ACQUIRING APPROXIMATE REPRESENTATIONS OF SOME SPATIAL RELATIONS

Vincent B. Robinson
(Goss.Ensuadmin@UNCA-MULTICS.MAILNET)
Department of Surveying Engineering
The University of Calgary
2500 University Drive, NW
Calgary, Alberta T2N 1N4
CANADA

Richard Wong
(Rnwhc@Cunyvm.BITNET)
Department of Computer Science
Hunter College - CUNY
695 Park Avenue
New York, NY 10021
USA

ABSTRACT

This paper reports on development of a Spatial Relations Acquisition Station (SRAS) and preliminary results of man-machine interactions. SRAS is based on a process for acquiring linguistic concepts through mixed-initiative man-machine interactions. It is unique by virtue of acquiring fuzzy representations while requiring only "yes/no" answers from the user. It is shown how SRAS can be used to acquire multiperson concepts for use in an organizational context. Results show significant interuser and interterm variation, and suggest that the size of the spatial database may not influence extent of interuser semantic variation. Results of multiperson concept formation show the importance of understanding the process.

INTRODUCTION

Robinson and Frank (1985) and Robinson et al (1985b) have identified several major areas where an understanding of NL concepts can contribute to our understanding of the nature and influence of uncertainty in geographic information processing. Like Robinson (1986), we are concerned primarily with the representation of NL concepts for use in the retrieval of geographic information from geographic data bases.

There have been several attempts at formulating spatial query languages. A number of spatial information query languages have been developed that are similar to Chamberlin and Boyce's (1974) SEQUEL (e.g., Frank, 1982; Barrera and Buchmann, 1981). They have much in common with other systems like the Map Analysis Package (Tomlin and Tomlin, 1981) that uses a subset of the English language to pass commands to a spatial information system. The meaning of each command must be unambiguous. Thus, the system cannot exploit the vagueness inherent in a natural language expression. One of the motivations behind development of a SEQUEL-like query language was the general lack of knowledge concerning retrieval of spatial information using natural language concepts (Frank 1982).

This paper reports on an effort to develop representations of NL concepts that preserve their approximate nature. Presented here is an algorithm for acquiring linguistic concepts through mixed-initiative, man-machine interactions operating on a spatial data base. Finally, selected results of man-machine interactions are presented and discussed.

REPRESENTATION FRAMEWORK

It is within the context of this framework that the approximation of linguistic entities is developed. Emphasis is placed on PRUF and test-score semantics, both of which have a basis in fuzzy set and possibility theory (Zadeh, 1978, 1981). PRUF and test-score semantics provide a general meaning representation and translation framework for development of linguistic approximations of spatial concepts for retrieval of spatial information from a spatial data base.

Briefly, Possibilistic Relational Uniform Fuzzy (PRUF) meaning representation language is based on the assumption that the imprecision intrinsic in natural languages is possibilistic in nature. Hence the logic underlying PRUF is a Fuzzy Logic. In PRUF a relational database is a collection of fuzzy relations which may be characterized in various ways by tables, predicates, recognition algorithms, generation algorithms, etc. Since an expression in PRUF is a procedure, it involves, only the frames not the relations in the database.

The semantics underlying PRUF are test-score semantics (TSS). The basic idea underlying TSS is that an entity in linguistic discourse has the effect of inducing elastic constraints on a set of objects or relations. The meaning of the entity may be defined by (a) identifying the constraints that are induced by the entity; (b) describing the tests that must be performed to ascertain the degree to which each constraint is satisfied; and (c) specifying the manner in which the partial test scores are to be aggregated to yield an overall test score.

Zadeh (1981) contends that the meaning of a linguistic entity in a natural language may be identified by testing elastic constraints that are implicit or explicit in the entity in question. The testing of constraints can be accomplished using tools afforded by test-score semantics and fuzzy logic can be used to assess the compatibility of a linguistic summary with a given database. The process of meaning representation in test-score semantics involves three distinct phases - (1) an explanatory database frame or EDF is constructed; (2) a test procedure is constructed which acts on relations in the explanatory data base (EDB) and yields test scores which represent the degree to which elastic constraints induced by the constituents of the semantic entity are satisfied; and (3) the partial test scores are aggregated into an overall test score that is a vector serving as measure of compatibility of the semantic entity with the EDB.

In PRUF, the translation of a proposition may be either focused or unfocused. The focused translation generally leads to a possibility assignment equation. An unfocused translation based on TSS is a collection of tests that are performed on a database induced by the proposition; and a set of rules for aggregating the partial test scores into an overall test score that represents the compatibility of the given proposition with the database. Robinson (1986a) provides examples of these translations and their relation to this problem.

