International Journal of Business Intelligence Research
Volume 13 • Issue 1

Hybrid Genetic Fuzzy System for Modeling Consumer Behavior

Priti Srinivas Sajja, Sardar Patel University, India*
https://orcid.org/0000-0002-9676-0885

ABSTRACT

Understanding consumer behavior is beneficial to a business in various aspects such as prediction of manufacturing quantity, new product launch, and aids in lock-in customers and lock-out competitors. The task is highly complex, and traditional models do not help in the absence of generalized decision-making logic. Further, such domains handle large amounts of data in unstructured formats. This article presents an intelligent system for modeling consumer behavior via a hybrid genetic fuzzy system from large source of data. The paper justifies and presents a literature survey with common observations. A four-phase generic architecture of genetic fuzzy system is presented for the modeling of consumer behavior. Detailed discussion on the architecture is also provided with an experiment. Technical details, fuzzy membership functions used in experiment, encoding strategy, genetic operators, and evaluation of rules using fitness function are also discussed in detail along with results. At the end, applications of the research work in other domains are listed with possible future enhancements.

KEYWORDS

Consumer Behavior, Fitness Function, Fuzzy Membership Functions, Genetic Fuzzy System, Genetic Operators, Rule Encoding, Rule Evolution

1. INTRODUCTION

For the survival of a business, stability and growth are very important aspects. To survive is one of the most important things for a business because of frequent and unexpected turbulence in the market. Particularly for a large business handling global competitions, awareness of users and ever-increasing consumers' demands knowledge of consumer behaviour has become critical. Knowledge of consumer behaviour can help in identifying new marketing opportunities, possibilities to gain competitive advantages, help consumers in getting the product and services they desire, and add value to the business operations. In the era of information and communication technology, lots of data are generated during operation through inventory, production, marketing, sales, finance, and other related functions. These data must be utilised for the betterment of the business. It has been observed that data for transactions related to a business come from varieties of locations, devices, and formats. Further, a business generally has multiple products and services. Hence, there are a significant amount of data about consumers, product, services, policies, banking and insurance details, and other related activities for a business. These data are offline as well as online. Such data need to be effectively utilised to help managers and business professionals to identify consumer needs,

DOI: 10.4018/IJBIR.301231 *Corresponding Author

understand consumer behaviour, and strengthen the business-consumer relationship. For a business, success is more than profit; its inclination is more towards the satisfaction of the consumers, which can be enhanced by improving offerings to the consumers. By efficiently managing data related to the domain, professionals and managers can improve their practices and refine their delivery of products and services with the right product and right message to the right consumer at the right time. It will be unfortunate if such big consumer data are not further analysed to improve a business.

Issues raised related to knowledge discovery, modelling, and handling voluminous data on various web platforms can be effectively handled by intelligent techniques compared to traditional data-based systems (Akerkar & Sajja, 2016). Artificial Intelligence and Machine Learning (AI and ML) methods are useful for handling a large amount of data related to consumers, products, and transactions and discovering knowledge about consumer behaviour patterns. Such consumer behaviour patterns do not follow a generic formula and logic but often follow evolutionary and natural approaches. Further, with the use of fuzzy logic, uncertainty related to the domain can also be handled effectively, as suggested by Priti Sajja (2020). The proposed work highlights the application of the hybrid genetic fuzzy system for effective consumer modelling and knowledge discovery. A generic architecture is proposed with multiple phases. The proposed work can be used in various domains where there is a need to model consumer behaviour and discover insights about consumers. Knowledge of consumers' behaviour leads to the success of the business and strengthens the business, and often leads to other businesses. Besides tracking the possible consumers, one can identify the possible categories of the consumers and future products, improve the quality of products and services, and improve the brand image of the company.

The architecture described in this work proposes the use of a genetic algorithm to encode fuzzy rules using the binary encoding strategy and demonstrates the evolution of these rules automatically. The article also illustrates how fitness can be calculated, and the consumer behaviour pattern is stored in the form of encoded fuzzy rules. Instead of manually deducing and generalising the consumer behaviour pattern, strong, valid, and necessary fuzzy rules are automatically evolved. This article describes work done so far in the area with common observations, presents a generic architecture, example rules, fuzzy membership functions, and evolution of strong rules with fitness function. The improvement between generations is also shown using the graph showing the fitness of the populations. In the end, the article is concluded with applications and possible future enhancements to work.

