1993013003

APPLICATION OF FUZZY SET AND DEMPSTER-SHAFER THEORY TO ORGANIC GEOCHEMISTRY INTERPRETATION

N93-22212

C. S KIM and G. H. Isaksen
Exxon Production Research Co.
Houston, Tx. 77001
(713) 965-7645

50406

p. 9

ABSTRACT

This paper presents an application of fuzzy sets and Dempster Shafer theory (DST) in modeling the interpretational process of organic geochemistry data for predicting the level of maturities of oil and source rock samples. This has been accomplished by (i) representing linguistic imprecision and imprecision associated with experience by a fuzzy set theory, (ii) capturing the probabilistic nature of imperfect evidences by a DST, and (iii) combining multiple evidences by utilizing John Yen's[1] generalized Dempster-Shafer Theory(GDST), which allows DST to deal with fuzzy information. The current prototype provides collective beliefs on the predicted levels of maturity by combining multiple evidences through GDST's rule of combination.

I. INTRODUCTION

Modeling the interpretation process of an expert requires representation and management of uncertain knowledge. This is because nearly every interesting domain contains knowledge that is inherently inexact, incomplete, or unmeasurable.

In this paper we explicitly treat two forms of uncertainties. One form of uncertainty is fuzziness related to linguistic imprecision. Based on fuzzy set theory, Zadeh[2] developed possibility theory to express this type of imprecision. The other form of uncertainty is the probability with which a certain evidence correctly predicts a subset of hypotheses. Dempster-Shafer Theory[3,4] (DST) deals with this type of uncertainty and provides a mechanism for combining multiple evidences for an overall belief in a subset of hypotheses. Unlike classical probability theory, DST enables the degree of ignorance to be expressed explicitly and does not fix hypothesis negation probability once occurrence probability is known.

In the past, several attempts[5,6] have been made to generalize DST to deal with fuzzy information. While these attempts fall short of fully justifying their approaches, John Yen[1] proposed a generalized Dempster-Shafer Theory (GDST), in which the important principle of DST is preserved: That the belief and the plausibility functions are treated as lower and upper probability bounds.

In this paper, we demonstrate representation and management of two types of uncertainties by GDST as applied to the interpretation of organic geochemistry data. In the following sections, we review the basics of GDST, and the development of a knowledge-based system for geochemistry interpretation

II. BASICS OF A GENERALIZED DEMPSTER-SHAFER THEORY

This review is not intended to describe detailed theory and developments of DST and GDST. Rather, we plan to describe their representation of imprecise information and the rule of combination in a qualitative way. More interested readers should refer to the references [1,3,4] cited.

In the DST, hypotheses in a frame of discernment must be mutually exclusive and exhaustive, meaning that they must cover all the possibilities and the individual hypothesis cannot overlap with others. An important advantage of DST over classical probability theory is its ability to express degree of ignorance associated with an evidence. Also, unlike classical probability theory, a commitment of belief to a hypothesis does not force the remaining belief to be assigned to its compliment. Therefore, the amount of belief not committed to any of the subsets of hypotheses represents the degree of ignorance. In DST, a basic probability assignment(bpa) m(A), as a generalization of a probability, indicates belief in a subset of hypotheses A. This quantity m(A) serves as a measure of belief committed to the subset A.

DST also provides a formal process for combining bpa's induced by independent evidential sources, which is called the rule of combination. This process is a tool for accumulating evidences to narrow the hypothesis set. If m_1, and m_2 are two bpa's from two evidential sources, a combined bpa is computed according to the rule of combination:

$$m_1 \oplus m_2(C) = \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)/k \qquad (1)$$

where k is a normalization factor,

$$k = 1 - \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j), \qquad (1a)$$

$m_1 \oplus m_2(C)$ is a combined bpa for a hypothesis C,
ϕ is a null set, and
A_i, B_j are hypotheses sets induced by the two evidential sources.

