Advances in Intelligent Systems and Computing 648

Patricia Melin
Oscar Castillo
Janusz Kacprzyk
Marek Reformat
William Melek *Editors*

Fuzzy Logic in Intelligent System Design

Theory and Applications

Springer

Advances in Intelligent Systems and Computing

Volume 648

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Patricia Melin · Oscar Castillo
Janusz Kacprzyk · Marek Reformat
William Melek
Editors

Fuzzy Logic in Intelligent System Design

Theory and Applications

Springer

Editors
Patricia Melin
Division of Graduate Studies and Research
Tijuana Institute of Technology
Tijuana, Baja California
Mexico

Oscar Castillo
Division of Graduate Studies and Research
Tijuana Institute of Technology
Tijuana, Baja California
Mexico

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warsaw
Poland

Marek Reformat
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, AB
Canada

William Melek
Laboratory of Computational Intelligence and Automation
University of Waterloo
Waterloo, ON
Canada

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-67136-9 ISBN 978-3-319-67137-6 (eBook)
DOI 10.1007/978-3-319-67137-6

Library of Congress Control Number: 2017952851

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

We describe in this book recent advances on the use of fuzzy logic in design of hybrid intelligent systems based on nature-inspired optimization and their application in areas such as intelligent control and robotics, pattern recognition, medical diagnosis, time series prediction, and optimization of complex problems. The book is organized into nine main parts, which contain a group of papers around a similar subject. The first part consists of papers with the main theme of theoretical aspects of fuzzy logic, which basically consists of papers that propose new concepts and algorithms based on type-1 fuzzy systems. The second part contains papers with the main theme of type-2 fuzzy logic, which are basically papers dealing with new concepts and algorithms for type-2 fuzzy systems. The second part also contains papers describing applications of type-2 fuzzy systems in diverse areas, such as time series prediction and pattern recognition. The third part contains papers that present enhancements to meta-heuristics based on fuzzy logic techniques describing new nature-inspired optimization algorithms that use fuzzy dynamic adaptation of parameters. The fourth part presents emergent intelligent models, which range from quantum algorithms to cellular automata. The fifth part contains papers describing applications of fuzzy logic in diverse areas of medicine, such as diagnosis of hypertension and hearth diseases. The sixth part contains papers describing new computational intelligence algorithms and their applications in different areas of intelligent control. The seventh part contains papers that present the use of fuzzy logic in different mathematic models. The eight part deals with a diverse range of applications of fuzzy logic, ranging from environmental to autonomous navigation. The ninth part deals with theoretical concepts of fuzzy models.

In the first part of theoretical aspects of type-1 fuzzy logic, there are four papers that describe different contributions that propose new models, concepts, and algorithms centered on type-1 fuzzy systems. The aim of using fuzzy logic is to provide uncertainty management in modeling complex problems.

In the second part of type-2 fuzzy logic theory and applications, there are four papers that describe different contributions that propose new models, concepts, and algorithms centered on type-2 fuzzy systems. There are also papers that describe different contributions on the application of these kinds of type-2 fuzzy systems to solve complex real-world problems, such as time series prediction, medical diagnosis, and pattern recognition.

In the third part of fuzzy logic for the augmentation of nature-inspired optimization meta-heuristics, there are six papers that describe different contributions that propose new models and concepts, which can be considered as the basis for enhancing nature-inspired algorithms with fuzzy logic. The aim of using fuzzy logic is to provide dynamic adaptation capabilities to the optimization algorithms, and this is illustrated with the cases of the bat algorithm, harmony search, and other methods. The nature-inspired methods include variations of ant colony optimization, particle swarm optimization, the bat algorithm, as well as new nature-inspired paradigms.

In the fourth part of emergent intelligent models, there are six papers that describe different contributions on the application of these kinds of models to solve complex real-world optimization problems, such as time series prediction, robotics, and pattern recognition.

In the fifth part of fuzzy logic applications in medicine, there are three papers that describe different contributions on the application of these kinds of fuzzy logic models to solve complex real-world problems, such as medical diagnosis.

In the sixth part of intelligent control, there are six papers that describe different contributions that propose new models, concepts, and algorithms for designing intelligent controllers for different plants. The aim of using these algorithms is to provide methods and solution to some real-world problem control areas, such as scheduling, planning, and robotics.

In the seventh part, there are five papers that are presenting the application of fuzzy logic in different mathematical models. There are also papers that describe different contributions on the application of these kinds of fuzzy models to solve complex real-world problems, such as in intelligent control.

In the eighth part, there are four papers dealing with applications of fuzzy logic, like in diagnosing air quality or vehicle navigation. In addition, theoretical contributions are presented in regard to how we can apply fuzzy logic.

Finally, in the ninth part, there are six papers presenting theoretical concepts of fuzzy models. The concepts range from fuzzy linear programming to fuzzy restricted Boltzmann machines.

In conclusion, the edited book comprises papers on diverse aspects of fuzzy logic, neural networks, and nature-inspired optimization meta-heuristics and their application in areas such as intelligent control and robotics, pattern recognition, time series prediction, and optimization of complex problems. There are theoretical aspects as well as application papers.

June 2017

Patricia Melin
Oscar Castillo
Janusz Kacprzyk
Marek Reformat
William Melek

Contents

Fuzzy Logic in Metaheuristics

Emergent Intelligent Models

Theoretical Concepts of Fuzzy Models

Theoretical Aspects of Fuzzy Logic

Can Multi-constraint Fuzzy Optimization *Bring* Complex Problems in Selecting Optimal Solar Power Generating System *into Focus*?