2. BACKGROUND AND LIMITATIONS OF EXISTING SOLUTIONS

Many researchers have worked on modelling consumer behaviour with traditional methods as well as intelligent methods. In the work of Francesco Nicosia (1966), the consumer decision process was first highlighted. Philip Kotler (1973) suggested the use of 'Atmospherics' as a marketing tool. Further, he had also published an article on utilising consumer obsession for marketing effectiveness (Kotler, 1977). Later, various models to analyse consumer behaviour were studied in the work of Pradeep Rau and Saeed Samiee (1981). Robert Donovan and John Rossiter (1982) suggested an environmental and psychological approach for improved store atmosphere. The work of Scott Armstrong (1991) documents characteristics of consumer behaviour modelling by common people as well as by experts. Gerry Crawford and T.C. Melewar (2003) studied the purchasing behaviour of customers. Rough sets were used by Malcolm Beynon et al. to discover knowledge about consumers (Beynon, Curry, & Morgan, 2001). These research works use the traditional non-intelligent technologies so far. Major technologies used for research are traditional data-based systems, statistical, mathematical and marketing-based models.

Intelligent technology is used in work proposed by Jorge Casillas and Francisco Lopez (2009). The work uses the hybridisation of the genetic algorithm and fuzzy logic to find out consumer behaviour patterns. This paper focuses on consumer trust, overall trust, and generic consumer aptitude for shopping. However, the work focuses on the knowledge discovery from the database. Sunran Jeon and

Min-su Kim (2012) analysed various effects of the services on customers' behaviour. In their work, they have considered services offered at the airport. However, the article could not utilise modern machine learning technologies. Further, the work is very domain-specific. In the work of Urshita Dastidar et al. (Dastidar, Ambekar, Hudnurkar, & Lidbe, 2021), retail industry consumer data are analysed to enhance shopping experiences. In this work also, emphasis is given to various possibilities of consumer data analysis. Here also, modern machine learning methods have not been used.

Data mining techniques are used by S. Vijayalakhmi et al. (Vijayalakshmi, Mahalakshmi, & Magesh, 2013) to identify and model consumer behaviour in the domestic appliances market in India using k-means techniques. Consumer related knowledge is modelled and illustrated by Ali Gohary and Bahman Hamzelu (2016). The main emphasis is given to the fact that how knowledge provided by a brand is affecting the customer perception regarding the product and mainly the service quality. The latest work presented by Farida Rasiwala and Bindya Kohli (2021) discusses intelligent advisory services to learn stakeholder perception. However, the article remains silent about the machine learning techniques used in the advisory system. The use of artificial intelligence is discussed by Dipali Dalvi (2021). The work discusses how modern technology can be used to identify and co-relate prospective employees. Work presented by M. Berndtsson et al. (Berndtsson, Lennerholt, Svahn, & Larsson, 2020) also discusses the effective use of data to enhance business. Similar work was done by Rasha Hadhoud and Salameh Walid (2020). The work discusses the impact of understanding the consumers on the success of the business. Here also, limited use of artificial intelligence is observed.

Work proposed by Renu Bala and Saroj Ratnoo (2016) discusses how genetic fuzzy systems can be used to discover fuzzy classification rules for exceptions between classes and within classes of entities. The accuracy of the system varies from 86% to 96% on different data sets. However, the fitness of fuzzy rules in terms of support and confidence is not discussed here. Further, the work presented here is domain-specific and cannot be used in multiple domains. In the work of Alexis Tymkiw (2016), emotions involved while purchasing activities are analysed. Sisi Jian et al. (Jian, Rashidi, & Dixit, 2017) analysed consumer who purchases vehicle considering non-discrete extreme value modes. Generic consumer behaviour is studied by Elena Vatamanescu et al. (Vatamanescu, Nistoreanu, & Miltan, 2017). A technique like the z-number concept is also being utilised to predict consumer behaviour. This technique is used by Gunay Sadikoglu (2017). The use of machine learning techniques can be seen in the work of Saavi Stubseid and Ognjen Arandelovic (2018). Such an approach is also utilised in the medical domain for pattern classification by Chin Tan et al. (Tan, et al., 2018). The recent work includes (i) an idea to design a product that could enhance the demand using machine learning approach (Malpica & Frías, 2019); (ii) Genetic fuzzy approach applied to discover patterns and estimate 'cloacal' temperature of chicks (Hernández-Julio, et al., 2020).

