DESIGN OF GEOGRAPHICAL INFERENCE ENGINE

Zhou Zhangfa
(Doctor Candicate, Peking University, China)
(ISPRS Commission III)

ABSTRACT

Geographical knowledge base and geographical inference engine are the two most important identifiers of Geographical Expert System (GES) . We have designed a geographical inference engine for Micro-Computer Geographical Expert System (MCGES),based on the methods of knowledge representation used in MCGES. As MCGES is a production system, the MCGES inference engine is based on IF-THEN inference.Fuzzy logic and grey system theory are used to imitate geographical inference in inference engine of MCGES. This paper concerns the design and principles of geographical inference engine.

KEY WORDS : Expert System, Geographical Information System, Inference Engine, Knowledge Base.

1. THE FUNCTIONS AND IMPORTANCE OF GEOGRAPHICAL INFERENCE ENGINE IN GEOGRAPHICAL EXPERT SYSTEM

As the combination of Artificial Intelligence (AI) and Geographical Information System (GIS) , Geographical Expert System (GES) is composed of three parts, that is,geographical data base, Geographical Knowledge Base (GKB) and Geographical Inference Engine (GIE) . GES is distinguished from GIS by its reasonning with geographical knowledge, and GIE and GKB are the most important identifiers of GES.

Inference engine,which is the logical core of infercnce procedure of expert system,controls the inference operation of the knowledge in the knowledge base on the data in the database and gains the conclusion, and is a kind of strategy program composed of inference controling algorithms and knowledge searching algorithms.

We have designed a GES on Micro-Computer (MCGES) .In MCGES,Micro-Computer Geographical Information System (MCGIS) is used as geographical data base, a geographical rule base is used as GKB,and GIE of MCGES connects MCGIS and GKB. By using knowledge in GKB to inference on data in MCGIS, GIE of MCGES can gain solution on the geographical problems without mathematical model and provide real time explanation for reasonning processes and inference results.There are two main function moudles in MCGES-GIE, one is reasonning module, the other is explaining module (refer to fig.1).

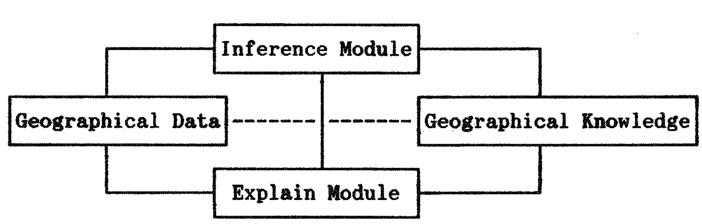

Fig. 1 The structure of MCGES-GIE

Most of geographical problems are inexact problems, which can not be or are very difficult to be expressed as mathematical models. Geographers use fuzzy reasonning to solve such geographical problems. In order to imitate geographers solveing inexact geographical problems, a kind of geographical knowledge representation method is presented, and fuzzy logic and grey system theory (Deng Julong, 1982) are adopted as reasonning model in MCGES-GIE.

2. KNOWLEDGE REPRESENTATION METHODS USED IN MCGES-GIE

As a branch of AI applications, expert system uses knowlege as logical kernel. Geographical knowledge is the logical association of geographical data, for example, rules of geographical planning, expressions of geographical phenomena.Therefore,geographical knowledge must be acquired from geographers. The acquisition and representation of geographical knowledge is an important factor to determine the level of GES.

Geographical knowlwdge representation may be defined as expressing geographical knowledge as abstract logical form that can be accepted by computer with logical analyses.

After analysising geographical research objects, we divide geographical knowledge into following three levels:

(1) The basic level, which describes and grades the geographical independent factors, is based on Geo-Code Model (GCM) (Ma Ainai, 1988). For example, in Kouhe Soil and Water Conservation Expert System, precipition, soil depth, soil type, vegetation coverage, slope etc. are selected as the basic factors, the rules to grade them are the basic knowledge.

(2) The medium level, which classifies integrative geographical factors, is a group of classification rules. Because the change and distribution of geographical factors are

successive, the medium rules always use fuzzy methods to classify them. Here is a sample example, from grassland to forest, tree gradualy substitute grass, we can not find a absolute boundary to delimite grass and forset, so in order to classify them, fuzzy mathematics are used to build fuzzy classification model.

(3) The advanced level is a series of integrative geographical planning rules, always is a set of experience of geographer.

