Automatic Generation of Explanatory Text from Flowchart Images in Patents

Hidetsugu Nanba, Shohei Kubo and Satoshi Fukuda

Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551 JAPAN

Abstract

This paper addresses the automatic generation of explanatory text from flowchart images in patents. The construction of an explanatory text generator consists of four steps: (1) automatic recognition of flowchart images from patent images, (2) extraction of text strings from flowchart images, (3) creation of data for machine learning, and (4) construction of an explanatory text generator using T5. In this study, a benchmark consisting of 7,099 images was constructed to determine whether an image in a patent is a flowchart. Furthermore, an explanatory text generator was constructed from the images using 11,188 flowchart image-explanatory text pairs. The experimental results showed that a recognition accuracy of 0.9645 was achieved for flowchart images. Although high-quality explanatory text could be generated from flowchart images, some issues remain for flowcharts with complex shapes.

Keywords
Flowchart, Image recognition, Text generation, Character recognition, Patent

1. Introduction

A procedural text is a description of a set of procedures to achieve a particular objective. Our goal is to automatically extract knowledge about a series of procedures in a wide range of fields from texts and systematize them. Here, we describe the automatic generation of explanatory text from flowchart images in patents.

In automatically generating explanatory text for the flowchart images, we focus on the abstract and selected figures of the patent. A selection figure enables us to grasp the outline of the invention quickly and accurately. The applicant usually selects a diagram from among the diagrams in the patent that they consider necessary for understanding the abstract contents. If a classifier that automatically determines whether an image in a patent is a flowchart or not is constructed and only those selected diagrams that are flowcharts are extracted, a large number of pairs of flowcharts and their explanatory texts (i.e., patent abstracts) can be generated automatically. Furthermore, using these pairs, we believe it is possible to construct a system that automatically generates explanatory text from flowchart images using machine learning.

The contributions of this paper are as follows:

- To determine whether an image in a patent is a configuration diagram, flowchart, or table, we constructed a benchmark consisting of 7,099 images. We used this benchmark to achieve a classification accuracy of 0.9645.
- We constructed 11,188 pairs of flowchart images and their descriptions automatically.
- Using these pairs, we constructed a system that automatically generates explanatory text from flowchart images through machine learning.

2. Related Work

2.1. Flowchart Analysis

Services that share flowcharts, such as myExperiment and SHIWA, have started recently, which has led to a demand for techniques to search for similarities between one flowchart and another flowchart [1]. A related research project in flowchart image analysis is CLEF-IP, which refers to a task targeting patents [2]. The Conference and Labs of the Evaluation Forum (CLEF) is a workshop on information retrieval held mainly in Europe. CLEF-IP recognizes shapes, detects text, edges, and nodes that are elements of flowcharts, and recognizes flowcharts. Herrera-Cámara also worked to recognize flowchart images [3]. In addition, Sethi et al. identified flowcharts from diagram images in deep learning-related papers and further analyzed the flowcharts to build a system that outputs the sources in Keras and Caffe [4]. This research differs from theirs in that we take a flowchart image as input and output its description as a natural language sentence. We considered the availability of resources such as the CLEF-IP for our work, but as it is too small to be used as training data for the generation of explanatory texts, this study started with the creation of training data.

2.2. Generating Text from Figures and Tables

Chart to text refers to the task of generating natural language sentences to describe the important information derived from charts and tables. Zhu et al. [5] addressed this problem by building a system,
3. Automatic Generation of Explanatory Text from Flowchart Images

The construction of the generator of explanatory text consists of the following four steps: (Step 1) automatic recognition of flowchart images; (Step 2) extraction of character strings from the flowchart image; (Step 3) creation of data for machine learning; and (Step 4) construction of an explanatory text generator using T5. Each procedure is described as follows.

(Step 1) Automatic recognition of flowchart images

Convolutional neural networks (CNNs) are used to recognize flowchart images in patents. Our method uses seven CNN models trained on a large image data set called “ImageNet” to construct a learning model by fine tuning, and its effectiveness is verified through the experiments described in Section 4.

(Step 2) Extraction of character strings from the flowchart image

An optical character recognition function in Google Cloud Vision (https://cloud.google.com/vision) is used to extract text strings from flowcharts. An example of a flowchart image and the character recognition result are shown in Figures 1 and 2 respectively. Here, “\n” indicates a line break.