Akash Dand[1(✉)], Chetankumar Patil[1], and Ashok Deshpande[2,3]

[1] Department of Instrumentation and Control,
College of Engineering Pune, Pune, India
akash.dand@gmail.com, cyp.instru@coep.ac.in
[2] Berkeley Initiative Soft Computing (BISC)-Special Interest Group (SIG),
Environment Management System (EMS),
University of California, Berkeley, USA
ashok_deshpande@hotmail.com
[3] College of Engineering Pune (COEP), Pune, India

Abstract. The debate on greenhouse gases (GHS), emissions from polluting sources & its health effects, climate Change, increased energy needs, and the use of non-conventional/renewable energy sources has reached a steady state. In country like India, apart from solar, high energy wind sources could also be used in selected locations. Therefore, not only renewable energy but energy mix is a viable proposition to meet increased energy needs. Though solar panels are installed all over the world to meet ever increasing energy needs and reduce carbon footprints, selection of Optimal Solar Power Generating System is a complex issue and could be labeled as multi constraints fuzzy optimization problem. The paper presents a novel method with a case study to address the issue of the optimal election strategy of solar energy system.

Keywords: Energy needs · Solar power generating system · Experts' knowledgebase · Cosine amplitude method · Goal · Multi-constraint fuzzy optimization

1 Introduction

Ever increasing energy needs and progressive depletion of natural resources, call for the use of renewable and nonpolluting energy. In country like India, apart from solar, high energy wind source could also be used in selected locations. In summary, not only renewable energy but energy mix is a viable proposition to meet increased energy needs. There are concerted efforts being made globally on solar energy. In this paper, we have made an attempt to address the issue based on optimal ranking of solar power generating system.

P. Melin et al. (eds.), *Fuzzy Logic in Intelligent System Design*,
Advances in Intelligent Systems and Computing 648, DOI 10.1007/978-3-319-67137-6_1

1.1 Commentary on Solar Power Generating System and Selection

Installation of solar panels is practiced all over the world. It has been observed that if solar panel faces sun perpendicularly, then it might gain the best power output. To face the panels throughout the day, there is a need of sun tracking systems (STS). Some of the researchers [1–4] are in favor of sun tracked solar panels which might give better power generation over fixed panels installation. Broadly speaking, STS [4–6] can be classified as single axis or dual axis and could be sensor based or time based. Should we go for sun tracking or not? The debate is on in the researcher community and industry experts. It is largely believed that for different longitude and latitude, different sun tracking system could be used [1, 3]. The selection process of STS invariably depends on multiple constraints such as cost, complexity, amount of power generation in different seasons of the year and area required for the installation, and alike. It can be argued in no uncertain terms that most of the constraints in decision making of such systems are imprecise/fuzzy. Decision processes with which fuzziness can be evaluated from many point of views [7–9]. The authors present the application of fuzzy logic based algorithm proposed by Bellman-Zadeh in their seminal paper [10] for the selection of optimal solar power generating system.

1.2 Objective

The overall objective is application of multi-constrain fuzzy optimization formalism in selecting optimal solar power generating system, while sub objective is to workout similarity of the domain experts as their belief/perception is used in fuzzy optimization algorithm.

2 Case Study

The case study relates to arrive at the optimal solar power generating system based on the experimental set up is installed at out the College of Engineering Pune (COEP) Pune India, located on longitude 18.5204° N and latitude 73.8567° E.

Figure 1 shows a dome like structure with 45 solar cells in series solar panels. in addition, there are traditional fixed solar panels and single axis tracking system.

Single axis sun tracking system is design with 25 solar cells of size 165 mm * 165 mm in series. Due to more area requirement in single axis, 36 solar cells are mounted as fixed panel in same footprint area.

All experimental solar power generating systems are connected to same quantity load. Voltage across load (V) and current flowing through load (I) is measured. From voltage and current, power generation calculations could be made using:

$$P = V * I \tag{1}$$

Power generation for one complete day was measured and is referred as a Goal in fuzzy optimization while the constraint could be cost, complexity in operation and

Fig. 1. Experimental setup – (a) Fixed panel, (b) Single axis STS and (c) Dome structure

maintenance and area required for the installation are the constraints used in Bellman - Zadeh formulation. Figure 1 shows the Installed Experimental setup.

2.1 Expert Knowledgebase

Authors have created domain expert's knowledgebase (assumed it as membership value based on partial belief concept used in fuzzy set theory for the constraints (cost, operation & maintenance, complexity and footprint area, and season independency). Out of 11 experts, 3 are from industry, 3 energy consultants and 5 research scholars working in field of solar power.

In summary, the opinion of all the experts will be considered as the constraints in Optimal ranking of solar power generating system using Bellman-Zadeh formalism. For dome structure, in the absence of bba/perception (membership values) for constrains, the authors have assumed the following values: 1. Cost 0.45, 2. Operation Maintenance 0.5, 3. Complexity 0.4, 4. Footprint area 0.65, 5. Season Independency 0.7.

3 Results and Discussion

3.1 Similarity Between Experts

Table 1 presents the membership grade of domain experts which is on two universes (membership value in row while column vector is expert). In order to compute similarity within and between the experts, the authors have used Cosine amplitude method [11], r_{ij} is similarity between expert i and j; k = 1…n are expert.

Table 1. Experts data as constraints in fuzzy optimization

	Fixed panels					Single axis				
	1	2	3	4	5	1	2	3	4	5
Expert1	0.1	0.45	0.2	0.5	0.45	0.45	0.55	0.5	0.45	0.6
Expert2	0.2	0.5	0.15	0.4	0.6	0.7	0.35	0.5	0.45	0.8
Expert11	0.4	0.45	0.2	0.7	0.45	0.6	0.55	0.7	0.5	0.65

$$r_{ij} = \frac{\sum_{k=1}^{n} x_{ik} * x_{jk}}{\sqrt{\left(\left(\sum_{k=1}^{n} x^2 ik\right)\left(\sum_{k=1}^{n} x^2 jk\right)\right)}} \tag{2}$$

The results in matrix form is invariably fuzzy tolerance relation which has been transformed to fuzzy equivalence relation using transitivity closure using (3) and dendrogram for various α cut values was drawn.

$$\mathrm{R}^{n-1} = \mathrm{R} \circ \mathrm{R} \circ \mathrm{R} \circ \mathrm{R} = \mathrm{R} \tag{3}$$

It can be inferred all experts agrees at 0.96 possibilities (Fig. 2).