MCGES expresses geographical knowledge with the form of production-rule. The BNF defination of MCGES knowledge is as following:

(1) The basic level

```
<factor>::=<factor name><factor grading>

<factor grading>::=<factor grade>
                   <minimum of grade>
                   <maximum of grade>

<factor grade>::=0 | 1 | 2 | ...
```

(2) The medium and advanced level

```
<rule>::=IF <evidence> THEN <conclusion>
         <effectiveness>

<effectiveness>::=
     <conclusion prior probability>
    {<evidence probability
               if conclusion exist>
     <evidence probability
               if conclusion not exist>}

<evidence>::=<condition | condition> AND
             <condition | condition>

<condition>::=<factor name><factor grade>

<conclusion>::=<temporary conclusion> |
               <final conclusion>

<temporary conclusion>::=<geographical type>
                         <geographical value>

<final conclusion>::=<geographical process>

<geographical type>::=geographical type1 |
                      geographical type2 | ...

<geographical value>::=geographical value1 |
                       geographical value2 | ...

<geographical process>::=
                 geographical process1 |
                 geographical process2 | ...
```

Here is an example of MCGES basic knowledge,

```
#factor-1: /* from MCGES for Kouhe Soil&Water
              Conservation Expert System */
{
    #name: slope; /* slope grading */
    1:     0,2  ;  /* grade 1 is 0 - 2 */
    2:     2,5  ;  /* grade 2 is 2 - 5 */
    3:     5,8  ;  /* grade 3 is 5 - 8 */
    4:     8,15 ;  /* grade 4 is 8 -15 */
    5:     15,25;  /* grade 5 is 15-25 */
    6:     25,- ;  /* grade 6,if >25 */
}
```

The effectiveness of rules points out the corelation between evidence and conclusion. P(C), P (E¦C) and P(E¦~C) are used to express conclusion prior probability, evidence probability if the conclusion exists and evidence probability if the conclusion does not exist.

Examples of rule are shown as following,

```
#rule-17:  /* from MCGES for Kouhe Soil&Water
              Conservation Expert System */
{
    #if: slope == 3,
         soil_depth == 3,
         erosion_type == water_erosion,
         erosion_density == 2,
         landuse == 1;
    /* landuse type is cultivated land */

    #then:   terrace ;
    /* Using terrace to conserve S&W */

    #effect: 0.22,
    /* terrace prior probability */

             0.40,0.45,
    /* slope : P(E ¦ C), p(E ¦ ~C) */

             0.65,0.70,
    /* soil_depth : P(E ¦ C), p(E ¦ ~C) */

             0.43,0.67,
    /* erosion_type : P(E ¦ C), p(E ¦ ~C) */

             0.21,0.74,
    /* erosion_density:P(E ¦ C),p(E ¦ ~C) */

             1.00,0.85;
    /* landuse : P(E ¦ C), p(E ¦ ~C) */

    #relative: rule-29,rule-36,rule-81,
               rule-90,rule-103;
    /* rules connected with #rule-17 */
}

#rule-81: /* from MCGES for Kouhe Soil&Water
             Conservation Expert System */
{
    #if: slope == 3,
         soil_depth == 3,
         erosion_type == water_erosion,
         erosion_density == 2,
         landuse == 8;
    /* uncultivated land */

    #then:   tree planting;
    /* planting tree to conserve S&W */

    #effect: 0.18,
    /* tree planting prior probability */

             0.15,0.35,
    /* slope : P(E¦C), p(E¦~C) */

             0.45,0.27,
    /* soil_depth : P(E¦C), p(E¦~C) */

             0.43,0.43,
    /* erosion_type : P(E¦C), p(E¦~C) */

             0.15,0.74,
    /* erosion_density : P(E¦C), p(E¦~C) */

             0.00,0.90;
    /* landuse : P(E¦C), p(E¦~C) */
```

```
    #relative: rule-36,rule-29,rule-17,
               rule-90,rule-103;
    /* rules connected with #rule-81 */
}
```

The RELATIVE of a rule is one or several other rules that have almost the same conditions, and is used to increase searching speed. This item may be got automatically by the system.

3. REASONNING METHODS USED IN MCGES-GIE

In accordance with the three levels of knowledge, a complete inference procedure of MCGES-GIE is composed of several operations in three levels.

