring sentences of images. Their neighboring sentences are extracted using the tags of HTML as follows:

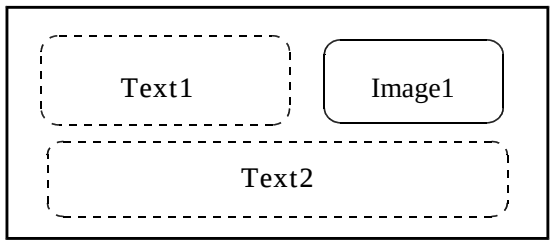
1. Extract <IMG> tag from a HTML source.
2. Extract sentences around the tag.
3. If "ALT" tag exists in <IMG>, examine the contents of "ALT", and extract sentences which include it.
4. Estimate the layout of the extracted sentences from tags.
   There are three tags about the layout: <TABLE>,
 and <HR>. The order of priority of the tags is as follows: "<TABLE>" > "
" > "<HR>".
   Extract the sentences which are the nearest to <IMG> tag.

The extraction process is exemplified in Figure 4. "Text1" and "Text3" are the neighboring sentences of "Image1" and "Image2", respectively.



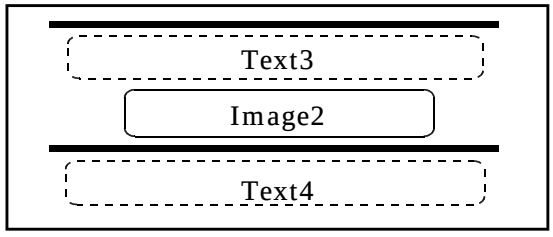
Type1 Type2

Figure 3. Image types




```
<TABLE>
  <TR>
    <TD>Text1</TD>
    <TD><IMG src="Image1"></TD>
  </TR>
  <TR>
    <TD colspan="2">Text2</TD>
  </TR>
</TABLE>
```

(a)




```
<HR>
Text3<BR>
<IMG src="Image2">
<HR>
Text4<BR>
```

(b)

Figure 4. Layout and HTML

(C) 2001 The Pacific Association for Computational Linguistics

3. Keyword Extraction and Weighting

We define keywords and weights to classify images. There are three kind of keywords, which are constructed by handwork, a TF*IDF method and a Bayes theorem based method. The system classifies images into 12 classes. The class names and their image examples are shown in Table 1. These class names denote the topics of "Sentence Generation" in Figure 1.

3.1 Handwork

We extract the keywords by handwork from the training data. The keywords extracted by handwork are exemplified in Figure 5. The keywords possess the weight 1 or 2. The system searches the keywords in the 1st sentence. If a keyword exists in this sentence, its score is multiplied by 2. If there are the tags, <B> and/or <FONT>, the score is multiplied by *n*. The value of *n* is shown in Table 2. The system classifies each image into the class showing the maximum of the total score.

3.2 TF*IDF Method

The TF*IDF weight of a term in one document is the term frequency(*tf*) divided by its document frequency(*df*). A document consists of neighboring sentences of an image. Keywords are extracted automatically using surface expressions. The process of the keyword extraction is as follows:

1. Eliminate hiragana expressions.
2. Eliminate 2-bytes symbols(ex. ◆,□).
3. Eliminate a word with 1 character.
4. Divide a word into several ones by symbols(ex. CD-R/RW → CD-R, CD-RW).
5. Eliminate a word including numerals. (exceptions: IEEE1394, MP3 and so on)
6. Eliminate stopwords.

Figure 6 is an example of the keyword extraction.

The weight is obtained by:

(画像|グラフィック)(を)?(扱う|あつかう|処理|処理する)?ボード
(Image|Graphic)(processing)? board

"?" denotes a word which produces once or nothing, "|" denotes "or"

Figure 5. Keywords extracted by handwork

Table 1. Class names

Class name	Content of image
Basic Spec.	CD-ROM, DVD-ROM
Extension	USB, PCI
Image Processing	IEEE1394, Painting soft
Display	TFT, CRT display
Operation	Keyboard, Mouse
Size	Comparison of size
Communication	Communication equipment
Sound	Speaker
Information	Service
Other models	Relational models
Soft	Attending soft
Others	Others

Table 2. The weight of tags

Tag	*n*
<FONT color=...>	1.5
<B>	1.5
<FONT size=+...>	1.75
<FONT size=－...>	0.75
<FONT size=+...> and <FONT color=...> or <B>	2.0