In the GDST proposed by Yen[6], a basic probability m(A) is assigned to a fuzzy subset of hypotheses. In this framework, each fuzzy subset of hypotheses has bpa m(A), and fuzzy membership function $\mu_A(x_i)$, where x_i's are elemental hypotheses in the frame of discernment. The rule of combination in GDST consists of two operations: a cross-product operation and a normalization process. Basic probabilities are first combined by performing a generalized cross-product including fuzzy set operations:

$$m'_{12}(C) = m_1 \otimes m_2(C) = \sum_{A_i \cap B_j = C} m_1(A_i)\, m_2(B_j) \qquad (2)$$

where $m'_{12}(C)$ is an unnormalized bpa induced by two evidences, and $\cap$ denotes a fuzzy intersection operator.

Then, a normalization is performed on fuzzy subsets of hypotheses whose maximum membership values are less than one. A detailed procedure and justification of this normalization process can be found in the reference [1]. Yen[1] also showed that this normalization can be postponed until the last evidence without affecting the computational results and the commutativity of the rule of combination.

In case of combining only two fuzzy bpa's, a combined bpa using GDST's rules of combination is:

$$m_1 \oplus m_2(C) = \sum_{(\overline{A \cap B}) \not\subset C} \underset{x_i}{Max}\ \mu_{A \cap B}(x_i)\ m_1(A) m_2(B)/k \qquad (3)$$

where

$$k = 1 - \sum_{A,B} (1 - \underset{x_i}{Max}\, \mu_{A \cap B}(x_i))\ m_1(A) m_2(B), \text{ and} \qquad (3a)$$

$\overline{A \cap B}$ is a normalized $A \cap B$.

As can be noticed in the equations above, GDST allows partially conflicting evidences, while DST only allows either conflicting or confirming evidences.

III. BIOMARKER INTERPRETATION SYSTEM

In exploration for oil and gas, it is important to be able to assess the maximum temperatures to which sediments or oils have been exposed in the subsurface. This is referred to as the level of thermal maturity. Organic chemical compounds known as biomarkers enable the geochemist to assess the level of maturity (LOM) of oils and sedimentary organic matter. In this paper, we focus our attention on modeling the process of interpreting biomarker data to predict LOM. The LOM scale ranges from 1 to 20, with LOM=1 being least mature and LOM=20 most mature. There exist more than 10 biomarkers whose intensities have definite links to the maturity with varying degrees of resolution and prediction power.

In our approach, these varying degrees of resolution among biomarker evidences are represented by fuzzy subsets of maturity intervals, and the probability with which an evidence correctly predicts a fuzzy maturity interval is represented by a basic probability in GDST. Therefore, evidential knowledge is represented in fuzzy rules, and the confidence for a specific rule is represented by a bpa. Moreover, GDST's rule of combination provide collective belief in the predicted level of maturity. In the following, detailed representation methods are presented along with actual application results.

(A) Representing Two Types of Imprecision

Interpretation of geochemical data is based on experience as well as theory. This interpretational knowledge is descriptive in nature, and best represented by fuzzy logic and possibility theory. For example, one may have an experience based correlation study between level of maturity (LOM) and $\%C_{29}20S$, which is a ratio of the intensities of several organic compounds. Then, the correlation curve in Figure 1 may be used by an interpreter as follow:

```
IF %C2920S is 40 %,
THEN expected LOM is about 8.
```

In the rule above, the concluding part is descriptive in that LOM = 8 is most possible, but LOM values of 6,7,9, and 10 are also possible with lesser degree as shown in Figure 2. Another example is the case where both premise and conclusion are best represented by fuzzy membership functions. Based on theory and experience, Heptane value can only predict maturity levels in four qualitative categories, such as immature, early mature, mature, and over mature. Examples of Heptane rules are:

```
IF Heptane value is medium,
THEN maturity is early mature

IF Heptane value is high,
THEN maturity is mature

IF Heptane value is very high,
THEN maturity is over mature
```

In the rules above, both the premise and the conclusions are descriptive and best represented by membership functions for Heptane value and maturity as depicted in Figure 3a and Figure 3b. From the fuzzy rules above and the membership functions in Figures 3a and 3b, observation of a Heptane value of 19 will result in the possibility values of 0.5, 1.0, 1.0, and 0.5 for LOM = 6, 7, 8, and 9 respectively:

$$\Pi_{LOM} = \{0.5/6,\ 1/7,\ 1/8,\ .5/9\} \qquad (4)$$

In the current system, LOM is predicted from 10 evidences each of which predicts LOM with different degree of resolution as shown by the two examples above.

