Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 38 (2011) 4460-4464

ISSN: 2229-712X

A concept graph for rough set theory

B.Sankara Subramanian

Department of CSE, SCSVMV University, Kanchipuram, TamilNadu.

ARTICLE INFO

Article history:
Received: 22 August 2011;
Received in revised form:
26 August 2011;
Accepted: 31 August 2011;

Keywords
Concept Graph,
Ontology,
Rough set theory,
Knowledge representation.

ABSTRACT

Conceptual graphs (CGs) are a system of logic based on the existential graphs of Charles Sanders Peirce and the semantic networks of artificial intelligence. They express meaning in a form that is logically precise, humanly readable, and computationally tractable. With a direct mapping to language, conceptual graphs serve as an intermediate language for translating computer-oriented formalisms to and from natural languages. With their graphic representation, they serve as a readable, but formal design and specification language. CGs have been implemented in a variety of projects for information retrieval, database design, expert systems, and natural language processing. The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). For example if objects are patients suffering from a certain disease symptoms of the disease form information about patients- Objects characterized by the same information are indiscernible similar in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory.

Introduction

Knowledge representation and reasoning [1][5]is an area of artificial intelligence whose fundamental goal is to represent knowledge in a manner that facilitates inferencing (i.e. drawing conclusions) from knowledge. It analyzes how to formally think - how to use a symbol system to represent a domain of discourse (that which can be talked about), along with functions that allow inference (formalized reasoning) about the objects. Generally speaking, some kind of logic is used both to supply formal semantics of how reasoning functions apply to symbols in the domain of discourse, as well as to supply operators such as quantifiers, modal operators, etc. that, along with an interpretation theory, give meaning to the sentences in the logic. The single most important decision to be made, is the expressivity of the KR. The more expressive, the easier and more compact it is to "say something".

In computer science and information science, an ontology [6] is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to describe the domain. What ontology has in common in both computer science and in philosophy is the representation of entities, ideas, and events, along with their properties and relations, according to a system of categories.

Knowledge Interchange Format

Knowledge Interchange Format gives the definition of conceptual graphs and the various notations for representing them,

KIF[7] is a language designed for use in the interchange of knowledge among disparate computer systems (created by different programmers, at different times, in different languages, and so forth).

The following categorical features are essential to the design of KIF.

1. The language has declarative semantics. It is possible to understand the meaning of expressions in the language without appeal to an interpreter for manipulating those expressions. In this way, KIF differs from other languages that are based on specific interpreters, such as Emycin and Prolog.
2. The language is logically comprehensive -- at its most general, it provides for the expression of arbitrary logical sentences. In this way, it differs from relational database languages (like SQL) and logic programming languages (like Prolog).
3. The language provides for the representation of knowledge about knowledge. This allows the user to make knowledge representation decisions explicit and permits the user to introduce new knowledge representation constructs without changing the language.

In addition to these essential features, KIF is designed to maximize the following additional features (to the extent possible while preserving the preceding features).

Implementability Although KIF is not intended for use within programs as a representation or communication language, it should be usable for that purpose if so desired.

Readability. Although KIF is not intended primarily as a language for interaction with humans, human readability facilitates its use in describing representation language semantics, its use as a publication language for example knowledge bases, its use in assisting humans with knowledge base translation problems, etc.

With this background settings, the paper is organized as follows.

Section 1 is introduction. Section 2 delineates conceptual graphs. Section 3 is devoted to Rough set theory. Section 4 gives the concept graph for Rough set theory and explains it. Section 5 is the conclusion.

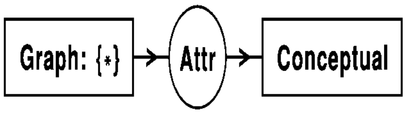
Conceptual Graphs

A conceptual schema [2]or conceptual data modelis a map of concepts and their relationships. This describes the semantics

Tele:
E-mail addresses: coolsankara@gmail.com

of an organization and represents a series of assertions about its nature. Specifically, it describes the things of significance to an organization (entity classes), about which it is inclined to collect information, and characteristics of (attributes) and associations between pairs of those things of significance (relationships).

