Deciphering the code of the novel "The Dancing Men"

Katsuhiko MURAKAMI
Fujitsu Ltd,
Kanagawa, Japan
murakami.ktk@gmail.com

Aoi KISHIDA
Kobe City Nishi-Kobe Medical Center,
Kobe, Japan
aoi.kishida@gmail.com

Hibiki ITO
University of Helsinki,
Helsinki, Finland
2001hibiki15@gmail.com

Sayaka MATSUMOTO
Office of Strategy and Planning,
Tokyo Institute of Technology,
Tokyo, Japan
syma@irds.titech.ac.jp

Kunihiko TAKAMATSU
Office of Strategy and Planning,
Tokyo Institute of Technology,
Tokyo, Japan
ktakamatu@irds.titech.ac.jp

Abstract-. This study examines a method for deciphering the handwritten dancing men graphic shown in "The Dancing Men." We attempted to reproduce and explain Holmes's reasoning for the graphic cipher by using a mathematical programming approach. Moreover, we studied the validity of Holmes's reasoning and demonstrated a procedure to decipher the cipher computationally. Herein, we propose a mathematical solution to clarify the ambiguity in Holmes' conjecture. Although the proposed method has not been implemented, its feasibility is confirmed through partial data creation and analysis.

Keywords-integer linear programming, image processing, simulation, Wasserstein distance

I. INTRODUCTION

The objective of this challenge (workshop) s to develop explainable artificial intelligence (XAI), that is, an AI system that answers questions with explanations. The task here was to infer the truth of a case (reproducing Holmes's reasoning), while providing a reasonable explanation based on the relevant open knowledge graph. Our analysis focused on the decipherment of the handwritten graphic cipher in the novel "The Dancing Men" by Arthur Conan Doyle. We did not use knowledge graphs in this study. Instead, we considered a method to reproduce and explain Holmes's reasoning by a computational and statistical approach.

In "The Dancing Men," six encoded messages are presented in sequence, as shown in Figure 1. The correct alphabet is shown below the images. One image code corresponds to one character, which is called a "substitution cipher" and is a premise within the novel. As the sixth sentence is not used for Holmes's reasoning, it is better to decipher without considering it. Holmes gave several reasons (hypotheses) for the plausibility of the deduction he reached: first, "the frequency of characters is similar in the cipher and in general writing;" second, "names of interested parties are likely to be used." These hypotheses seem to be effective because of their high objectivity.

First, we question the extent to which the character frequency information can be used for decoding.

To establish the frequency of a character or image, it is necessary to determine whether similar images indicate the same characters. However, as the images of the dancing men are handwritten and Holmes's position is that he does not know the identification rules, it is unknown whether two images are the same just because they are similar; in a sense, all the images for different characters are similar. In the novel, this is considered self-evident. However, it is unknown to the machine. Therefore, the second item to consider is how to determine whether the pattern of the dancing men in the graphic cipher is the same character.

I. METHOD

To test the inference by character frequency, we first obtained the frequencies of 26 characters of the general text alphabet. The text data were obtained from Wikipedia (approximately 11 billion characters), then the probability of E being the character with the highest frequency in the string in question and the probability of being another character were calculated by running 1,000 simulations.

Next, to examine the similarity of the images in terms of representing the same character, the features of the images were examined by the human eye. The assumption was that if one can find rules to identify characters using the features, the judgement for identity and explanation is valid.

Fig. 1 Graphic code (image data from [1])

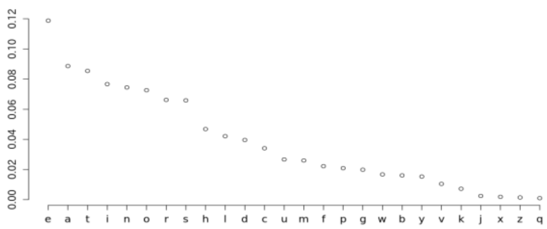

Fig. 2 Frequencies of alphabet in Wikipedia (11,055,558,613 characters) by our analysis: x-axis is alphabet, and y-axis is frequency.

In addition, when determining and explaining that a certain image is a certain character, a characteristic marker and its rule are necessary. Therefore, the items that can be used as landmarks for each image are prepared as data in table format.