ACQUISITION OF LINGUISTIC APPROXIMATIONS

Jain (1980), Hersh et al. (1979), Leung (1982), and Lundberg (1982) have all bemoaned the lack of a clear methodology for determining compatibility functions. One of the problems has been the inability to discriminate between measuring the ability of subjects to use fuzzy logic as opposed to specifying the membership functions (e.g., Lundberg, 1982). This assumption underlies Yager's (1982) document retrieval system based on fuzzy logic and places an impossible cognitive load upon the user. The methodology outlined here reduces cognitive load while capturing the vagueness inherent in natural language concepts.

Using a mixed-initiative methodology similar to that suggested by Nakamura and Iwai(1982), compatibility functions are acquired by the Spatial Relations Acquisition Station (SRAS) (Robinson, 1984, 1986c). The process described below is designed for the problem of determining the meaning of a spatial relation such as NEAR using some base variable such as distance. It is composed of four major components. First, the process is initialized. Second, a question is chosen according to some criteria. Third, response to a question is used to infer adjustments to the representation of the linguistic variable. Finally, before repeating the second and third steps above a decision is made as to whether or not the question-answer process should continue.

In this question-answering scheme it is assumed that XA is the computer's universe of discourse on a base variable and XB the user's universe of discourse. The concept to be learned is denoted as C. The computer learns C through a process of question and answer (QA) by constructing fuzzy set FS, the learned concept, which is a replica of C in XA. The computer selects a question unit (QU) out of units of XA based on a selection criteria and asks the user whether x[i] belongs to C ($x[i] \in C$) or not ($x \notin C$). The user answers YES or NO.

Initialization

This process begins from a position of maximum uncertainty. This is tantamount to the machine possessing no preconception. We use the definition of maximum uncertainty as derived from the fuzzy information-theoretic measure described by De Luca and Termini (1972). In essence, each tuple in the relation receives a compatibility score of 0.5. Since this measure plays another important role in the process, it is described later in more detail.

Concept Formation Process

Let FS[k-1] denote the learned concept of the computer just before the k-th QA step. When reply of the user to the k-th question of whether x[i] belongs to set C is YES, the computer constructs FS(x) in its knowledge space XA. As a result of the k-th QA step, learned concept C[k-1] is changed to C[k] given by

$$C_k = C_{k-1} \cup FS(x). \tag{1}$$

When the reply is NO, the computer constructs nFS(x) in knowledge space XA and constructs C[k] from C[k-1] using nFS(x) (1-FS(x)) where

$$C_k = C_{k-1} \cap nFS(x). \tag{2}$$

As an initial approximation, FS(x) and nFS(x) are assumed to be defined as

$$FS(x) = \exp(-\alpha\, d_{ik}) \quad \text{where} \quad \alpha > 0, \quad \text{and} \tag{3}$$

$$nFS(x) = 1 - FS(x) \tag{4}$$

where d[ik] is the distance of x[i] from x[k].

In the above specification the parameter α determines the spread of both functions. The parameter α is adaptive. Initially

$$\alpha = \ln(0.5) / \max(d_{ik}) \tag{5}$$

which means that those locations farthest from location k would be assigned a membership value of 0.5 and those in between will range from 0.5 through 1.0 depending upon the value of the base variable, distance.

During the acquisition process α is adaptively changed. First, α [k] is changed according to the following rules -

if the answer to x[k] is NO
then $\alpha[k] = \ln(0.2) / x[k]$ (6a)

if the answer to x[k] is YES then $\alpha[k] = \ln(0.85) / x[k]$. (6b)

These rules have the effect of drawing an analogy between the 'nearness' or 'not nearness' of x[k] to the key location and those locations similar distances from x[k]. In Eq. 6b, if x[k] is closer than the previous YES-related x[k] then the previous value is used. Thus, the spread of membership function (see Eq. 3) resulting from an affirmative response is never constricted.

To allow for previous answers let x[i] be the k-th QU and x[j] is a QU used before the k-th question-answering step. If x[j] exits in the neighborhood of x[i] in knowledge space XA, and the reply to x[i] is opposite to x[j] then it is supposed that the boundary of user's concept exists between x[i] and x[j]. Thus, the value of α is increased; that is FS(x) is made narrow so FS(x[j]) becomes below (or above) a prescribed value ξ (or $1-\xi$). Before the next QA step α is set to $\alpha[k]$.

<u>Selection of Questions</u>

Each place other than z is considered to be a candidate as a question unit. Candidates for the k-th question unit are limited to only the units whose grade of membership are above a prescribed small value (e.g. 0.1), which have not been used as the question units before, and have not been in close proximity to another question unit (ie receiving a membership value greater than or equal to 0.80).

It is desirable that QU's be selected so as to conform learned concept C[k] to the user's concept of interest C in knowledge space XB. In order to decide on the appropriate QU,a measure of the uncertainty of a fuzzy concept is used. An information-theoretic measure (De Luca and Termini, 1972) is used to measure the uncertainty of a fuzzy relation. It is defined as -

$$I_k = - \sum_{i=1}^{n} [(\mu_i \ln \mu_i) + ((1 - \mu_i) \ln (1 - \mu_i))]. \qquad (7)$$

This measure takes on the value of zero(0) if and only if $\mu[i] = 0$ or $\mu[i] = 1$ for all i. It is maximized when $\mu[i] = 0.5$ for all i. This latter condition occurs when the dominant truth value of any tuple cannot be distinguished.