In addition to the imprecision in the knowledge represented by possibility theory above, there exists another type of uncertainty associated with evidences. For example, rules associated with $\%C_{29}20S$ have higher probability of being true than the Heptane rules. In our approach, the probability with which a proposition " If A is a1 Then B is b1" is true is represented by bpa assigned to the fuzzy subset of hypotheses induced by the proposition. The compliment of this probability is assigned to the degree of ignorance associated with the proposition, since our system generates only one fuzzy subset of hypotheses for each evidence.

(B) Test Result

In order to validate the system, thirty interpretations were tested to see if the system's interpretations conformed to those of the expert. With reference to the test results listed in Table 1, one can notice that the system interpreted maturities are biased towards higher LOM. However, these errors are all higher than they should be and consistent by itself, and can be traced to the membership function definitions. We are currently fine tuning these membership functions to correct the problem and plan to test the system with additional field data..

V. CONCLUSIONS

We presented a knowledge-based system in which linguistic imprecisions and uncertainties associated with fuzzy rules are modeled in the frame work of a generalized Dempster-Shafer Theory. This development is significant in that many application problems in oil exploration requires a mechanism of combining fuzzy information from various sources.

Even though the current biomarker interpretation system has been tested on only 30 data sets, the system will be further tested with additional field data and expanded to handle interpretations for other characteristics such as source facies, depositional environments, and the degree of biodegradation.

REFERENCES

1. Yen, J., Generalizing the Dempster-Shafer theory to fuzzy sets, IEEE Trans., Sys., Man, & Cyb., Vol. 20, No, 3, May/June 1990
2. Zadeh, L.A., Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets & Systems 1(1978) 3-28, North-Holland Publishing Co., 1978
3. Dempster, A.P., Upper and lower probabilities induced by a multivalued mapping, Annals Math. Statistics, Vol. 38, No. 2, 1967, pp.325-339

4 Shafer, G., A mathematical theory of evidence, Princeton Univ. Press, Princeton, N.J., 1976

5. Ishizuka, M., K.S. Fu, and J.T.P. Yao, Inference procedures and uncertainty for the problem-reduction method, Inform. Sci., Vol. 28, 1982, pp. 179-206
6. Yager, R., Generalized probabilities of fuzzy events from fuzzy belief structure, Inform. Sci., Vol. 28, 1982, pp. 45- 62

Table 1. Comparison of interpretations

Data Set Number	Interpreted LOM	System Generated LOM
1	8-9	9-10
2	9	10
3	9	10
4	9	10-11
5	9	10
6	9	10-11
7	8.5-9	9-10
8	>10	11
9	9	9-10
10	9	9-10
11	9	9-10
12	7.5-8	7
13	>10	11-11.5
14	>10	11-11.5
15	10-11	11
16	11	11
17	9	10
18	7.5-8	8-9
19	8	9-10
20	10	11-11.5
21	10	11-11.5
22	10	11
23	10	11-11.5
24	9	9
25	9	9.5-10
26	10	11-11.5
27	9-10	11
28	9	9.5-10
29	10-11	11
30	10-11	10-11

