Expert System for the Identification of Korean Butterflies

Hwang-Yong Kim and Kun-Suk Woo
School of Agricultural Biotechnology, Seoul National University

How do you identify insect specimen? If you have to identify any insect, you would attempt to use dichotomous diagnostic key. Have you ever used it conveniently?

The identification keys of this type is not only hard to construct but also hard to understand. It maybe contains correct information needed for identification, and there is no conflict in it. However almost all users do not familiar with the terminology in it. And they cannot help removing the frequent hesitation.

Expert system is the imitation of the process of experts. Now, we have to consider one important fact that the experienced is not dependent on the dichotomous key. When Experts do identify a specimen, they can limit the result promptly by size, color, shape and the peculiar character of specimen. And then they usually check illustrations or photographs of candidates in detail. Comparison of the description of species is the final step. So, the process of experts is differ from that of the identification key.

Expert systems are easy to use and easy to transfer. In addition, sooner or later, it could be applied as remote identification, and some parts of expert system will be automated by proper machine sensor.

In this research, we develop the expert system for the identification of Korean butterflies. Domain experts are the personal experiences and the public knowledge printed in articles and books. OOP(Object Oriented Programming) by C++ is adopted instead of AI language(for example CLIPS, LISP).

Knowledge represented as production rule-based type with certainty factor, such as 『if(condition) → then(act)』 The examples of rule are as follows.

Rule 1 : If the colour of wing is white then the result is *Parnassius stubbendorfii* (Ménétriès) 모시나비 (CF = 0.4)

Rule 2 : If the colour of wing is white then the result is *Papilio bianor* Cramer 제비나비 (CF = -0.5)

Negative certainty factor means that the rule is not true.

Because we forbid the connectives, total number of rules are reached to several thousands. So Rete Pattern Matching Algorithm was chosen as the control strategy of rules. And combine the certainty factors with following methods.

both positive : $CF_1 + CF_2 (1 - CF_1)$
both negative : $CF_1 + CF_2 (1 - CF_1)$
otherwise : $(CF_1 + CF_2) / [1-\min (|CF_1|, |CF_2|)]$

For the inference engine, backward chaining was used. In working memory facts with certainty factor are exist. The assessment of this expert system in practice is being prepared.