... USB を 2 つ装備している .....
(... equipped with USB ports[2] ...)
↓
Keywords: USB, 装備
(USB, equipped)

Figure 6. Keyword extraction

(C) 2001 The Pacific Association for Computational Linguistics

1. Calculate the weight of each document.

$$dw(d_k, i) = w \times tf_i \times log\frac{N}{df_i}$$

$dw(d_k, i)$: The weight of $term_i$ in a $document_k$
w: The weight for tf_i
N: The number of documents

2. Calculate the weight of each class.

$$cw(C_j, i) = \sum_{i \in Cj} dw(d_k, i) \times \frac{df(C_j, i)}{df_i}$$

$cw(C_j, i)$:The weight of $term_i$ in $Class_j$
$df(C_j, i)$: The number of documents including $term_i$ in $Class_j$

3.3 Bayes Theorem Based Method

The Bayes theorem based method is a probabilistic one. The probability that $term_i$ belongs to class C_k is given by:

$$Pr(C_k|i) = \frac{Pr(C_k) \times Pr(i|C_k)}{\sum_{j=1}^{r} Pr(C_j) \times Pr(i|C_j)}$$

r: the number of classes(r=1, ... ,12).

We handle the probabilities as the weights of the keywords.

4. Evaluation

In this section, first, we evaluated the proposed image extraction method using 787 images. Next, we used three keyword sets(Handwork, TF*IDF method, Bayes theorem based method) and evaluated their performance.

4.1 Image Extraction

We used 337 images(the training data) for constructing the conditions to extract images, and 450 images(the test data). The experimental results are shown in Table 3.

In our system, a recall rate is important because the image extraction is the pre-processing for the image classification. The conditions to extract images produced the high recall rates for both the training and test data.

4.2 Image Classification

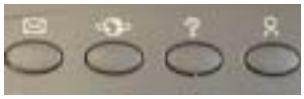
We used 325 images possessing their neighboring sentences. The images and their neighboring sentences are exemplified in Figure 7. First, we used the three keywords sets, which were constructed by handwork, the TF*IDF method and the Bayes theorem based method, and evaluated their performance. The number of keywords which were constructed by handwork was 184. The number of keywords for the TF*IDF method and the Bayes theorem based method was 1111. The training data consists of 182 documents. The experimental results are shown in Table 4. The symbol "w" in Table 4 denotes the weight of each keyword in the 1st sentence. For the TF*IDF method, we normalized the weights because the frequency of each class is different.

Table 3. Results of image extraction

	Training data	Test data
Recall	98.0%	97.2%
Precision	91.5%	94.0%

●背面（the back）
1. PC カード（PCMCIA support）
2. USB
3. シリアルポート（Serial ports）

Class name: Extension



◆イージー・アクセスボタン（Easy Access Button）
使いやすいインターネット対応ボタンやメール対応ボタンを装備しています。(PC1 is equipped with some buttons for the internet, e-mail and so on)

Class name: Operation

Figure 7. Examples of images and their neighboring sentences

(C) 2001 The Pacific Association for Computational Linguistics

Table 4. Results for the training data

		w=1	*w*=2	*w*=5	*w*=10	*w*=20
TF*IDF	Standard	89.0%	90.1%	91.2%	90.1%	87.9%
	Normalized1	90.1%	90.7%	91.8%	89.5%	90.1%
	Normalized2	90.7%	92.9%	92.3%	91.8%	90.1%
Bayes theorem		97.3%	98.9%	99.5%	99.5%	99.5%
Handwork		78.0%				

Table 5. Results for the test data

		w=1	*w*=2	*w*=5	*w*=10	*w*=20
TF*IDF	Standard	70.6%	72.0%	68.5%	68.5%	66.4%
	Normalized1	70.6%	73.4%	70.6%	69.9%	67.8%
	Normalized2	78.3%	79.0%	76.2%	73.4%	70.6%
Bayes theorem		74.1%	74.8%	75.5%	74.8%	72.7%
Handwork		66.4%				

$$Normalized1 = 1 - \frac{df_c}{N} \qquad Normalized2 = \log \frac{N}{df_c}$$

N: The number of documents
df_c: The frequency of a class

The Bayes theorem based method produced the best performance. The TF*IDF method obtained the high accuracy. The accuracy of TF*IDF method was lower than that of the Bayes theorem based method because the TF*IDF method overtrained the weight of each keyword.