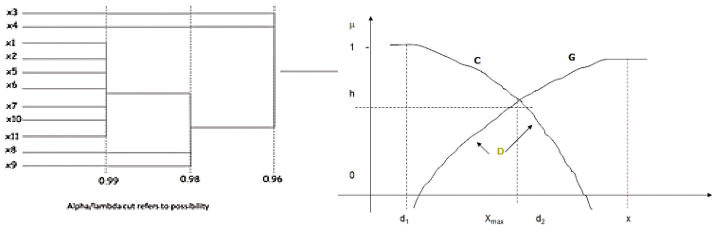

Fig. 2. (a) Dendrogram for various alpha cut values (b) Fuzzy goal G, constraint C, decision D, optimal decision Xmax

3.2 Multi Constraint Fuzzy Optimization

Optimization result will give optimal sun tracking system. Consider fuzzy sets G (goal) and C (constraint) with membership function μG (x); and μC (x), where x is an element of the crisp set of alternatives. Let fuzzy set D as decision with membership function μD (x). This will result in multiple decision from alternatives. Using the membership functions as an operation-intersection [10].

$$\mu_D(x) = \min(\mu_G(x); \mu_C(x)) \tag{4}$$

Mostly decision need to be in crisp and this requires defuzzification of D. It is natural to adopt for that purpose the value x from the selected set [d1; d2] with the highest degree of membership in the set D. That is maximizing decision μD (x).

$$x_{max} = \{x | \max \mu_D(x) = \max \min (\mu_G(x); \; \mu_C(x))\} \tag{5}$$

The experiments were carried out from 24 May 2017 to 2 June 2017. Power generation of 25 cell fixed panels, 36 cells fixed panels, Dome and single axis sun tracking system was measured for 10 days. Power generation values were normalized and membership grade for the Goal (G) was worked out. Table 2 represents the statistical analysis of 10-day power generation in watts. In fuzzy optimization, authors have use 36 cells fixed panel system as footprint area is the constraint (Fig. 3).

Table 2. 10 days total power generation data

	25 cell Fixed panel	36 cell Fixed panel	Dome	Single axis sun tracker
Minimum	251	324	288	299
Maximum	295	381	335	351
Mean (μ)	269.7	349.4	308.5	322.5
S.D. (σ)	12.6	16.4	14.1	15.2
95% confidence level μ ± 2σ	244–295	316–382	280–337	292–353

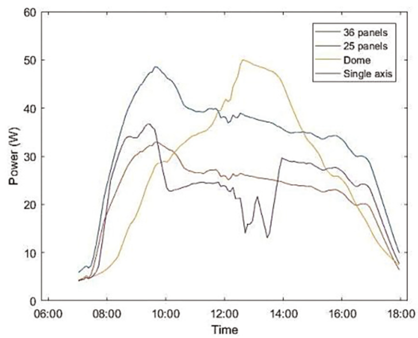

Fig. 3. Power generation 29 May 2017, 36 cell, 25 cell, Dome and Single axis sun tracker

A typical computation for optimal ranking (based on observation May 29, 2017) is C1 = {0.31, 0.45, 0.52}, C2 = {0.46, 0.5, 0.50}, C3 = {0.07, 0.4, 0.55}, C4 = {0.57, 0.65, 0.32}, C5 = {0.47, 0.7, 0.66} and G = {0.55, 0.52, 0.5}

$$max\begin{pmatrix} min(0.55, 0.31, 0.46, 0.5, 0.57, 0.47), \\ min(0.52, 0.45, 0.5, 0.25, 0.65, 0.7), \\ min(0.5, 0.52, 0.50, 0.54, 0.2, 0.66) \end{pmatrix} = max\begin{pmatrix} \mathbf{0.31}, \\ 0.25, \\ 0.2 \end{pmatrix} = 0.31 \quad (6)$$

0.31 refers to membership value of fixed panel. **It can be stated that in this particular case optimal sun power generation system is Fixed panel.**

According to geographical location of Pune, in a year, 60 days are assumed to be cloudy or partially cloudy days. In all conditions Dome gives better results than single axis sun tracking system. Single axis sun tracking system generates 322 W in a day in which it consumes 48 W in rotation of panels, so all over it power output is 20% to 40% less then Dome. Comparing from cost point of view; Single axis system is costlier as cost of motor and rotating design is more which is approximately twice of the panel cost. But Dome is less costly, due to no rotating parts and with much lesser maintenance. Dome requires 20% extra solar cells, but as cost of solar cells are decreasing every day, the total cost of design will come down in future - this may not be a case with single axis system as cost of motor, bearing and allied mechanical components have reached steady state and will tend to increase in future. Because of more area requirement for Dome configuration, Fixed panel system is preferred to Dome. If we consider same footprint area in which 25 cell Single axis sun tracking system is designed; Dome can accommodate 45 solar cells. while in fixed panel system 36 cells can be installed. Cost of fixed system is less and maintenance is very low.

Concluding Remarks
Decision making in fuzzy environment is demonstrated in possible selection of *optimal* solar power generating system. The Authors believe that more studies in this regard should be carried out. However, approach delineated in the paper can be used in any other system having different goals and constraints.