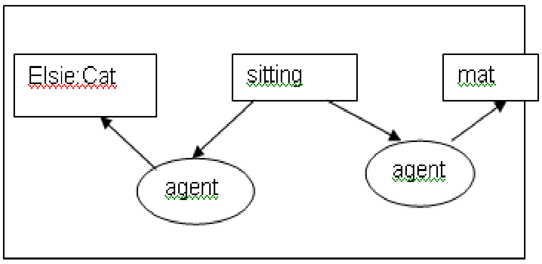
Conceptual graphs (CGs) are a formalism for knowledge representation. Conceptual graphs are formally defined in an abstract syntax that is independent of any notation, but the formalism can be represented in several different concrete notations



A Grpahical interface to First-Order Logic

In this approach, a formula in first-order logic (Predicate Calculus)[3] is represented by a labeled graph.

A linear notation, called the Conceptual Graph Interchange Format (CGIF), has been standardized in the ISO standard for common logic.



The diagram on the right is an example of the display form for a conceptual graph. Each box is called a concept node, and each oval is called a relation node. In CGIF, this CG would be represented by the following statement:

```
[cat Elsie] [sitting *x] [Mat *y] (agent
?x Elsie) (location ?x ?y)
```

In CGIF, brackets enclose the information inside the concept nodes, and parentheses enclose the information inside the relation nodes. The letters x and y, which are called coreference labels, show how the concept and relation nodes are connected. In the Common Logic Interchange Format (CLIF) [4], those letters are mapped to variables, as in the following statement:

```
(exists ((x Sitting ) (y Mat)) (and (Cat
Elsie) (agent x Elsie) (location x y)))
```

variables in CLIF, and the question marks on ?x and ?y map to bound variables in CLIF. A universal quantifier, represented @every*z in CGIF, would be represented for all (z) in CLIF.

Reasoning can be done by translating graphs into logical formulas, then applying a logical inference engine.

Rough set theory

Rough set theory [8] is still another approach to vagueness. Similarly to fuzzy set theory it is not an alternative to classical set theory but it is embedded in it. Rough set theory can be viewed as a specific implementation of Frege's idea of vagueness, i.e., imprecision in this approach is expressed by a boundary region of a set, and not by a partial membership, like in fuzzy set theory.

Rough set concept can be defined quite generally by means of topological operations, interior and closure, called approximations.

Let us describe this problem more precisely. Suppose we are given a set of objects U called the universe and an indiscernibility relation $R \subseteq U \times U$, representing our lack of knowledge about elements of U. For the sake of simplicity we assume that R is an equivalence relation. Let X be a subset of U. We want to characterize the set X with respect to R. To this end we will need the basic concepts of rough set theory given below.

- The lower approximation of a set X with respect to R is the set of all objects, which can be for certain classified as X with respect to R (are certainly X with respect to R).
- The upper approximation of a set X with respect to R is the set of all objects which can be possibly classified as X with respect to R (are possibly X in view of R).
- The boundary region of a set X with respect to R is the set of all objects, which can be classified neither as X nor as not-X with respect to R.

Now we are ready to give the definition of rough sets.

- Set X is crisp (exact with respect to R), if the boundary region of X is empty.
- Set X is rough (inexact with respect to R), if the boundary region of X is nonempty.

Thus a set is rough (imprecise) if it has nonempty boundary region; otherwise the set is crisp (precise).

This is exactly the idea of vagueness proposed by Frege. The approximations and the boundary region can be defined more precisely. To this end we need some additional notation.

The equivalence class of *R* determined by element *x* will be denoted by *R*(*x*).