Next, we evaluated their performance by the test data. The test data consists of 143 documents. The experimental results are shown in Table 5. The TF*IDF method was effective in case of insufficiency of the keywords in documents because each weight of the TF*IDF method was larger than that of the Bayes theorem based method. We calculated the weights of the TF*IDF method and the Bayes theorem based method using the test data. As a result, the accuracy of the TF*IDF method(*Normalized2* and *w*=2) and the Bayes theorem based method(*w*=5) were 89.5% and 97.2%, respectively.

We evaluated the number of keywords. We reduced the number of the keywords by a threshold using the TF*IDF value. The number of the keywords was 659. As a result, the accuracy using the keywords was 84.6%(the training data). For the test data, it was 63.4%. The reduction of the number of keywords using the TF*IDF value was not effective, especially for the test data.

On the whole, the TF*IDF method and the Bayes theorem based method calculated the valid weights of the keywords and outperformed the handwork for both the training and test data. They were suited to this classification. The reason why the results by the handwork is not good is that the weight of each keyword is 1 or 2 only. It is, however, difficult to determine the continuous weight by handwork. The Bayes theorem based method performed better than the TF*IDF method if there were sufficient keywords. The system requires increasing the number of keywords for the high accuracy, especially for the Bayes theorem based method.

5. Conclusions

In this paper, we propose a method for classification of images using their neighboring sentences and evaluated the performance for three keyword sets, which were constructed by handwork, a TF*IDF method and a Bayes theorem based method. In particular, the keywords by the TF*IDF method and the Bayes theorem based method outperformed those by the handwork. We showed the effectiveness of image classification using neighboring sentences of images by way of the experiments. Our system can classify low-resolution images by using their neighboring sentences. Future

(C) 2001 The Pacific Association for Computational Linguistics

work will include improvement of an algorithm to extract the keywords for the TF*IDF method and the Bayes theorem based method.

References

[1] K. Yamada, K. Sugiyama and H. Nakagawa. Learning relations between linguistic expressions and pictures in newspaper, Technical report of IEICE, NLC97-63, pp.65-70, 1998(in Japanese).

[2] Y. Watanabe, Y. Okada, K. Kaneji and Y. Sakamoto. Retrieving related TV news reports and newspaper articles, IEEE Intelligent SYSTEMS, Vol.14, No.5, pp.40-44, 1999.

[3] Y. Watanabe and M. Nagao. Diagram understanding for pictorial book of flora using integration of pattern information and natural language information, JSAI, Vol.11, No.6, pp.888-895, 1996(in Japanese).

[4] Y. Watanabe and M. Nagao. Image analysis using natural language information extracted from explanation text, Vol.13, No.1, pp.66-74, 1998(in Japanese).

[5] K. Shimada and T. Endo. Extracting important data from specifications, IPSJ SIG, NL133-15, pp.107-113, 1999(in Japanese).

[6] K. Shimada and T. Endo. Sentence generation from table structure of extracted important data, Technical report of IEICE, TL99-29, pp.25-31, 1999(in Japanese).

[7] K. Shimada, T. Ito and T. Endo. Image identification and classification using surface expressions, IPSJ SIG, NL139-8, pp.55-60, 2000(in Japanese).

[8] W T. Chuang and J. Yang. Extracting sentence segments for text summarization: A machine learning approach, SIGIR 2000, pp. 152-159, 2000.

[9] M. Pazzani and D. Billsus. Learning and revising user profiles: The identification of interesting web sites, Machine Learning 27, pp.313-331, 1997.

[10] J. Favela and V.Meza. Image-retrieval agent: integrating image content and text, IEEE Intelligent SYSTEMS, Vol.14, No.5, pp36-39, 1999.

[11] R. Houghton. Named Faces: Putting names to faces, IEEE Intelligent SYSTEMS, Vol.14, No.5, pp.45-50, 1999.

[12] R. Srihari. Computational models for integrating linguistic and visual information: a survey, Artificial Intelligence Review 8, pp.349-369, 1995.

[13] R. Srihari. Use of captions and other collateral text in understanding photographs, Artificial Intelligence Review 8, pp.349-369, 1995.

(C) 2001 The Pacific Association for Computational Linguistics