Let Iyes(k) and Ino(k) be the measures of uncertainty of the fuzzy sets given by (1) and (2) respectively. Iyes(k) corresponds to the uncertainty of concept C[k] in the k-th QA step if the user replies YES. Ino(k) corresponds to the uncertainty associated with concept C[k] in the k-th QA step if the user replies NO. So, let

E(I(k)) be the expectation of I(k) which is given by -

$$E(I_k) = [\nu I_{yes(k)} + \omega I_{no(k)}] \qquad (8)$$

where ν and ω are weights that may be used to reflect the relative likelihood a response will be YES or NO. In this process

$$\nu = \exp[-2 \alpha d_{ik}] \quad \text{and} \quad \omega = 1.0 - \nu \qquad (9)$$

which has to effect of weighting the average expected uncertainty as function of the distance from the key location. This is more consistent with the manner users weight their expectations than the simple averaging as reported in Robinson et al (1985a, 1985b).

E(I[k]) is calculated only with respect to candidate units. Thus, the optimal question unit for the k-th QA step is that which

$$\text{maximizes} \quad | I_{k-1} - E(I_k) | . \qquad (10)$$

As a result of simulations this was found to provide better boundary finding behavior than that described in Robinson et al (1985a).

Stopping the Process

One of the major issues in specifying this process is that of when does the process stop. Here use is made of Kaufmann's index of fuzziness (K).

Index of Fuzziness. The index of fuzziness suggested by Kaufmann (1975) is defined as

$$K_k = \min_S (2/[XA\#]^{\frac{1}{2}}) \left[\sum_{x \in XA} (\mu_{C_k}(x_i) - \mu_S(x_i))^2 \right]^{\frac{1}{2}} \qquad (11)$$

where S is any ordinary subset in XA, XA# is the number of units in XA, μ[Ck](x[i]) = membership function of fuzzy set C[k] and μ[S](x[i]) is the characteristic function of ordinary set S. K[k] is a normalized distance in XA# dimensional space between C[k] and ordinary set SC[k] (ε {S}) nearest to C[k], and does not become over 1.

Now consider that d(x[i]) is the projection of an Euclidean Distance between C[k] and SC[k] into the i axis :

$$d(x_i) = \mu_{C_k}(x_i) - \mu_{SC_k}(x_i) . \quad (12)$$

The value of d(x[i]) is a measure of fuzziness of x[i] with respect to membership in C[k] . The definition of SC[k] -

$$\mu_{SC_k}(x_i) = 0 \text{ for } \mu_{C_k}(x_i) < 0.5 \quad \text{and} \quad (13a)$$

$$\mu_{SC_k}(x_i) = 1 \text{ for } \mu_{C_k}(x_i) \geq 0.5 \quad (13b)$$

Thus, K[k] indicates how strongly correlated the fuzzy subset representation is with a crisp, or regular, subset representation. The index gives us an indication of how closely the fuzzy concept fits a 'crisp', or nonfuzzy, representation. When K = 0 then there no longer exists a difference between the fuzzy concept and the crisp concept.

Stopping Rule. In this work it is assumed that the computer has accomplished learning the user's subject of interest C when the index of fuzziness of C[k] falls under a prescribed value, say K[ke]. The computer finishes when the K[k] becomes less than some specified proportion (ρ) of the maximum of the values in the previous steps. That is to say that the process stops at step k[e] where it is the step satisfying for the first time the following relation -

$$K_{k_e} < \rho [\max_{k=0,1,...,k_e} K_k]. \quad (14)$$

In the case of FAR relations the algorithm remains substantially the same only making use of the complement of the results. That is to say that when the user responds with a YES to the question "is city_z FAR from city_x" the process above treats it as a negative response. Upon modification of the concept the complement is used as the representation of FAR. This has the nice property of representing the complement of NEAR if the responses of the users are consistent with FAR being strictly a complement of NEAR.

MULTIPERSON CONCEPT FORMATION

The procedures used in SRAS essentially acquire a personal definition of a spatial relation. It has meaning with regard to the semantics of the single user. However, geographic information systems are typically used within organizations that arrive at definitions by committee. We suggest here that the concepts acquired by SRAS can subsequently be used in arriving at a consensual representation of a spatial relation. Furthermore, the process by which the concensual representation is arrived at can clearly and rigorously defined.

There are several approaches to constructing multiperson concepts. We draw upon the work of Gaglio et al (1985) and discuss four methods of constructing multiperson concepts from SRAS. They are the agreement, global evidence, combined agreement and global evidence, Zimmermann's (1983) and the weighted-mean method.