The indiscernibility relation in certain sense describes our lack of knowledge about the universe. Equivalence classes of the indiscernibility relation, called *granules* generated by *R*, represent elementary portion of knowledge we are able to perceive due to *R*. Thus in view of the indiscernibility relation, in general, we are able to observe individual objects but we are forced to reason only about the accessible granules of knowledge.

Formal definitions of approximations and the boundary region are as follows:

- *R-lower approximation* of *X*

$$R_*(x) = \bigcup_{x \in U} \{R(x) : R(x) \subseteq X\}$$

- *R-upper approximation* of *X*

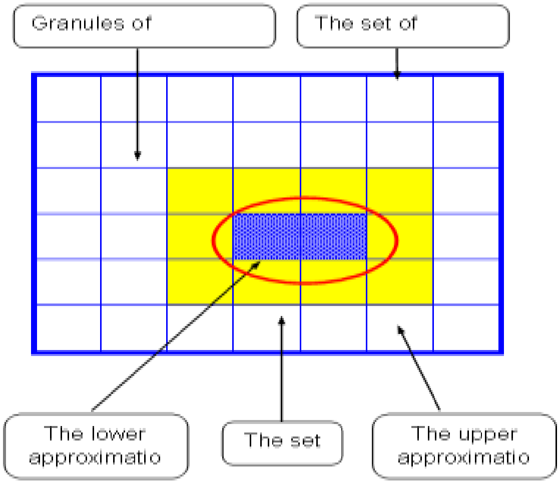
$$R^*(x) = \bigcup_{x \in U} \{R(x) : R(x) \cap X \neq \varnothing\}$$

- *R-boundary region* of *X*

$$RN_R(X) = R^*(X) - R_*(X)$$

As we can see from the definition approximations are expressed in terms of granules of knowledge.

The lower approximation of a set is union of all granules which are entirely included in the set; the upper approximation – is union of all granules which have non-empty intersection with the set; the boundary region of set is the difference between the upper and the lower approximation.



It is interesting to compare definitions of classical sets, fuzzy sets and rough sets. Classical set is a primitive notion and is defined intuitively or axiomatically. Fuzzy sets are defined by employing the fuzzy membership function, which involves advanced mathematical structures, numbers and functions. Rough sets are defined by approximations. Thus this definition also requires advanced mathematical concepts.

Approximations have the following properties:

1) $R_*(X) \subseteq X \subseteq R^*(X)$

2) $R_*(\emptyset) = R^*(\emptyset) = \emptyset; R_*(U) = R^*(U) = U$

3) $R^*(X \cup Y = R^*(X) \cup R^*(Y)$

4) $R_*(X \cap Y) = R_*(X) \cap R_*(Y)$

5) $R_*(X \cup Y) \supseteq R_*(X) \cup R_*(Y)$

6) $R^*(X \cap Y) \subseteq R^*(X) \cap R^*(Y)$

7) $X \subseteq Y \rightarrow R_*(X) \subseteq R_*(Y) \,\&\, R^*(X) \subseteq R^*(Y)$

8) $R_*(-X) = -R^*(X)$

9) $R^*(-X) = -R_*(X)$

10) $R_*R_*(X) = R^*R_*(X) = R_*(X)$

11) $R^*R^*(X) = R_*R^*(X) = R^*(X)$

It is easily seen that approximations are in fact interior and closure operations in a topology generated by data. Thus fuzzy set theory and rough set theory require completely different mathematical setting.

Rough sets can be also defined employing, instead of approximation, rough membership function [5]

$\mu_X^R : U \rightarrow < 0,1 >$

where

$$\mu_X^R(x) = \frac{|X \cap R(x)|}{|R(x)|}$$

and $|X|$ denotes the cardinality of X.

The rough membership function expresses conditional probability that x belongs to X given R and can be interpreted as a degree that x belongs to X in view of information about x expressed by R.

The meaning of rough membership function can be depicted as shown in Fig.2.