The above models and techniques for modelling consumer behaviour are mainly traditional methods and domain-specific. Non-intelligent traditional methods do not provide advantages such as handling vagueness and the ability to learn from data. A few systems used fuzzy logic to handle vagueness and uncertainty (Khan & Khan, 2013), (Howells & Ertugan, 2017), (Azzini, Marrara, & Topalovic, 2019), (Moyce, 2018). However, these methods lack fine-tuning and flexibility along with evolutionary properties. Further, modern business involves lots of data, which are very difficult to handle with either traditional database/file-based generic logic or with the help of typical artificial intelligence techniques. Analysing consumer behaviour is an example case where lots of online and offline data are available; however, it is difficult to deduce generalised logic. This leads to the need for the evolutionary hybrid genetic fuzzy system for the required knowledge discovery in identifying consumer modelling patterns. The common findings from the literature survey are listed as follows.

- The non-intelligent methods used to model consumer behaviour do not offer advantages of the machine learning techniques.

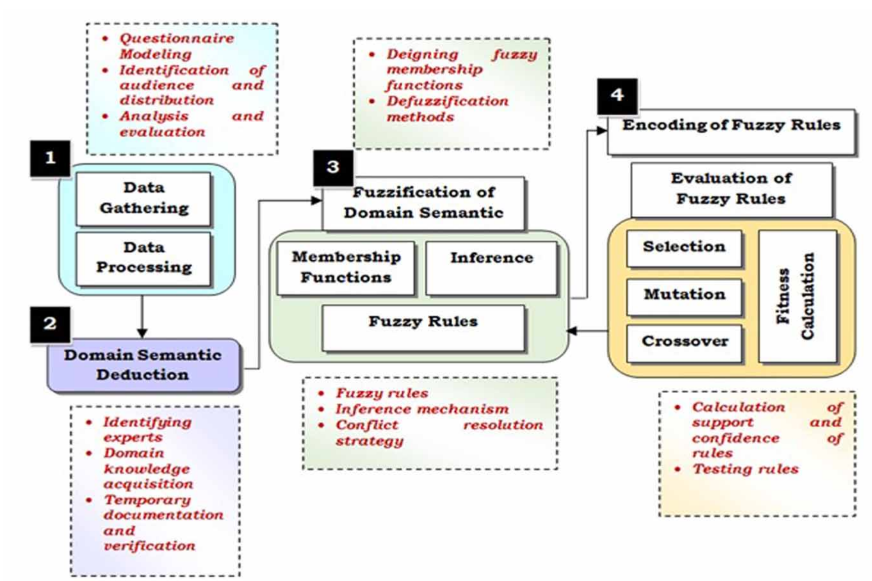
- A very few applications use genetic-fuzzy hybridisation. The systems which employ such hybrid techniques are limited to a few parameters or a few products. Further, the process of measuring the fitness of the evolved rules in terms of support and confidence is still needed to be enhanced.
- The majority of the solutions are domain-specific and have limited applications;
- etc.

Based on the findings, there is a need for a domain-independent (generic) architecture for modelling consumer behaviour and knowledge discovery. Hence, the research question is to design a generic architecture that can use a hybrid intelligent mechanism to model consumer behaviour with good quality evolved rules and test a prototype by applying domain-specific data and rules. The next section describes the methodology by discussing the architecture and its working.

3. METHODOLOGY AND ARCHITECTURE OF THE SYSTEM

To model consumer behaviour, as discussed above, the traditional artificial intelligence methods are of less use. Soft computing or modern artificial intelligence methods such as artificial neural network, fuzzy logic, and genetic algorithms are also contributing little when used stand-alone. Precisely stating, an artificial neural network can be extensively useful while lots of data are available and generalised logic is missing. The artificial neural network requires precise and normalised data. Above that, such models also document knowledge in connections and can not offer evolutionary advantages. Genetic algorithms offer evolutionary properties, however, they lack self-learning from data in the absence of generalised logic like an artificial neural network. Hence, to get the dual advantages of both techniques, a domain-independent generic architecture of a genetic fuzzy system is proposed. This architecture can later be used for modelling consumer behaviour. Figure 1 illustrates the architecture.

