An Intelligent Tool for Selecting Abstracts in MEDLINE

Miyako Tanaka [1]   **Sanae Nakazono** [2]

Hiroshi Matsuno [2]   **Satoru Miyano** [3]

miyako@ube-k.ac.jp   b1835@sty.sv.cc.yamaguchi-u.ac.jp

matsuno@sci.yamaguchi-u.ac.jp   miyano@ims.u-tokyo.ac.jp

[1]   Department of Business Adminstration, Ube National College of Technology, 2557, Tokiwadai, Ube 755-8555, Japan

[2]   Faculty of Science, Yamaguchi University, 1677-1 Yoshida, Yamaguchi 753-8512, Japan

[3]   Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minatoku 108-8639, Japan

1   Introduction

Without intensive reading of abstracts by experts, it is almost impossible to decide if a given MEDLINE record is a target article or not. However, there may be a way to reduce the hard task of experts in selecting right articles from MEDLINE. For this purpose, we have developed an intelligent tool for selecting MEDLINE abstracts whose mechanism is based on the iterative method proposed in the paper [3]. This paper reports that, with this intelligent system, 90% of target abstracts can be selected while leaving half amount of abstracts unread with the assistance of the machine learning system BONSAI [1].

2   System Overview

Fig. 1 shows the overview of the system working behind the tool. The system searches MEDLINE and gets a set A of abstracts which contain the keywords previously inputted by the expert. Moreover, the experts select more specific set S of abstracts from the set A. Then, the experts read the abstracts in the set S carefully in order to divide the set S into two sets POS and NEG and reflect their knowledge by assigning a character to each of words in the set S (this task is called "rough reading"[3]). The converted abstract file *Converted K* is going to be divided into two files *POS candidate* and *NEG candidate* by applying the rule found by BONSAI from the two files, *Converted POS* and *Converted NEG*. The tool lets the file *POS candidate* be the file S and lets the file *NEG candidate* be the file K and repeats the actions above until the aim of experts is satisfied. Fig. 2 is a screen shot of the tool.

One of our future work is to develop the tool as shown in [2] which can explicitly indicate sentences in the abstract in which experts will be interested.

References

[1] Shimozono, S., Shinohara, A., Shinohara, T., Miyano, S., Kuhara, S., and Arikawa, S., Knowledge acquisition from amino acid sequence by machine learning system BONSAI, *Trans. Inf. Process. Soc. Japan*, 35:2009–2018, 1994.

[2] Tanaka, M., Usuzaka, S., Matsuno, H., and Miyano, S., Gene relation extracting in literature database assisted by agents, *IPSJ Symposium Series*, 98(8):355–360, 1998.
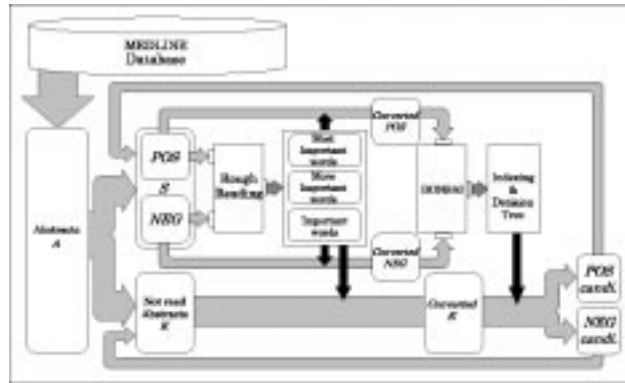
Figure 1: System overview.

Figure 2: A screen shot of the tool.

[3] Usuzaka, S., Sim, K.L., Tanaka, M., Matsuno, H., and Miyano, S., A machine learning approach to reducing the work of experts in article selection from database: a case study for regulatory relations of *S. cerevisiae* genes in MEDLINE, *Genome Informatics 1998*, 91–101, Universal Academy Press, 1998.