II. RESULTS AND DISCUSSION

Of the 26 characters, the most frequent were, as in the novel, e, followed by a, t, i, and n, in that order. The simulation results of the generation of a 15-character string based on this showed that e ranked highest (appeared the most frequently) in only 50% of cases, and a and t ranked highest in more than 40% of cases: there were many cases in which multiple characters were ranked highest, that is, tied. Therefore, Holmes' method (assumption) where the character with the highest number of occurrences is probably e, is only true in half of the cases. Therefore, other possibilities should be explored for a good general image cipher deciphering method.

Next, we examined the criteria for the different characters in the images. Figures 2(a) and 2(b) are considered to be the same, although some of their features are different. In decoding the images, Holmes inferred that the flag is a word separator, although this is a self-serving assumption we believe is not convincing. Figure 2(c) is a different image. As the correct interpretation allows one to ignore the flag in images, it is unsurprising that we consider that there is no difference between the two images, resulting in the code being interpreted as the same. As mentioned above, it is difficult to determine whether these graphic images should be considered identical or distinct. To address this issue, we constructed a feature table of the dancing men images. For the right and left arms, we coded three features: presence/absence, bend/extension (if present), and up/down (if curved); for the right and left legs, we constructed two features—hip bend/extension and knee bend/extension; for body orientation, one feature—up/down—for the flag, and two features—presence/absence and right/left. A total of 13 features were assigned to all 76 characters, all of which could be coded with binary values. If all these possibilities are explored then decoded, it will be possible to correctly determine and explain which features are used to produce the same character.

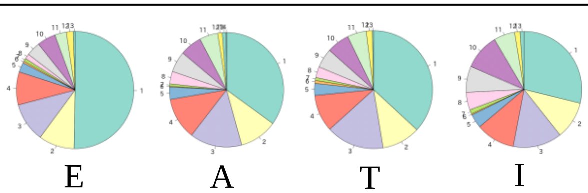

Fig. 3 Statistics of ranks of high frequency by the characters.

Fig. 4 Three pairs of similar images. Are these identical characters?

An alternative solution to determine the identity of images is to calculate the similarity measure between each pair of images, and use the similarity measure in reasoning. This type of approach does not require the manual discrete annotation described above and has the advantage that there are no arbitrary factors in terms of the presence or absence of features, except for threshold values and parameters of the similarity measures. The similarity measures applicable here require 1) robustness for rotation, scales, and thickness of lines, and 2) flexibility to determine rules to identitcal images depending on the variety of all images. We consider that Wasserman distance might be more suitable considering the above requirements.

Although we can find a correspondence between an image and a character with the above conditions alone, we have not yet verified whether it is a meaningful string or not. Therefore, we should make a relatively small dictionary (a list of candidate words). We will use a method to check highly possible strings in the dictionary.

Based on the above, we propose a decoding method using an integer optimization problem. The score function to be maximized is

$$f(x) = M + D$$

where M is the agreement score based on the correspondence between image features (or similarity) and characters, and D is the word score based on the used word in the dictionary. The agreement score consists of binary variable $m_{i,j}$, which indicates that feature j is used by image i (26 characters and 76 images), and binary variable $u_{k,l}$, which indicates that image i corresponds to character j. By maximization of $f(x)$, the solution leads to the deciphering and good explanation.

In this proposal, we outline a procedure for computationally decrypting a cipher. The part of the mathematical procedure for solving the cryptographic problem is an existing optimization method. We have defined and mapped the cryptanalysis problem solved by Holmes in the novel to a mathematical problem that is as general as possible. The features of the method are as follows: the calculation is squeezed by the dictionary to achieve a low amount, context consideration is included in the score of likely words, and the method can answer the question "What are the partial features to identify the dancing men character?” Theoretically, Holmes' estimated answer should be obtained as a somewhat high-scoring solution. At present, we have not yet implemented the system. We would like to continue working on it and verify its applicability through experimentation

REFERENCES

[1] The Arthur Conan Doyle Encyclopedia, "The Adventure of the Dancing Men," 1903. https://www.arthur-conan-doyle.com/index.php?title=The_Adventure_of_the_Dancing_Men (accessed Nov. 17, 2022).