Figure 1. Architecture of the Genetic Fuzzy System



The architecture comprises four major phases, namely (i) data collection, deduction of domain knowledge, (iii) fuzzification of the semantics, and (iv) evolution of the rules. During the data collection phase, important parameters and metadata are identified. The data about the product, data about target consumers, data about services and service providers, etc., needed to be acquired through online or offline means. After the collection and pre-processing of necessary and valid data, domain semantic needs to be deduced with the help of experts. Here, there is a scope of adding some domain heuristics also. During the third phase, membership functions for the selected parameters are identified. This is the phase where fuzzy logic is used. Fuzzy rules can be applied by using these fuzzy membership functions. With the help of fuzzy logic, vagueness and uncertainty can be managed effectively; however, to gain evolutionary advantages, there is a need for a genetic algorithm. The fourth phase of the system employs the genetic algorithm technique to get the evolutionary advantages. The encoding, genetic operators, and fitness mechanisms are explained in detail later in this section.

The four main phases of the architecture are described below.

3.1 Phase 1: Data Collection

This phase discusses the important parameters for which data are to be collected and means of the data acquisition, possible target audience, and other contributors.

Identification of parameters: As shown in Figure 1, the first development phase starts with data collection. To collect the data, traditional models and frameworks can also be utilised. Besides means to collect data, it is also important to decide which data need to be collected. To model consumer behaviour regarding a product or service parameters such as ease of use, the robustness of the product/service, security aspects of the product, data involved in related services, social benefits, return on investment, maintenance, etc., are considered. Questions covering these parameters are to be designed with the help of domain experts, and a questionnaire is designed. Free web scraper tools, such as webscraper.io, are available which crawl through different related websites and collect data that can be exported through CVV files. Organisations such as Pew Research Centre[1] also provides various online consumer related data.

Identification of the possible contributors and target audience: Experts help is needed to evaluate the shortlisted parameters and identify their impact on the domain. This acts as a cross-verification test on the selected parameters before distributing the questionnaire to the audience. The experts can also suggest the right audience and method of distribution.

3.2 Phase 2: Domain Semantic Deduction

Once important data are collected, and the significance of the parameters is analysed, domain semantics and generic decision logic are identified with the help of experts. For example, in the case of pre-sensing users' shopping behaviour for a given product, the consumer's information such as age, gender, profession, search history, etc., can be considered, and logic can be formed. See the following examples.

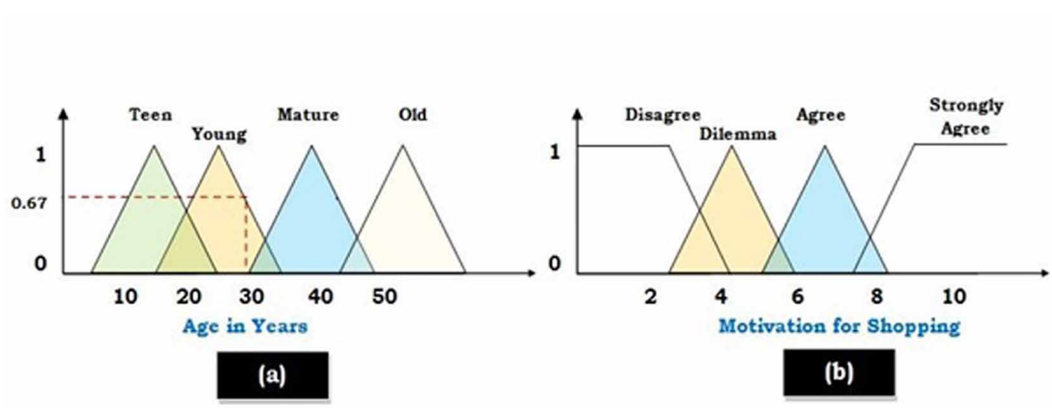
- *If the user is 'Young' and the profession is 'Study' with economic background is 'Rich', the user is interested in a sports car with value 'High'.*
- *For products like Air Conditioning machine, after-sale service is considered 'Late' if it is after 3 days.*
- *etc.*

Deduction of domain semantics can not be effective and applicable unless necessary & sufficient domain parameters are identified, and data about them are collected properly. As stated earlier in this section, parameters such as ease of use, the robustness of the product/service, security aspects of the product, data involved in related services, social benefits, return on investment, and maintenance, etc. are generic parameters and play an important role in modelling consumer behaviour for future decision making. It is to be noted that initially, some domain logic samples need to be identified by experts as seeds/initial population of a genetic algorithm, which is discussed in section 3.4.