The rough membership function can be used to define approximations and the boundary region of a set, as shown below:

$R_*(X) = \{x \in U : \mu_X^R(x) = 1\},$

$R^*(X) = \{x \in U : \mu_X^R(x) > 0\},$

$RN_R(X) = \{x \in U : 0 < \mu_X^R(x) < 1\}.$

It can be shown that the membership function has the following properties [5]:

1) $\mu_X^R(x) = 1$ *iff* $x \in R_*(X)$

2) $\mu_X^R(x) = 0$ *iff* $x \in U - R^*(X)$

3) $0 < \mu_X^R(x) < 1$ *iff* $x \in RN_R(X)$ $R(x)$

4) $\mu_{U-X}^R(x) = 1 - \mu_X^R(x)$ *for any* $x \in U$

5) $\mu_{X \cup Y}^R(x) \geq max\ (\mu_X^R(x), \mu_Y^R(x))$ *for any* $x \in U$

6) $\mu_{X \cap Y}^R(x) \leq min\ (\mu_X^R(x), \mu_Y^R(x))$ *for any* $x \in U$

From the properties it follows that the rough membership differs essentially from the fuzzy membership, for properties 5) and 6) show that the membership for union and intersection of sets, in general, cannot be computed – as in the case of fuzzy sets – from their constituents membership. Thus formally the rough membership is a generalization of fuzzy membership. Besides, the rough membership function, in contrast to fuzzy membership function, has a probabilistic flavour.

Now we can give two definitions of rough sets.

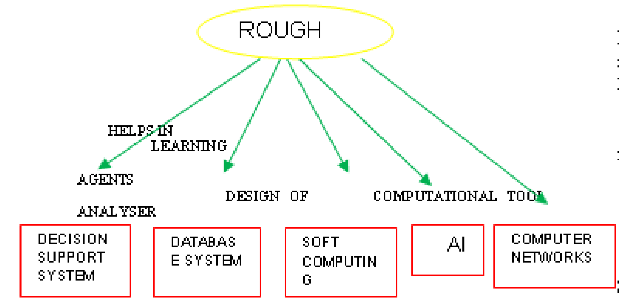
Definition 1: Set X is *rough* with respect to R if $R_*(X) \neq R^*(X)$.

Definition 2: Set X *rough* with respect to R if for some x, $0 < \mu_X^R(x) < 1$.

It is interesting to observe that definition 1 and definition 2 are not equivalent [5], but we will not discuss this issue here.

Let us also mention that rough set theory clearly distinguishes two very important concepts, vagueness and uncertainty, very often confused in the AI literature. Vagueness is the property of sets and can be described by approximations, whereas uncertainty is the property of elements of a set and can expressed by the rough membership function.

Rough set-concept graph



Rough set theory implication:

Decision Support System

Data mining and/or knowledge discovery is a very hot issue nowadays, as more and more information is being stored digitally the ability to collect data far outweighs the ability for a person to analyze it.

Therefore data mining and knowledge discovery is a very important part of today's business.

The purpose of this is to find descriptive and predictive model and/or pattern from the data.

Descriptive pattern are used when trying to classify a new data that will be introduced into the system and predictive pattern are used to forecast possible future outcome.

Database Systems

Knowledge discovery in databases, or dala mining, is an important direction in the development of data and knowledge-based systems. Because of the huge amount of data stored in large numbers of existing databases, and because the amount of data generated in electronic forms is growing rapidly, it is necessary to develop efficient methods to extract knowledge from databases.