<u>Agreement Method</u>

Fuzzy intersection of μ[i]'s defined as

$$\mu(x_k) = \bigcap_i \mu_i(x_k) \qquad (15)$$

forms the basis of this method. This corresponds to a group decision procedure where decisionmakers have a sort of veto power. That is, the degree of acceptance assigned to each truth value is equal to the lowest among those assigned by the various committee members.

<u>Global Evidence Method</u>

In this method the "positive" opinions prevail because it is based on fuzzy union of μ[i]'s

$$\mu(x_k) = \bigcup_i \mu_i(x_k) \qquad (16)$$

Gaglio et al (1985) suggest that this method may be suitable when the procedure for obtaining multiperson concepts does not have a feature similar to veto power of some member.

<u>Combined Agreement and Global Evidence Methods</u>

There are several possible ways of combining the previous two methods. We discuss two that are particularly relevant to the kind of committees typical of the organizational context of geographic information systems.

<u>Method I</u>. Committee member i has veto power over decisions of others and is defined

$$\mu(x_k) = \mu_i(x_k) \cap \left(\bigcup_{j \neq 1} \mu_j(x_k) \right). \qquad (17)$$

<u>Method II</u>. The second method defined as

$$\mu(x_k) = \mu_i(x_k) \cup \left(\bigcap_{j \neq 1} \mu_j(x_k) \right) \qquad (18)$$

corresponds to the situation where committee member i only has "acceptance" power not "veto" power. Both (17) and (18) describe situations where there is asymmetric decisionmaking power among the committee members. Asymmetry in committee situations generally is a function of the type of chair a committee has. Therefore, we suggest that (17) is a reasonable method for modelling a "strongly chaired committee" while (18) can be used to model a "weakly chaired committee".

Zimmermann's Method

In the method suggested by Zimmermann (1983) we combine the agreement with the global evidence method using a "compensatory and" operator. This method may be defined as

$$\mu(x_k) = [\bigcup_i \mu_i(x_k)]^{1-\gamma} \quad [\bigcap_i \mu_i(x_k)]^{\gamma} \qquad (19)$$

where $0 \leq \gamma \leq 1$.

This method preserves symmetry among decisionmakers while striking a compromise between the global evidence and the agreement methods. The parameter γ determines the nature of that compromise. As the value of γ increases the greater the influence of the global evidence method increases.

Weighted-Mean Method

The weighted-mean method can defined by

$$\mu(x_k) = \sum_i w_i \mu_i(x_k) \qquad (20a)$$

where

$$\sum_i w_i = 1. \qquad (20b)$$

In this method the w[i]'s can be used to weight the importance of each committee member's concept. However, we see many problems with trying to formalize the specification of the w[i]'s, thus use of this method should be used only after some additional research has been conducted in this arena. In addition, use of some of the above methods is implicitly using some form of the weighted-mean method. We say this because giving some members of a committee, in effect, "veto" power represents in a very practical manner the assignment of importance not given other members.

MAN-MACHINE INTERACTIONS

In previous papers we presented results of simulated man-machine interactions (Robinson, 1984; Robinson et al, 1985a; Robinson, 1986a). The simulations illustrated the behavior of the process but did not provide an opportunity to investigate semantic variation among and between users of geographic information systems. Recently we presented results of a session with SRAS by a so-called expert (Robinson et al, 1986c) and showed that there was significant semantic variation within a single, expert user. In addition, it was shown explicitly that intransitivities exist in the definition of a simple concept by an expert.

Man-Machine Sessions

In this paper we will discuss premliminary results obtained from 5 subjects, one of which was the expert referred to above. There were a total of 16 sessions with each subject. Two geographic databases were used, each providing a distinctly different spatial context. One database contained 29 settlements drawn from the 1:250,000 USGS map sheet for Waycross, Georgia. The other database consisted of 112 settlement locations drawn from the 1:250,000 USGS map sheet for Hartford, Connecticut.

For each database there were a total of 8 sessions. Each session was concerned with acquiring a spatial relation expressed as one of terms in Table I. For example, in a session operating on the Hartford database the subject was asked, by SRAS -

Is Port_Ewen Close_to Waterbury ?

Each subject responded with "yes" or "no." It is important to realize that the subject did have the map sheet available for reference. Terms and databases were presented to the subject in a randomized order. Generally sessions were separated by 24 hours or more. Before starting the sessions the subject was not informed of the term set and the subject was not aware of what term was to be covered in the next session. The general problem remained the same, so the only variables were terms and database.

Results and Discussion

For each subject and session we can generate the question-answer tree such as one shown in Figure 1. Each of these trees tells us how many steps the session took, which settlements were used as question units, in what order they were asked, and the response of the subject. If a subject answered in exactly the same manner as another subject, the tree will be exactly the same for both subjects. Furthermore, for concepts that are compliments such as "near" and "far," exactly complimentary responses yield the same tree differing only in responses.