References

1. Abu-Khader, M.M., Badran, O.O., Abdallah, S.: Evaluating multi-axes sun-tracking system at different modes of operation in jordan. Renew. Sustain. Energy Rev. **12**(3), 864–873 (2008)
2. Suthar, M., Singh, G.K., Saini, R.P.: Performance evaluation of sun tracking photovoltaic systems: a case study. In: Fifth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2013), Bangalore, pp. 328–335 (2013)
3. Mehrtash, M., Quesada, G., Dutil, Y., Rousse, D.: Performance evaluation of sun tracking photovoltaic systems in Canada. In: ISME 2012, Shiraz, Iran (2012)
4. Kumar, M.N., Singh, K., Anjaneyulu, K.S.R.: Solar power analysis based on tracking simulation. Int. Electr. Eng. J. (IEEJ) **5**(5), 1398–1403 (2014)
5. Mousazadeh, H., Keyhani, A., Javadi, A., Mobli, H., Abrinia, K., Sharifi, A.: A review of principle and sun-tracking methods for maximizing solar systems output. Renew. Sustain. Energy Rev. **13**(8), 1800–1818 (2009)
6. Ponniran, A., Hashim, A., Joret, A.: A design of low power single axis solar tracking system regardless of motor speed. Int. J. Integr. Eng. **3**(2), 5–9 (2011)
7. Chang, S.S.L.: Fuzzy dynamic programming and the decision making process. In: Proceedings of 3D Princeton Conference on Information Sciences and Systems, pp. 200–203 (1969)
8. Fu, K.S., Li, T.J.: On the behavior of learning automata and its applications. Technical report TR-EE 68-20, Purdue University, Lafayette, Indiana, August 1968
9. Zadeh, L.: Fuzzy algorithms. Inf. Control **12**, 94–102 (1968)
10. Bellman, R., Zadeh, L.: Decision making in a fuzzy environment. J. Manag. Sci. **17**, 141–164 (1970)
11. Ross, T.: Fuzzy Logic with Engineering Applications, 3rd edn, pp. 69–71. Wiley, Hoboken (2010)

Relating Fuzzy Set Similarity Measures

Valerie Cross(✉)

Computer Science and Software Engineering,
Miami University, Oxford, OH 45056, USA
crossv@miamioh.edu

Abstract. Measuring similarity is an important task in many domains such as psychology, taxonomy, information retrieval, image processing, bioinformatics, and so on. The diversity of domains has led to many different definitions of and methods for determining similarity. Even within fuzzy set theory, how to measure similarity between fuzzy sets presents a wide variety of approaches depending on what characteristic of a fuzzy set is emphasized, for example, set-based, logic-based or geometric-based views of a fuzzy set. First similarity is examined from a psychological viewpoint, and how that perspective might be applicable to fuzzy set similarity measures is explored. Then two fuzzy set similarity measures, one set-based and the other geometric-based, are reviewed, and a comparison is made between the two.

Keywords: Fuzzy set similarity · Set-based similarity · Geometric-based similarity · Dissemblance index

1 Introduction

Comparing two concepts or objects is a necessary process in many domains such as biology, psychology, taxonomy, statistics and artificial intelligence. This comparison operation attempts to determine a relationship between the two concepts. One such type of relationship that is frequently determined is their similarity. Because of the diversity of domains, the general meaning of similarity is ambiguous with many different definitions and approaches to measuring similarity. As presented in psychological theory [1], a warning on assessing similarity is given, “Like most powerful and widespread ideas, it [similarity] is not amendable to a ready and precise definition; indeed, this very resistance to definition probably goes far to explain its usefulness as a supposed explanatory principle. Ideas that are imprecise are also dangerously versatile when it comes to accounting for the complexities of human behavior.”

Even within one domain such as fuzzy set theory, a wide variety of methods exist for assessing similarity [2], many of which are extensions of similarity measures that are well-known in their respective research domains. The more recent research area of ontological knowledge representation for the Semantic Web has also had a proliferation of semantic similarity measures for various tasks such as ontology alignment, information extraction, and semantic annotation. The objective of a semantic similarity measure, also referred to as an ontological similarity measure, is to calculate the degree to which one concept is similar to another concept within the context of an ontology.

P. Melin et al. (eds.), *Fuzzy Logic in Intelligent System Design*,
Advances in Intelligent Systems and Computing 648, DOI 10.1007/978-3-319-67137-6_2

Although several major categories of semantic similarity measures exist such as path-based, information content-based and feature-based, those measures using information content have been the emphasis of much study and evaluation especially in the bioinformatics and biomedical domains. In [3] many of these semantic similarity measures are shown as related to fuzzy set similarity measures if a concept is represented as a fuzzy set consisting of itself and all its ancestor concepts and the membership degrees are based on the information content of each concept within the context of the ontology.

The focus of this paper is that of similarity in fuzzy set theory. This paper examines some "respects for similarity" [4] from the domain of psychological theory and their general applicability to the measurement of similarity in fuzzy set theory. "Respects for similarity" refers to the ways in which two things can be similar. The term frame of reference [5] is also used for respects for similarity. The comparison process has intrinsic factors that determine the respects. As pointed out in [4], asking "How similar are X and Y?" can be viewed as asking a slightly different question, "How are X and Y similar? The process for fixing the respects is a crucial facet of similarity comparisons.

Correspondingly, similarity in other domains is investigated to better understand how fuzzy set similarity measures have been extended from these domains. Two specific similarity measures, one used very early in taxonomy and the other used in calculating distances between intervals on the real line are reviewed and their fuzzy set extensions analyzed. These two similarity measures are compared to determine any relationships between them. To begin, Sect. 2 looks at similarity as an empirical and theoretical psychological construct and attempts to elicit correspondences to fuzzy set similarity. These correspondences might suggest other views and uses for fuzzy set similarity measures. Just like different characteristics of a concept in the context of an ontology are considered important to constructing a semantic similarity measure, a variety of characteristics of a fuzzy set are considered in the construction of a fuzzy set similarity measure. Section 3 presents the taxonomic related fuzzy set similarity measures and its relation to Tversky's psychological model of similarity. The fuzzy extension of the distance between real number intervals to the similarity between fuzzy set intervals is described in Sect. 4. Section 5 compares and contrasts these two fuzzy set similarity measures and establishes a relationship between them. A summary and plans for future research are provided in Sect. 6.

2 Respects for Similarity

In [4] the researchers examine similarity as an explanatory construct in psychological theory where humans are comparing two objects or things. The things being compared ranged from two simple linguistic terms to two visual forms. Their experiments indicate that similarity is highly flexible and in some ways troublingly flexible. Their experimental observations, however, are used to argue that the flexibility is reasonable as long as systematic changes in the process of similarity assessment can be established.

The research of Tversky [5] has played a major role in shaping the understanding of similarity in psychological research. Tversky's research informs the research in [6] where it is noted that "the relative weighting of a feature (as well as the relative importance of common and distinctive features) varies with the stimulus context and the task; so that there is no unique answer to the questions of how similar is one object to another." This quote on similarity again emphasizes its "resistance to definition" and that its assessment method is "dangerously versatile."