3.1 Basic inference

We can use the knowledge in the first level to match values of factors got from MCGIS,and grade the factors. The results of basic inference are the classes of basic geographical factors of objects.

3.2 Medium inference

There are two goals in this step, one is to calculate attribute values of integrated geographical factors, another is to classify the integrated geographical factors based on the conclusions of the basic inference and the medium knowledge (rules).

Now we can get a great deal of remote sensing data,but much less survey data.Remote sensing data may be considered as a kind output of geographical phenomena, and those relatively less survey data may be considered as the really geographical features. Fuzzy mathematics is a suitable tool to distinguish successive varibles and can be conveniently used to build geographical classification model with remote sensing data.For example,if we have got MSS-4,5,6 and 7 image data of the same district, we can construct a fuzzy function to classify landuse types in this district according to the image data of four bands,

object set: $X=\{x \mid x^T=(x1,x2,x3,x4)\}$

where x1,x2,x3,and x4 represent grey values of MSS-4,5,6,and 7;

fuzzy classification model is

$$\underset{\sim}{Ai}=\underset{\sim}{Ai1} \cap \underset{\sim}{Ai2} \cap \underset{\sim}{Ai3} \cap \underset{\sim}{Ai4} \quad (1)$$

where $\underset{\sim}{Aij}$ is normal fuzzy set,

jurisdiction function of Ai on X is

$$Ai(x)= \min_{1<=j<=4} \underset{\sim}{Aij}(xj) = \exp\left[- \max_{1<=j<=4} \left(\frac{xj-aij}{bij}\right)^2\right] \quad (2)$$

where $aij=Exij$, $bij=\sqrt{Dxij}$;
i=1, 2,..., n are type code;
j=1, 2, 3, 4 are band code.

then we have following classification formula,

$$a. \quad \underset{\sim}{A}_k(x) = \max_{1<=i<=4} \min_{1<=j<=4} \underset{\sim}{Aij}(xj)= \max_{1<=i<=n} \left\{ \exp\left[- \max_{1<=j<=4} \left(\frac{xij-aij}{bij}\right)^2\right]\right\} \quad (3)$$

$$b. \quad (\underset{\sim}{Ai},\underset{\sim}{B})= \bigwedge_{j=1}^{4} (\underset{\sim}{Aij},\underset{\sim}{Bj})= \min_{1<=j<=4} (Aij,Bj) \quad (4)$$

if let $ej=Exj$, $dj=\sqrt{Dx}$, $xj \in \underset{\sim}{Bj}$,

and $\underset{\sim}{B}= \bigcap_{j=1}^{4} \underset{\sim}{Bj}$ is the fuzzy subset that need be classified, then we have,

$$(\underset{\sim}{Ak},\underset{\sim}{B})= \max_{1<=j<=4} (\underset{\sim}{Ai},\underset{\sim}{B}) = 0.5*\left\{1+\exp\left[- \min_{1<=i<=4} \max_{1<=j<=4} \left(\frac{aij-ej}{bij+dj}\right)^2\right]\right\} \quad (5)$$

and $\underset{\sim}{B}$ can be considered as type k.

Grey system theory can forcast the developing trendency of varibles with relatively less known condition, and can describe the unbalanced relationship between main varible and subordinated varibles. Helped by grey system theory, we can build geographical grey model based on remote sensing data and survey data.A general Grey Model (GM)may be shown as GM(n, h) , which is a n factorial, h varibles differential equation,its expression is,

$$\frac{d^n x_1^{(1)}}{dt^n} + a_1 \frac{d^{n-1} x_1^{(1)}}{dt^{n-1}} + \cdots + a_n x_1^{(1)} = b_1 x_2^{(1)} + b_2 x_3^{(1)} + \cdots + b_{h-1} x_h^{(1)} \quad (6)$$

After getting integrated geographical factor values,using medium knowledge and conclusions of basic inference to match with them, we can gain the geographical classification results, which will be used in advanced inference.

3.3 Advanced inference

Based on the conclusions of the basic and medium inference and the advanced knowledge (rules) ,we can gain the decision measures or divisions and planning scheme on a certain geographical problem, which is the last step in the inference.