As a rough indication of the level of semantic variation we can look at variations in length of sessions without regard to similarity of pattern. Table 2 shows how often sessions of particular length occurred. What is surprising is the great range in session length. Of 80 sessions there were 17 different lengths of question-answer sessions. This implies a fair amount of variation in question-answer patterns.

Given the considerable difference in database size we might expect the variations in the question-answer trees to be greater with larger databases. Of the 40 sessions per database, there were 17 unique question-answer trees resulting from use of the Waycross, Georgia database and 18 unique trees from use of the Hartford, Connecticut database. These results suggest two things. One is that size of the database may not influence the overall semantic variation. The other is that regardless of database size, there is a roughly even chance that one user's question-answer tree will be the same as anothers. This variation becomes even more pronounced when we break the results down according to the terms in Table 1.

Tables 3 and 4 illustrate how often there is agreement by the subjects on the definition of a term. Even if there is agreement, there is never more than 3 of the 5 in agreement. Some interesting observations can be made regarding those terms on which there is no agreement. Using the Georgia database, each subject had different definitions for the terms Far, Distant_from, and Short_distance_from. There is no agreement on the definition of in_the_Vicinity_of, Remote_from, and Close_to when using the Connecticut database. It is apparent from these results that one can expect little agreement on the exact definition of simple spatial relations. This leads us naturally into the topic of multiperson concept construction.

Table 5 shows the results of sessions for five subjects regarding the specification of the term Close_to. This term was chosen as a subject of special attention because there is an intransitivity imbedded in the concept of Close_to for subject 1. Also, notice that subjects 2 and 3 agreed exactly on the specification of Close_to. Specification of Close_to by Subject 4 is the most liberal of the five subjects. Note the preponderance of membership values greater than 0.5 and absence of any membership values equal to 0.00. The resulting pattern of membership values for Subject 5 resemble the results for Subject 1, but contains some differences and lacks the aforementioned intransitivity.

Responses to the questions regarding Close_to Douglas lead to identification of an intransitivity. In the question/answer process of Subject 1, Nicholls received a no response and Pearson received a yes response. Pearson is farther from Douglas than is Nicholls, yet the subject said that Nicholls is not close while Pearson is close. Since the map sheet was available it is appropriate to consider whether transport routes and/or major landscape features may have influenced

Table 1. Term Set Used in Man-Machine Interaction Sessions

Nearness Terms	Farness Terms
Near	Far
in_the_Vicinity_of	Remote_from
Close_to	Distant_from
Short_distance_from	Long_distance_from

Table 2. Length of Sessions and Their Frequency of Occurrence.

Number of Steps	Number of Sessions
2	9
4	5
6	3
9	1
10	6
11	16
12	1
14	14
15	2
16	6
17	4
37	4
40	1
43	2
45	1
47	4
48	1

Source: Author's calculations.

Table 3. Frequency of Matching Question-Answer Trees by Linguistic Term for Sessions Using the Georgia Database.

Term	Number of Agreements	Number of Steps
Near	2	14
in_the_Vicinity_of	2	14
Remote_from	2	10
Close_to	2	14
Long_distance_from	3	14

Source: Author's calculations.

Table 4. Frequency of Matching Question-Answer Trees by Linguistic Term for Sessions Using the Connecticut Database.

Term	Number of Agreements	Number of Steps
Near	2	11
Far	2	11
Distant_from	3	11
Short_distance_from	2	47
Long_distance_from	3	11

Source: Author's calculations.

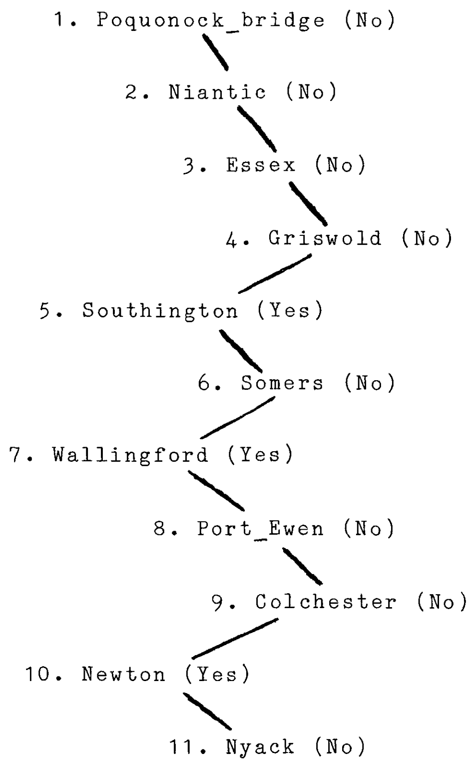

Figure 1. The Question-Answer Tree For A Subject's Definition of Near_Waterbury, Connecticut.

the subject. We find no major landscape features and both settlements are on direct, straight routes from Douglas. Furthermore, their routes are nearly orthogonal to one another. Depending on how one composes a multiperson concept, variations such as this intransitivity may be incorporated into multiperson concept.