Pattern classification	Dependency Analysis	Granular Computing	Machine Learning	Intrusion detection system
Intelligent Information Retrieval	Normalization	Fuzzy Logic	Neural Networks	Traffic Analysis
Multi-criteria decision Analysis.	Construction of Data-mining Query Languages	Artificial Neural Network Genetic Algorithm	Support Vector Machines K-Nearest Neighbour	Optimization Techniques
Knowledge Discovery from data	Rough set model for Databases.	Evolutionary computing	Naive Bayes Classifier Decision Tree Association rule Inductive Learning	

Soft computing

Soft computing [9]differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation. In effect, the role model for soft computing is the human mind. The guiding principle of soft computing is: Exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost. At this juncture, the principal constituents of Soft Computing (SC) are Fuzzy Logic (FL), Neural Computing (NC), Evolutionary Computation (EC) Machine Learning (ML) and Probabilistic Reasoning (PR), with the latter subsuming belief networks, chaos theory and parts of learning theory. What is important to note is that soft computing is not a melange. Rather, it is a partnership in which each of the partners contributes a distinct methodology for addressing problems in its domain. In this perspective, the principal constituent methodologies in SC are complementary rather than competitive. Furthermore, soft computing may be viewed as a foundation component for the emerging field of conceptual intelligence.

Artificial Intelligence

It is the science and engineering[10] of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. Intelligence involves mechanisms, and AI research has discovered how to make computers carry out some of them and not others. If doing a task requires only mechanisms that are well understood today, computer programs can give very impressive performances on these tasks. Such programs should be considered ``somewhat intelligent".

Computer Networks

A computer network, often simply referred to as a network, is a group of computers and devices interconnected by communications channels that facilitate communications among users and allows users to share resources.

Computer networks can be used for several purposes:

- *Facilitating communications.* Using a network, people can communicate efficiently and easily via email, instant messaging, chat rooms, telephone, video telephone calls, and video conferencing.
- *Sharing hardware.* In a networked environment, each computer on a network may access and use hardware resources on the network, such as printing a document on a shared network printer.
- *Sharing files, data, and information.* In a network environment, authorized user may access data and information stored on other computers on the network. The capability of providing access to data and information on shared storage devices is an important feature of many networks.
- *Sharing software.* Users connected to a network may run application programs on remote computers.

Conclusion

This paper addresses the problem of developing concept graph for rough set theory.This paves the way for linking various fields of computer science engineering with the rough set theory and also generates a meaningful and coherent conceptual framework for rough logic.

All kinds of knowledge (ontology, rules, constraints and facts) are labeled graphs, which provide an intuitive and easily understandable means to represent knowledge,

- reasoning mechanisms are based on graph notions, basically the classical notion of graph homomorphism; this allows, in particular, to link basic reasoning problems to other fundamental problems in computer science (problems on conjunctive queries in relational databases, constraint satisfaction problem, ...),
- the formalism is logically founded, i.e., it has a semantics in first-order logic and the inference mechanisms are sound and complete with respect to deduction in first-order logic,
- from a computational viewpoint, the graph homomorphism notion was recognized in the 90's as a central notion, and complexity results and efficient algorithm have been obtained in several domains.

Acknowledgment

I sincerely thank all those who helped me in producing this manuscript.

References

1. Chein, M., Mugnier, M.-L. (2009),Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs, Springer, 2009.
2. Dau,F. - The Logic System of Concept Graphs with Negation and Its Relationship to Predicate Logic, volume 2892 of LNCS, Springer, 2003.
3. Sowa, John F. (1984), Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, Reading, MA, 1984.
4. Sowa, John F. (1976), "Conceptual Graphs for a Data Base Interface", IBM Journal of Research and Development 20(4), 336–357, July 1976.
5. www.jfsowa.com/cg/ -
6. T. Gruber (2001) . What is an Ontology? Online entry. Accessed Nov 9, 2009.
7. logic.stanford.edu/kif/-A DARPA project
8.R. Deja: Conflict analysis, rough set methods and applications, in: L. Polkowski, S Tsumoto, T.Y. Lin (eds.), Studies in

Fuzzyness and Soft Computing, Physica-Verlag, A Springer-Verlag Company, 2000, 491-520.

9. www.soft-computing.de/def.html

10.www-formal.stanford.edu/jmc/whatisai/node1.html