The argument [4] is made that instead of viewing similarity assessment as constrained by the perceptual process, similarity assessment is flexible and the comparison process itself methodically sets the respects. Assessing similarity is assumed to be based on matching and mismatching of properties. Things are similar to the degree they share properties and dissimilar to the degree that properties apply to one but not the other. The issue is that two things share a subjective number of properties and likewise they differ in a subjective number of properties. Before similarity can be computed, a prior process must occur that determines what properties are to be used in the similarity computation.

Others [7] argue that the respects for determining how two things are judged as similar are set not by the comparison process but by the goals motivating the comparison process. This view is the result on research to determine the requirements for a similarity measure for use in the automatic generation of textual comparisons. Comparison between objects is categorized into six different types. For example, a clarificatory comparison is a domain-based comparison with the goal of distinguishing one object from another object that is highly similar to it. Domain-based comparisons are used to establish explicit relationships between an object and other objects existing in the same domain.

Regardless of how and when respects are established, most agree that similarity assessment cannot be performed without them. The problem still remains as to the process of selecting the respects as so aptly described by Tversky [5], "When faced with the a particular task (e.g. identification or similarity assessment) we extract and compile from our data base a limited set of relevant features on the basis of which we perform the required task."

Another important issue discussed in [4] is the effect of context in similarity assessment. Setting the context for comparison contributes to the selection of the respects to which similarity is being assessed. Two objects may be judged less similar when no explicit context is given than when one is given because the context tends to make salient the context-relevant properties to be used in the similarity assessment. The similarity of the two objects is increased based on the degree to which the two objects share values for these now salient properties.

The extension effect of context is also important in similarity assessment. When in one context, properties that are shared by all objects are not useful in similarity assessment; however, if the context is extended or broadened to include objects not sharing these properties, then these properties become more salient. In the extended context, two objects sharing those properties are perceived as more similar than they were in the original context. To summarize, depending on the context, two objects may vary in their similarity, but this variability becomes systematic when incorporated into the specification of a similarity comparison.

Analogy also plays a role in similarity assessment. Instead of focusing on similarity in values for simple properties of objects, it looks for relational or structural similarities. An example given in [4] is "an atom is like the solar system" where the analogy relies on relations such as "revolves around" and not property values such as "hot" or "yellow". The research in [4] argues the importance of incorporating relational structure since relational structures can significantly affect the process of determining the correspondences between objects when assessing similarity.

Similarity assessment involves comparing two objects but this comparison process may be directional. The example given in [4] is very informative: "surgeons are like butchers" as compared to "butchers are like surgeons". The former is critical of surgeons whereas the latter is favorable of butchers. In the contrast model of similarity [5], the less salient or less prominent object is compared to the more salient or prominent object as evidence by the results of human experiments where the less salient object is consider more similar to the more salient object than vice versa.

The direction of comparison also affects the properties selected for assessing similarity as shown in their experimental results [4]. The selected properties may be more closely related to the base object to which a comparison is being made. The common properties used in comparing two objects may vary as a function of the direction of a comparison and the bias is to select properties more strongly associated with the base object. To summarize, similarity is more than identity since similarity comparisons may encompass properties of one object becoming the candidate properties of the other in performing the similarity assessment.

Since similarity assessment usually involves multiple properties, research in cognitive psychology has concentrated on how multiple pieces of information are integrated into a single assessment of similarity. Similarity assessment is affected by both the selection of the applicable properties and the constraints that the integration method places on the process. As part of the integration method, weighting may be involved that favors certain properties over others. In [4] experiments have shown that this weighting procedure is not independent of the outcome of the comparison process.

One last interesting aspect brought out in [4] is the notion of experience and learning affecting the process of similarity assessment. Children, for example, judge similarity in a more holistic manner and are less like to analyze individual components, but as they mature, they base their similarity judgements more on abstract, relational, and less on superficial properties.

To summarize the research in [4] for the domain of psychology, similarity assessment is dynamic and highly variable but connected to the details of the comparison process. The details that are focused on in their research are the fixing of the properties or respects to which objects are similar, the context, the direction of the comparison, the kind of properties whether simple attributes or relational structures, the integration of multiple information and the weighting of this information in the process and human experience. Many of these details of the comparison process in similarity assessment can be found in fuzzy set similarity measurements.

In the following two sections, a set based fuzzy similarity measure from taxonomy related its related measures and a then a geometric based fuzzy set similarity measure from distance between real line intervals are described. Their details are examined from the viewpoint of similarity assessment in the domain of psychology.

3 Set-Based Similarity

One of the early set based similarity measures for crisp sets is the Jaccard index [8], which was used in taxonomic classification. A specimen is represented by a set of attributes describing it. Two specimens are judged to be similar based on the similarity between their set of attributes. In taxonomy, the Jaccard index has also been referred to as the "coefficient of similarity" [9] and in psychology, it is the unparameterized ratio model of similarity [5]. Its formula where X and Y are sets is expressed as

$$S_{jaccard}(X,Y) = \frac{f(X \cap Y)}{f(X \cup Y)}. \tag{1}$$

The function f is an additive function and is typically the cardinality of the set. The Jaccard index is easily extended when X and Y are fuzzy sets by using fuzzy set operators to perform the intersection and the union on the two fuzzy sets and the function f is fuzzy set cardinality, which is simply the sum of the membership degrees for all elements in the fuzzy set. A fuzzy Jaccard dissimilarity measure can be derived by subtracting the Jaccard similarity from 1, i.e., $D_J = 1 - S_{jaccard}(X,Y)$.

From the psychological analysis of similarity assessment, the fuzzy sets X and Y are being compared based not only on the elements making up each set but the degree of membership of each element in the set. The selection of properties in this similarity measure is natural; that is, all elements in the support of a fuzzy set describe it. The selected properties for the comparison process, therefore, include both the support of X and the support of Y.