Because most of geographical inference are shown as under certain conditions to gain certain results,in this step,MCGES-GIE adopts following production strategy in reasonning:

```
RULE :           if A then B
PREMISE :        A is true
CONCLUSION :     B is true
EFFECTIVENESS:   possibility
```

The advanced inference uses Bayes theorem and fuzzy logic to reason. If we record the probability of conclusion with evidence existing as P(C¦E) ,then Bayes theorem may be show as,

$$P(C|E)=P(E|C)*P(C)/(P(E|C)*P(C) + P(E|\sim C)*(1-P(C) \qquad (7)$$

and

$$P(C|\sim E)=(1-P(E|C))*P(C)/((1-P(E|C))*P(C) + (1-P(E|\sim C))*(1-P(C) \qquad (8)$$

If evidence E expressed as E1 AND E2 AND ... AND En, then

$$P(E|C)=\min\{P(Ei|S)\},\ i=0,\ 1,\ 2,\ \ldots,\ n \qquad (9)$$

The advanced inference of MCGES-GIE uses the algorithm of backward chaining, which is shown as following:

bkwdchain(knowledge_base: KB, goal: C)
[BEGIN]

(1) scan KB to find rules with C as conclusion, set NOTUSED tag to those rules and put them into rule set IRSET;

(2) calculate P(C¦E) of all rules in IRSET, and sequence them according to their P(C¦E);

(3) select the most important rule with NOTUSED tag from IRSET, choose the evidence with the biggest P(E¦C)as new goal C1;

(4) [IF] C1 is unknow
    [THEN] manage to calculate C1 or call bkwdchain(KB,C1);
    [ELSE]
    a. calcute the probability of C if all other evidence without C1 support it, and record it as Pmax;
    b. calcute the probability of C if all other evidence without C1 do not support it, and record it as Pmin;

(5) [IF] Pmin <= maxium of threshold value, Pmax >= minium of threshold value
    [THEN]
    a. change the NOTUSED tag to HASUSED in current rule;
    b. GOTO (3);

(6) [IF] Pmin > maxium of threshold value
    [THEN] C is true, exit;

    [IF] Pmax < minium of threshold value
    [THEN] C is not true, exit;

[END]

The three level inference make up a reasonning net, which can be illustrated as figure 2.

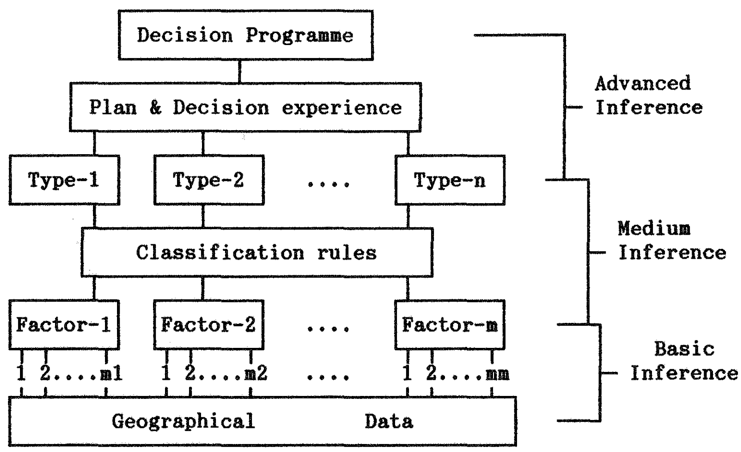

Fig. 2 The Reasonning Net of MCGES-GIE

4. THE EXPLAINING MOTHODS OF MCGES-GIE

The explaining module is another important part of MCGES-GIE. In the inference procedure, MCGES-GIE allocate a buffer to store all of the reasonning nodes, reasonning direction and temoprary results. With user's different requirement, MCGES-GIE presents three level explanation based on those information,

(1) explanation for inference results, including what the results mean, how the system to get them, which rules are involved;

(2) explanation for rules involved in the inference procedure;

(3) explanation for all geographical factors.

5. CONCLUSIONS

MCGES-GIE can iminate the inexact inference of geographers to solve a great deal of inexact geographical problems, and can conveniently present explanation in different levels to meet the user's requirements. MCGES-GIE has been successfully used in Kouhe Soil and Water Conservation Expert System.

REFERENCE

[1] Deng Julong, 1982, "Control Problems of Grey System", Systems & Control Letters, Vol.1, No.5.

[2] Ma Ainai, 1988, "A Geo-code Model for the Use of GIS", The 16th ISPRS, Kyoto, 1988, Vol.27, B4, pp.585-591.