In Table 6 we show what the results of several methods of multiperson concept formation using the membership values in Table 5 and using Subject 1 as the "chair of the committee." As one might expect, Subject 4 had an inordinate influence on the concept formed using the global evidence method. This was due to the preponderonce of high membership values. To dampen the influence of just such a situation the Combination Methods I and II were devised.

By inspecting the results in regard to the membership values of tifton, lenox, nicholls, and alma, one can see why method I is said to give the chairman 'veto' powers whereas method II is more like 'acceptance' powers. We chose Subject 1 as

Table 5. Fuzzy Membership Values for the Spatial Relation Close_to_Douglas Acquired from Subjects Using the Waycross, Georgia Database.

Settlement Name	Subjects 1	2	3	4	5
cordele	0.50	0.00	0.00	0.67	0.50
ashburn	0.53	0.40	0.40	0.70	0.50
sylvester	0.50	0.50	0.50	0.66	0.50
doerun	0.50	0.50	0.50	0.63	0.50
moultrie	0.52	0.52	0.52	0.64	0.50
coolidge	0.50	0.50	0.50	0.62	0.50
poulan	0.50	0.52	0.52	0.67	0.50
tifton	0.00	0.62	0.62	0.72	0.51
sycamore	0.48	0.44	0.44	0.70	0.50
lenox	0.43	0.65	0.65	0.71	0.51
adel	0.44	0.63	0.63	0.70	0.44
nashville	0.00	0.51	0.51	0.74	0.00
alapaha	0.00	0.00	0.00	0.76	0.00
ocilla	0.05	0.00	0.00	0.78	0.00
fitzgerald	0.00	0.00	0.00	0.78	0.00
rochelle	0.52	0.39	0.39	0.74	0.50
abeville	0.54	0.51	0.51	0.76	0.54
broxton	0.95	0.95	0.95	0.95	0.95
pearson	0.95	0.95	0.95	0.81	0.00
lakeland	0.00	0.00	0.00	0.73	0.00
willacoochee	0.52	0.52	0.52	0.79	0.00
homerville	0.60	0.60	0.59	0.77	0.00
nicholls	0.05	0.95	0.95	1.00	0.95
lumber_city	0.00	0.00	0.00	0.90	0.00
hazlehurst	0.00	0.00	0.00	0.93	0.00
waycross	0.05	0.05	0.05	0.05	0.05
blackshear	0.05	0.05	0.05	0.05	0.05
alma	0.00	0.72	0.72	0.92	0.56
baxley	0.00	0.05	0.05	1.00	0.00

Source: Author's calculations.

chairman to illustrate the importance of the chair in each of the methods. In Combination Method I the intransitivity regarding pearson and nicholls is perserved due to the veto power of the chair, whereas using method II eliminates the intransitivity. Also, note that in 24 cases the methods produced the same membership value, differing in those cases where substantial disagreement existed between the chairman and the remainder of the 'committee.'

We used the Zimmermann method with a number of different values of γ . Results the Zimmermann method are reported in Table 7. What is most disturbing about this method of multiperson concept formation is the influence that a membership value of zero has on it. Thus, we feel that this method is more appropriate when the membership values are

Table 6. Membership Values Resulting from Four Methods of Multiperson Concept Formation Using Results Reported in Table 5.

Settlement Name	Agreement Method	Global Evidence Method	Combination* Method I	Combination* Method II
cordele	0.00	0.67	0.50	0.50
ashburn	0.40	0.84	0.53	0.53
sylvester	0.50	0.79	0.50	0.50
doerun	0.50	0.74	0.50	0.50
moultrie	0.50	0.75	0.52	0.50
coolidge	0.50	0.70	0.50	0.50
poulan	0.50	0.80	0.50	0.50
tifton	0.00	0.89	0.00	0.51
sycamore	0.44	0.85	0.48	0.48
lenox	0.43	0.84	0.43	0.51
adel	0.44	0.80	0.44	0.44
nashville	0.00	0.84	0.00	0.00
alapaha	0.00	0.76	0.00	0.00
ocilla	0.00	1.00	0.05	0.05
fitzgerald	0.00	0.94	0.00	0.00
rochelle	0.39	0.83	0.52	0.52
abeville	0.51	0.83	0.54	0.54
broxton	0.95	0.95	0.95	0.95
pearson	0.00	0.95	0.95	0.95
lakeland	0.00	0.73	0.00	0.00
willacoochee	0.00	0.87	0.52	0.52
homerville	0.00	0.77	0.60	0.60
nicholls	0.05	1.00	0.05	0.95
lumber_city	0.00	0.90	0.00	0.00
hazlehurst	0.00	0.93	0.00	0.00
waycross	0.05	0.05	0.05	0.05
blackshear	0.05	0.05	0.05	0.05
alma	0.00	0.92	0.00	0.56
baxley	0.00	1.00	0.00	0.00

* Subject 1 was used as the "committee chairperson."
Source: Author's calculations.

non-zero. As noted above, one can plainly see that as increases so does the influence of the global evidence method. However, it remains unclear what guidelines should be followed when deciding on a value for γ.