Figure 1: Example of Flowchart Image Included in a Patent

1 Figures 2 and 3 show examples of character recognition results and a manually written explanatory text (patent abstract). In this case, the similarity between them is so high that we use them as machine learning data.
半導体装置の製造方法は、半導体層の上にフォトレジストを塗布する工程と；第1波長の紫外線を用いてフォトレジストを露光した後にフォトレジストを現像することによって、開口部の内側へと迫り出したネガ型レジストパターンを、形成する工程と；第1波長より短い第2波長の紫外線をネガ型レジストパターンに照射することによって、ネガ型レジストパターンを硬化させる照射工程と；照射工程を行った後、ネガ型レジストパターンの開口部から露出する半導体層の上に、ニッケル（Ni）から主に成る金属膜を形成する工程と；ネガ型レジストパターンを半導体層から除去する工程とを備える。

【翻訳】
The method of manufacturing a semiconductor device comprises the steps of: applying a photoresist onto a semiconductor layer; forming a negative resist pattern, which is pressed inwards into an aperture, by developing the photoresist after exposing the photoresist using a first wavelength of ultraviolet light; forming a negative resist pattern by irradiating the negative resist pattern with ultraviolet light of a second wavelength that is shorter than the first wavelength; and The negative resist pattern is hardened by irradiating the negative resist pattern with ultraviolet light of a second wavelength shorter than the first wavelength; and The negative resist pattern is hardened by pressing it into an aperture. By developing the photoresist after exposing it using a first wavelength of ultraviolet light; forming a negative resist pattern by irradiating the negative resist pattern with ultraviolet light of a second wavelength that is shorter than the first wavelength; and The negative resist pattern is hardened by pressing it into an aperture. By developing the photoresist after exposing it using a first wavelength of ultraviolet light; forming a negative resist pattern by irradiating the negative resist pattern with ultraviolet light of a second wavelength that is shorter than the first wavelength; and The negative resist pattern is hardened by pressing it into an aperture. By developing the photoresist after exposing it using a first wavelength of ultraviolet light; forming a negative resist pattern by irradiating the negative resist pattern with ultraviolet light of a second wavelength that is shorter than the first wavelength; and The negative resist pattern is hardened by pressing it into an aperture.
Results

The results of Recall, Precision, and F-measure for ROUGE-1, ROUGE-2, ROUGE-L, and BERTScore.

Table 2
Evaluation Results for the Generation of Explanatory Texts

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th>Precision</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROUGE-1</td>
<td>0.47</td>
<td>0.72</td>
<td>0.55</td>
</tr>
<tr>
<td>ROUGE-2</td>
<td>0.26</td>
<td>0.46</td>
<td>0.32</td>
</tr>
<tr>
<td>ROUGE-L</td>
<td>0.41</td>
<td>0.64</td>
<td>0.49</td>
</tr>
<tr>
<td>BERTScore</td>
<td>0.74</td>
<td>0.77</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Discussion

For simple geometries with no branches in the flowchart (see Figure 4), we obtained good analytical results. Figures 5 and 6 show the explanatory text and the patent summary (correct answer) generated by our method, respectively.

Figure 4: Example of Target Image for Generation

Figure 5: Exploratory Text Automatically Generated from the Image in Figure 1

Figure 6: Patent Summary for the Image in Figure 1 (Correct Answer)

Flowcharts with complex shapes, such as the one shown in Figure 7, tended to generate low-quality explanatory text. The dash line boxes in the figure were added by the author for the purpose of explanation. The description generated by the process in Figure 7 is shown in Figure 8.

Figure 7: Example of a Flowchart with Conditional Branching
Consider the coordinate information of the necessary to perform preprocessing such as input to T5 as they are, but in the future, it will be branch is “When the predetermined time elapses (yes figure are not considered at all. The first conditional because this time the coordinates of each string in the images, it was found that high-quality explanatory text could be generated, although some issues remain for flowcharts with complex shapes. In the future, we will examine the possibility of generating appropriate explanatory text for flowcharts with complex shapes, such as those containing multiple conditional branches, by considering the positional information of each character string in the image, rather than using the character strings in the flowchart as is.

5. Conclusions

In this study, 11,188 flowchart image-description pairs were obtained from patents and these data were used to construct a system that automatically generates descriptions of flowchart images using T5. The experimental results showed that for the detection of flowchart images, an accuracy of 0.9645 was achieved with a fine-tuned model using DenseNet121. In the generation of explanatory text from flowchart images, it was found that high-quality explanatory text could be generated, although some issues remain for flowcharts with complex shapes. In the future, we will examine the possibility of generating appropriate explanatory text for flowcharts with complex shapes, such as those containing multiple conditional branches, by considering the positional information of each character string in the image, rather than using the character strings in the flowchart as is.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP22K12154 and JP20H04210.

References