The correspondence or alignment between the properties of the two fuzzy sets is automatic since each element in the fuzzy set is considered a property and the constraint on a fuzzy intersection is an exact match on each element in the intersection. The weighting, however, for a property (element) in this comparison process is its degree of membership or agreement with the fuzzy concept being represented by the fuzzy set. In addition to the required exact match on the aligned property values is the constraint on the integration between their two membership degrees using a fuzzy set intersection operator, which is typically *min*. The result is that multiple pieces of information exist since there are multiple elements (properties) and further integration, referred to as aggregation in fuzzy set theory, must occur to assess the overall similarity of the two fuzzy sets. With the Jaccard index, the aggregation operator is summation, that is, the cardinality of the fuzzy set intersection.

The numerator of the Jaccard index provides an assessment of the agreement of properties between the two fuzzy sets but does not take into consideration, properties in one fuzzy set that are not contained in the other fuzzy set and vice versa. The denominator, which is the union of the fuzzy sets, typically using the *max* operator, does consider this and thus normalizes the overall similarity assessment in [0, 1].

Psychological similarity considers direction of comparison as a critical aspect in the process. The Jaccard index does not account for presupposing a direction for the comparison. An inclusion index, however, can and is a version of the parameterized ratio model of similarity [5], which is given as

$$S_{Tverksy-ratio}(X,Y) = \frac{f(X \cap Y)}{f(X \cap Y) + \alpha f(\mathrm{X} - \mathrm{Y}) + \beta f(Y - X)}. \tag{2}$$

where *(X − Y)* is set difference operator. Setting $\alpha = 1$, $\beta = 1$ produces the Jaccard index. Setting $\alpha = 1$, $\beta = 0$ produces the degree of inclusion for *X*, that is, the proportion of *X* overlapping with *Y,* given as

$$S_{inclusion}(X,Y) = \frac{f(X \cap Y)}{f(X)}. \tag{3}$$

In the parameterized ratio model, the value *f(X)* for object *x* is considered a measure of the overall salience of that object. In psychology, the factors adding to an object's salience include "intensity, frequency, familiarity, good form, and informational content" [5]. Although the cardinality of a fuzzy set is a very simple way to measure the "salience" of a fuzzy set, i.e., the larger the cardinality, the less salience, other ways might be more useful depending on the application. Both fuzzy entropy [10] and a function of the distance of a set to its complement [11] have been used as fuzziness measures. One could consider that a fuzzy set is more salient than another fuzzy set if it has less fuzziness.

For fuzzy rule-based reasoning systems, salience of the two fuzzy sets being compared is not relevant. One approach that is used is to set the comparison direction from the observation fuzzy set as compared to the rule antecedent fuzzy set, which becomes the base for comparison to. The objective is to determine how certain is it that the observation satisfies the antecedent. The more the observation is included within the antecedent, the more certain that the antecedent is satisfied. If the observation fuzzy set is a subset of the antecedent fuzzy set, the inclusion measure produces a one. Not every fuzzy rule base system, however, uses an inclusion measure to assess agreement between the rule antecedent and the observation fuzzy sets.

In fuzzy applications that are to mimic human directional comparison judgments, the use of the more salient fuzzy set as the base for comparison might be more appropriate. Here the properties of the more salient fuzzy set *S* become the selected properties for the comparison process, and those properties in the less salient fuzzy *L* set that are not in *S* are simply ignored. Here similarity is more than an identify as described in [4]. In this use of similarity, the inclusion index measures the proportion of the properties of *S* found in *L* to all the properties of *S* and is given as

$$S_{inclusion}(S,L) = \frac{f(S \cap L)}{f(S)}. \tag{4}$$

Image processing applications [12] using fuzzy set theory tested two different versions of the inclusion index with other fuzzy set similarity measures in a shape classification experiment. The denominator of the inclusion index is replaced by either *min(f(X), f(Y))* or *max(f(X), f(Y))*. In the experimental results, the error rate for the *max* version of the denominator were much smaller than that of the *min* version. The comparison direction and how to choose that direction makes a difference and is application dependent.

For an example of the extension effect discussed with psychological research on similarity, consider another fuzzy set similarity measure used in [12] which follows the formula of the Jaccard index but instead measures the similarity between the complements of the fuzzy sets, i.e. *X'* and *Y'*. The more the complements of the two sets are similar, then the more the two fuzzy sets are similar. With this approach, if the context or the universe of discourse for the two fuzzy sets is extended, i.e., its size increased, then the Jaccard index for the two fuzzy sets would not be affected by the extension since the properties considered salient would still be those in the union of the two fuzzy sets. The Jaccard index as measured using the complements of the two fuzzy sets, however, would be affected and would be greater in the extended context than in the original context. Intuitively, in the extended context the complements of the fuzzy sets share more properties.

4 Geometric-Based Similarity

Geometric based similarity relies on the dissemblance index, which provides a normalized distance between two real intervals. If V = [v1, v2] and W = [w1, w2], the dissemblance index is given as

$$D(V, W) = \frac{(|v1 - v2| + |w1 - w2|)}{2 * (\beta_2 - \beta_1)}. \tag{5}$$

where $[\beta_1, \beta_2]$ is the smallest interval that contains both the V and W intervals. The factor $2 * (\beta_2 - \beta_1)$ is necessary to produce a normalized dissemblance in [0, 1].

The dissemblance index consists of two components, the left and right distance between the two intervals and may be generalized to fuzzy intervals. A pair of boundary functions L_N and R_N and parameters $(r_1, r_2, \lambda, \rho)$ define a fuzzy interval. The core of N is $[r_1, r_2]$ and λ and ρ are parameters of the boundary functions L_N and R_N such that the support of N is in the interval $[r_1 - \lambda, r_2 + \rho]$. If L_N and R_N are positively and negatively sloping linear functions, respectively, then N is represented by a trapezoidal fuzzy set membership function. Figure 1 illustrates two fuzzy trapezoidal fuzzy sets X and Y and labels for left and right boundaries.