Table 7. Results of Zimmermann's Method of Multiperson Concept Formation Using Membership Values in Table 5.

Settlement Name	Zimmermann Method γ = 0.9	0.7	0.5	0.3	0.1
cordele	0.00	0.00	0.00	0.00	0.00
ashburn	0.43	0.50	0.58	0.68	0.78
sylvester	0.52	0.57	0.63	0.69	0.76
doerun	0.52	0.56	0.61	0.66	0.71
moultrie	0.52	0.56	0.61	0.66	0.72
coolidge	0.52	0.55	0.59	0.63	0.68
poulan	0.52	0.57	0.63	0.70	0.77
tifton	0.00	0.00	0.00	0.00	0.00
sycamore	0.47	0.53	0.62	0.70	0.80
lenox	0.46	0.52	0.59	0.68	0.78
adel	0.47	0.53	0.59	0.67	0.76
nashville	0.00	0.00	0.00	0.00	0.00
alapaha	0.00	0.00	0.00	0.00	0.00
ocilla	0.00	0.00	0.00	0.00	0.00
fitzgerald	0.00	0.00	0.00	0.00	0.00
rochelle	0.42	0.49	0.57	0.66	0.77
abeville	0.53	0.59	0.65	0.72	0.79
broxton	0.95	0.95	0.95	0.95	0.95
pearson	0.00	0.00	0.00	0.00	0.00
lakeland	0.00	0.00	0.00	0.00	0.00
willacoochee	0.00	0.00	0.00	0.00	0.00
homerville	0.00	0.00	0.00	0.00	0.00
nicholls	0.06	0.11	0.21	0.39	0.73
lumber_city	0.00	0.00	0.00	0.00	0.00
hazlehurst	0.00	0.00	0.00	0.00	0.00
waycross	0.05	0.05	0.05	0.05	0.05
blackshear	0.05	0.05	0.05	0.05	0.05
alma	0.00	0.00	0.00	0.00	0.00
baxley	0.00	0.00	0.00	0.00	0.00

Source: Author's calculations.

CONCLUDING COMMENTS

We demonstrated that approximate representations can be acquired from exact user responses using mixed-initiative man-machine interactions. In previous reports (Robinson et al, 1985a; Robinson et al, 1986a) we contented that human users might not respond in as deterministic manner as did the simulations. Results reported here and in Robinson et al (1986c) support that contention. In fact, our results suggest that even with a relatively small, simple spatial database, significant semantic variation does exist. This result has significant implications for maintaining semantic integrity within and between geographic databases. Furthermore, these results question the wisdom of using simple term matching to represent 'fuzzy' queries (eg. Chang and Ke, 1979).

These results show that there exists significant interuser and interterm variation. Results of this study suggest that size of the database may not influence the overall interuser semantic variation as much as one might suspect. In fact, this study suggests that regardless of database size, there is a roughly even chance that one user's question-answer tree will be the same as anothers. However, given our sample size, we refrain from making any strong probabilistic statements.

Of particular importance in this study has been the consideration of multiperson concept formation. Most spatial information systems are used in an organizational context where group (multiperson) concepts are the norm. We suspect that some of the work dealing with of multiperson concept formation will become valuable in dealing with semantic variations found in the distributed geographic information systems of the future. Regardless of the application domain, we showed how important it is to understand the process of multiperson concept formation. Future research on this topic will, most likely, be cast within the context of knowledge acquisition for use in expert systems.

Finally, we believe this study illustrates that this avenue of research has implications for developing systems for the detection and representation of ill-defined spatial entities as well as spatial relations that are by their nature fuzzy concepts. The mixed-initiative man-machine interaction coupled with approximate reasoning appears to us to be a particularly attractive approach for acquiring an approximate representation of ill-defined geographic concepts or features for subsequent use in an expert geographic information system.

ACKNOWLEDGEMENTS

Partial support was provided by National Science Foundation Grant IST-8412406 and a Research Grant from PSC-CUNY. We also wish to acknowledge the support provided by The University of Calgary and the City University of New York.

REFERENCES

Barrera, R. and A. Buchmann. 1981, "Schema Definition and Query Language for a Geographical Database System", IEEE Trans. on Computer Architecture, Pattern Analysis and Image Database Management, 250-256.

Buckles, B.P. and F.E. Petry. 1983, "Information-Theoretical Characterization of Fuzzy Relational Databases", IEEE Trans. Syst., Man, and Cybernetics, SMC-13, 74-77.

Chamberlain, D.D. and R.F. Boyce. 1974, "SEQUEL: A Structured English Query Language", in Proceedings of ACM SIGFIDET Workshop on Data Description, Access, and Control, Ann Arbor, MI., 713-775.