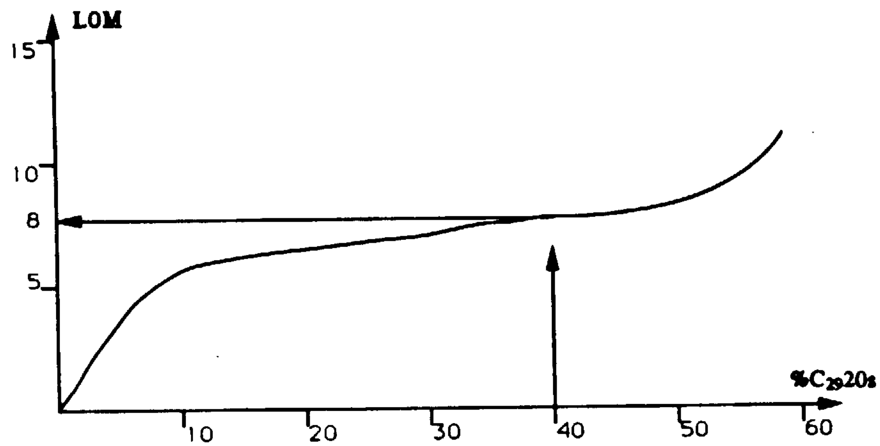

Figure 1: Experience based correlation curve between %C29 20s and LOM

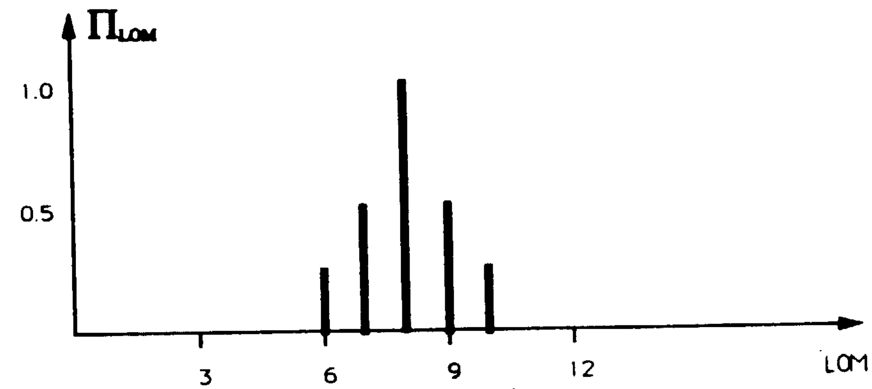

Figure 2: Possibility of LOM's induced from observing %C29 20s = 40

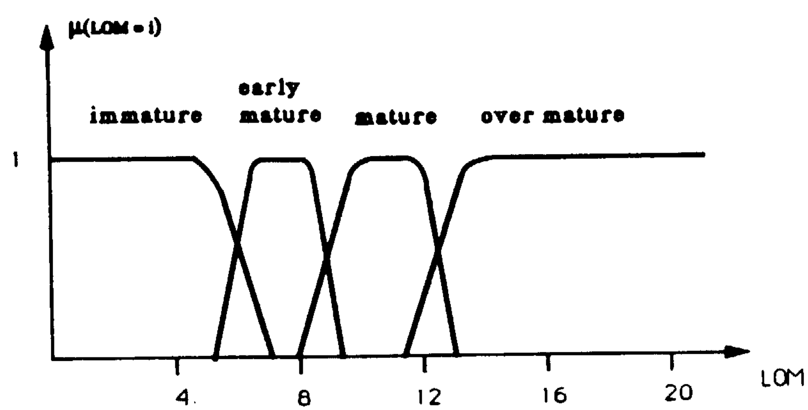

Figure 3a: Membership functions for maturity

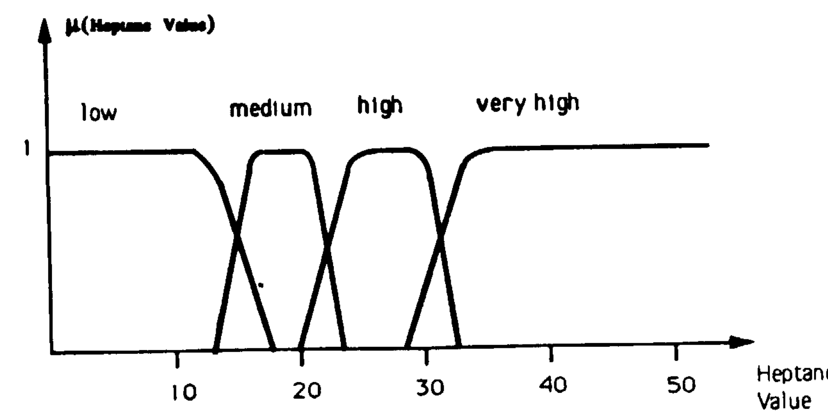

Figure 3b: Membership functions for heptane values