To calculate the fuzzy dissemblance index between two fuzzy intervals X and Y, the formula uses integration over the α-cuts of the fuzzy intervals as

$$fD(X, Y) = \frac{1}{2(\beta_2 - \beta_1)} \int_0^1 (|L_X(\alpha) - L_Y(\alpha)| + |R_X(\alpha) - R_Y(\alpha)|) d\alpha. \tag{6}$$

where $[\beta_1, \beta_2]$ is the smallest interval that contains both the support of the X and Y fuzzy intervals. *FD* calculates a fuzzy dissimilarity measure between two fuzzy intervals based on a normalized distance and can be converted into a fuzzy similarity measure as $S_{fD}(X, Y) = 1 - fD(X, Y)$.

With the fuzzy similarity measure S_{fD}, also referred to as a geometric fuzzy similarity [2], the alignment between properties is not based on identical property values as for the Jaccard fuzzy similarity measure but on identical α values. The comparison is measured between the property values at the identical α values for the left and the right components of the fuzzy interval. This geometric similarity differs from the Jaccard fuzzy similarity measure since the comparison is done on the α values and resolved using a fuzzy set intersection operator. Correspondingly, both have a normalizing factor that includes the support of both X and Y.

Some similarity research have been proposed to approximate *fD* to avoid the computationally expensive integration over α, the value [13]. These approximations use only the distance obtained from a single a-cut, for example, only the distance between the core intervals of the fuzzy sets. This approximation does not incorporate information about the proximity of the support intervals. Thus, the approximation result may be much smaller than *fD*. A summarization technique was introduced in [14]. First, the distance between the support intervals is determined as

$$fD_0(X,Y) = \frac{1}{2(\beta_2 - \beta_1)}(|L_X(0) - L_Y(0)| + |R_X(0) - R_Y(0)|) \tag{7}$$

and similarly for the core intervals, fD_1. The summarized distance is the average core and support distances given as

$$fD_{@} = \frac{fD_0 + fD_1}{2}. \tag{8}$$

For trapezoidal fuzzy sets in which L_X does not intersect L_Y and R_X does not intersect R_Y, this summarization technique produces equivalent results as *fD*. When L_X does intersect L_Y at a_L the left distance for the support interval must be factored by a_L and the left distance for the core interval must be factored by $(1 - a_L)$ and similarly if R_X does intersect R_Y at a_R. This factor represents the height of the triangle created at the intersections.

The geometric fuzzy similarity measure $S_{diss}(X,Y) = 1 - fD(X,Y)$ and its use in fuzzy reasoning is presented in [14] since using this distance based measure allows a fuzzy conclusion to be determined using the left and right distances between the fuzzy rule antecedent and the fuzzy observation even when there is no overlap between the two. The details of this fuzzy reasoning approach are not examined here but instead a relationship between the fuzzy Jaccard similarity and the fuzzy geometric similarity measures are explored.

5 Relating Set and Geometric Similarity

When extending the similarity measures of psychology and taxonomy to similarity measures for fuzzy sets, it is natural to see how features of objects are replaced by elements of the fuzzy sets, crisp set cardinality replaced with fuzzy set cardinality and set operators replaced with fuzzy set operators. However, not all equalities using crisp set operators are true for all possible fuzzy set operators. For example, when X and Y are crisp sets and not disjoint, $f(X \cup Y) = f(X) + f(Y) - f(X \cap Y)$. This equality is true for fuzzy sets only when members of Frank's family of dual t-norms and t-conorms [15] are selected for the union and intersection operators. A more methodical method of creating a framework for fuzzy set similarity measures is based on developing a set of properties that they should satisfy. In order to develop the relationship between the Jaccard and geometric fuzzy similarity measures the theoretical foundation for the fuzzy Jaccard similarity measure is first presented [16].

One of the properties established for a fuzzy set similarity measure between X and Y is that $S(X, Y) = 1$ if and only if the symmetric difference between the two, $(X\Delta Y)$ is the empty set. Another property is if X and Y have disjoint sets, then $S(X, Y) = 0$. To meet these conditions a fuzzy set similarity measure is derived using relative cardinality on the negation of the symmetric difference between X and Y, $g\left((X\Delta Y)'\right)$ where g is relative cardinality and

$$X\Delta Y = (X \cup Y) \cap (X' \cap Y') = (X \cap Y') \cap (X' \cap Y)$$

Here is another example of an equality being true for crisp sets but only true for fuzzy sets when minimum is used for intersection and maximum is used for union.

The fuzzy similarity measure should be in the interval [0, 1] so the range for $g\left((X\Delta Y)'\right)$ must be found to produce a normalized value. The maximum value for $g(X\Delta Y)$ occurs when the two fuzzy sets are disjoint, which is $g(X \cup Y)$. The minimum value for $g\left((X\Delta Y)'\right)$, therefore, occurs for $g\left((X \cup Y)'\right)$. The range for $g\left((X\Delta Y)'\right)$ is $[\left(g((X \cup Y)'\right), 1]$. The fuzzy similarity measure can be derived as

$$S(X, Y) = \frac{g\left((X\Delta Y)'\right) - g\left((X \cup Y)'\right)}{1 - g\left((X \cup Y)'\right)}.$$

This equation can be rewritten as

$$S(X, Y) = \frac{g(X \cup Y) - g(X\Delta Y)}{g(X \cup Y)} = 1 - \frac{g(X\Delta Y)}{g(X \cup Y)}$$

since $g(X) = 1 - g(X')$ for relative cardinality. From the above equation, the fuzzy similarity measure produces a 0 if and only if $g(X\Delta Y) = g(X \cup Y)$, that is the fuzzy sets X and Y are disjoint. The fuzzy set similarity measure produces a 1 if and only if the symmetric difference produces the empty set. When X and Y are crisp and $X = Y$, all the symmetric difference operators produce an empty set. When X and Y are fuzzy sets, however, the only symmetric difference operator to produce an empty set when $X = Y$ is derived using $(X \cap Y') \cup (X' \cap Y)$ with bold intersection, $\max(0, u_X(v) + u_Y(v) - 1)$ and bold union, $\min(1, u_X(v) + u_Y(v))$.