Chang, S. and J. Ke. 1979, "Translation of Fuzzy Queries for Relational Database System," IEEE Trans. on Pattern Analysis and Machine Intelligence, 1, 281-294.

De Luca, A. and S. Termini. 1972, "A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory," Information and Control, 20, 301-312.

Frank, A. 1982, "MAPQUERY: Data Base Query Language for Retrieval of Geometric Data and their Graphical Representation", Computer Graphics, 16, 199-207.

Gaglio, S., R. Minciardi, and P.P. Puliafito. 1985, "Multiperson Decision Aspects in the Construction of Expert Systems," IEEE Trans. on Systems, Man, and Cybernetics, 15, 193-204.

Hersh, H.M., A. Caramazza, and H.H. Brownell. 1979, "Effects of context on fuzzy membership functions", in Advances in Fuzzy Set Theory and Applications, M.M. Gupta, R.K. Ragade, R.R. Yager (Eds.), North Holland Publishing, New York, 389-408.

Jain, R. 1980, "Fuzzyism and real world problems," in Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems, S.K. Chang and P.P. Wang (eds.), Plenum: New York, 129-133.

Kaufmann, A. 1975, Introduction to the theory of fuzzy subsets, Academic Press, New York.

Leung, Y. 1982, "Approximate characterization of some fundamental concepts of spatial analysis," Geographical Analysis, 14(1), 29-40.

Lundberg, G. 1982, "Modeling constraints and anticipation: linguistic variables, foresight-hindsight, and relative alternative attractiveness," Geographical Analysis, 14(4), 347-355.

Nakamura, K. and S. Iwai. 1982, "Topological Fuzzy Sets as a Quantitative Description of Analogical Inference and Its Application to Question-Answering Systems for Information Retrieval", IEEE Trans. on Systems, Man, and Cybernetics, 12, 193-204.

Robinson, V.B. 1983, "Integrating mathematical spatial algorithms in an interactive microcomputing environment," Monograph on Spatial Mathematical Algorithms for Land Data Systems,Lincoln Institute of Land Policy Colloquium Series: Cambridge, Massachusetts.

Robinson, V.B. 1984, "Modeling Inexactness in Spatial Information Systems", _Proceedings_ of Pittsburgh Conference on Modeling and Simulation, Pittsburgh, PA.

Robinson, V.B. 1986, "Implications of Fuzzy Set Theory for Geographic Databases," _Computers, Environment, and Urban Systems_, in press.

Robinson, V.B. and A.U. Frank. 1985, "About Different Kinds of Uncertainty in Geographic Information Systems,", _Proceedings_, AUTOCARTO 7, Washington, D.C.

Robinson. V.B., D. Thongs, and M. Blaze. 1985a, "Machine Acquisition and Representation of Natural Language Concepts for Geographic Information Retrieval," _Proceedings_, Pittsburgh Modeling & Simulation Conference, University of Pittsburgh.

Robinson. V.B., M. Blaze, and D. Thongs. 1985b, "Natural Language Concepts in Geographic Data Processing Systems," _Proceedings_ Intern'l Conf on Adv. Technology for Monitoring and Processing Global Environ'l Data, London, UK.

Robinson, V.B., M. Blaze, and D. Thongs. 1986a, "Man-Machine Interaction for Acquisition of Spatial Relations as Natural Language Concepts," in B. Opitz (ed.) _Geographic Information Systems in Government_, A.Deepak Press: Hampton, VA.

Robinson, V.B., A.U. Frank and M. Blaze. 1986b, "Expert Systems And Geographic Information Systems: Review and Prospects," _Jrnl of Surveying Engineering_, 112(2): 119-130.

Robinson, V.B., M. Blaze, and D. Thongs. 1986c, "Representation and Acquisition of a Natural Language Relation for Spatial Information Retrieval," _Proceedings_ 2nd Intern'l Symp on Spatial Data Handling, 472-487.

Tomlin, C.D. and S.M. Tomlin. 1981, "An Overlay Mapping Language", presented at Symp on Regional Landscape Planning, Am. Soc. of Landscape Architects.

Yager, R.R. 1981, "Measurement of Properties on Fuzzy Sets and Possibility Distribution," _Proceedings_ 3rd Intern'l Seminar on Fuzzy Set Theory, E.P. Klement (ed), J. Kepler Univ, Linz, Austria, 211-222.

Zadeh, L. 1978, "PRUF-A Meaning Representation Language for Natural Languages," _Int. Jrnl. Man-Machine Studies_, 10, 395-460.

Zadeh, L. 1981, "Test-Score Semantics for Natural Languages and Meaning Representation via PRUF," in _Empirical Semantics_, B. Rieger (ed.), Brockmeyer, Bochum, 281-349.

Zimmermann, H.J. 1983, "Using Fuzzy Sets in Operational Research," _European Jrnl of Operational Research_, 13(3): 201-216.