Using this symmetric difference operator and replacing relative cardinality with fuzzy set cardinality since the cardinality of the universe of discourse may be cancelled out in the numerator and denominator

$$\frac{g(X\Delta Y)}{g(X \cup Y)} = \frac{\sum_v \min(1, (\max(0, u_X(v) - u_Y(v)) + \max(0, u_Y(v) - u_X(v))))}{|X \cup Y|}.$$

Since the differences in the membership degrees cannot be larger than 1 the min operation can be removed to produce

$$\frac{g(X\Delta Y)}{g(X\cup Y)}=\frac{\sum_v \max(0,u_X(v)-u_Y(v))+\max(0,u_Y(v)-u_X(v))}{|X\cup Y|}$$

Since either the membership of v in X is greater than or equal to its membership in Y,

$$\frac{g(X\Delta Y)}{g(X\cup Y)}=\frac{\sum u_X(v)-\min(u_X(v),u_Y(v))+u_Y(v)-\min(u_X(v),u_Y(v))}{|X\cup Y|}$$

Now rewriting by distributing the summation operator over each component in the summation and using set intersection $\cap$ for minimum produces

$$\frac{g(X\Delta Y)}{g(X\cup Y)}=\frac{(|X|+|Y|-2|X\cap Y|)}{|X\cup Y|}=\frac{(|X\cup Y|-|X\cap Y|)}{|X\cup Y|}=1-\frac{|X\cap Y|}{|X\cup Y|}$$

since for the maximum and minimum operators, $|X|+|Y|=|X\cup Y|+|X\cap Y$, therefore, resulting in

$$S(X,Y)=1-\left(1-\frac{g(X\cap Y)}{g(X\cup Y)}\right)=\frac{g(X\cap Y)}{g(X\cup Y)}$$

which is the original proposed "similarity of coefficient" used in taxonomic classification. If the fuzzy Jaccard similarity measure is converted to a dissimilarity measure by subtracting from 1, then

$$D_J(X,Y)=\frac{g(X\Delta Y)}{g(X\cup Y)}$$

which also incorporates the symmetric difference.

There is a strong relationship between the fuzzy dissemblance measure and the Jaccard dissimilarity measure. The fuzzy distances calculated for the left and right components of dissemblance dissimilarity measure when added together include the symmetric difference between X and Y.

To establish the relationship between the two fuzzy dissimilarity measures, first consider two cases, (1) the fuzzy sets X and Y do not intersect and (2) the fuzzy sets X and Y do intersect. Case 1 is easier since when they do not intersect, $D_J(X, Y) = 1$ because the symmetric difference produces the same as the union of the two sets. Thus $fD(X,Y) \leq D_J(X, Y)$. Case 2 has two subcases: (1) the cores of the fuzzy sets intersect and (2) the cores of the fuzzy sets do not intersect.

Subcase 1 is easier since with overlap in the cores of the fuzzy sets, the dissemblance dissimilarity only includes the symmetric difference as in D_J. Thus, $fD(X, Y) \leq D_J(X, Y)$ since both have the same numerator $g(X\Delta Y)$ but the normalization factor in the denominator for *fD* is $2 * (\beta_2 - \beta_1)$ which is always greater than or equal to $|X\cup Y|$.

Subcase 2 is most difficult since when R_X (α) intersect L_Y(α) at α_I there is a distance between [$R_X(1)$, $L_Y(1)$]. Figure 1 illustrates this. The dissemblance dissimilarity measure in addition to the symmetric difference, includes this distance as twice the area of the top triangle *T* with base of ($L_Y(1) - R_X(1)$) and height of $(1 - \alpha_I)$ value since α_I is the point of intersection. The $(1 - \alpha_I)$ value represents the height of triangle since the triangle is formed above the α_I intersection point. This triangle area is included twice because both the distance between the left boundary functions of *X* and *Y* and between the right boundary functions are include this triangle area. Rewriting the fuzzy dissemblance measure and using symbol T in the equation,

$$fD(X,Y) = \frac{g(X\Delta Y) + 2 * T}{2 * (\beta_2 - \beta_1)}$$

To analyze this, the starting point is when R_X(α) intersect L_Y(α) at $\alpha_I = 0$. Since *X* and *Y* are disjoint, $fD(X,Y) \leq D_J(X, Y)$, the case 1 scenario. When R_X(α) intersects $L_Y(\alpha)$ at α_I, two triangles are formed the top triangle T and the bottom triangle *B*. The area of *B* is $|X \cap Y|$. The area of *T* is at a maximum when $\alpha_I = 0$ since its height, therefore, would be 1. However, this is case 1 and $fD(X,Y) \leq D_J(X, Y)$ for this case. As α_I increases, the area of T shrinks. As the area of the intersection grows, the corresponding area of T shrinks. In comparing to $D_J(X, Y)$, even at the maximum area for T, the fuzzy dissemblance similarity is still smaller than the fuzzy Jaccard dissimilarity measure. Twice the area of triangle T cannot produce a large enough value to cause $fD(X, Y)$ to surpass $D_J(X, Y)$.

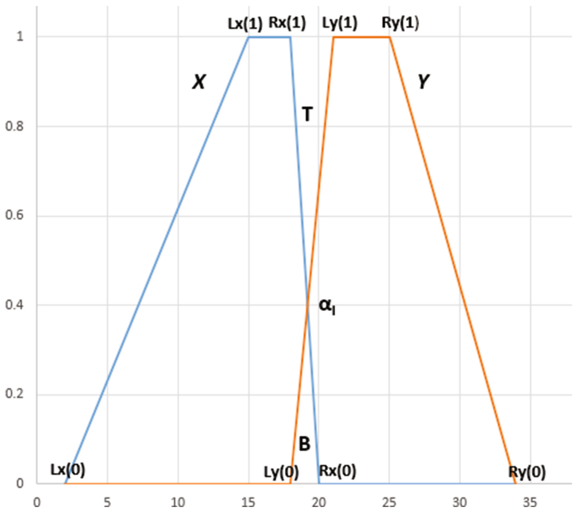

Fig. 1. Trapezoidal fuzzy sets X and Y with B intersection area and T dissemblance overlap.