3.3 Phase 3: Fuzzification of the Semantics

There are parameters related to the case of modelling consumers' behaviour, which can not be modelled in a crisp manner. Parameters such as economic background, shopping habits, age, the worthiness of product against its cost, speed, and quality of after-sale services, etc., are some example parameters where vagueness is inevitable. These parameters can be modelled with the help of fuzzy logic to efficiently accommodate their vagueness into the computerised system. The selected fuzzy parameter can be defined either on actual values or on the scale of 10 on X-Axis, showing vagueness between 0 and 1 on Y-Axis. In the case of Figure 2(A), the age of a possible consumer is a fuzzy variable and takes linguistic values such as Teen, Young, Mature, and Old. As per fuzzy set behaviour, it is possible that a consumer of age 32 simultaneously belongs to the set of 'Young' consumers as well as a set of 'Mature' consumers with different membership values. In the case of the motivation of shopping, as illustrated in Figure 2(B), a consumer may Disagree or Confuse (Dilemma) for shopping a particular item if the score on a 10 base scale is 5.2. Such a score is derived from an interactive mechanism. This is a kind of expressing and achieving higher generality and better flexibility with the use of fuzzy logic. For each vague parameter, a fuzzy membership function is designed, using which fuzzy rules can be generated. This type of representation is very much like a human way of decision making and reasoning. The membership graphs shown above work to represent the associated vagueness of parameters mathematically to process them by machine. Fuzzy rules help model the relationship between various critical parameters for decision-making and problem-solving. Identification of such a relationship (linear and non-linear) is critical for building a generic model for consumer behaviour. As stated, the use of fuzzy logic provides advantages like working with fuzzy and partial data, offers the virtue of approximate and human-like reasoning, higher generality, and flexibility at a time. For the discovery of knowledge in a domain such as a consumer behaviour, these advantages help a lot.

Figure 2. Fuzzy Membership Functions for (a) Age of Consumer and (b) Motivation of Shopping



3.4 Phase 4: Machine Learning with Genetic Algorithm

For further automation and discovery of knowledge about consumer behaviour, there is a need for an algorithm to automatically generate and select strong fuzzy rules with desirable properties through suitable fitness functions. Consumer data are voluminous, and it is challenging to deduce generalised logic from the data. In this case, these rules are automatically generated and checked with the fitness function. For this purpose, the genetic algorithm is used here. The genetic algorithm has the property to find acceptable solutions from large problem spaces, where typical models are of less use. The typical genetic algorithm has a flow of control, as shown in Table 1.

As illustrated in Table 1, the genetic algorithm begins with an initial population of randomly selected (and a random number of candidates) from the solution space. Desirable characteristics of solutions are modelled in the fitness function. Candidates selected in the initial population are evaluated against the fitness function, and better (stronger) candidates are preserved in the next generation. The next generation is now consisting of stronger candidates toward solutions. On these candidates, operations such as genetic mutation (changing one or more gene values with other possible gene values to represent different characteristics) and crossover (interchanging substrings of genes from two selected individual solutions) are applied. These candidates are further evaluated against fitness function to check their strengths. Stronger candidates are again selected in the new generation, which offers even better solutions if proper operators are applied. Candidates with poor fitness values should be dropped immediately. Gradually better solutions would be evolved by carrying out this procedure repetitively.

Table 1. Flow of Control in Typical Genetic Algorithm

Begin		
	Iteration I¬ 0 Continue ¬ True Initialise Population Pop (I) Evaluate Pop (I) While (Continue) do	
		I¬ I + 1 Pop (I)¬ Selection(Pop (I-1)) Pop (I)¬ Crossover(Pop (I)) Pop (I)¬ Mutation (Pop (I)) Evaluate Pop (I) Update Continue
	End While	
End		

4. EXPERIMENT AND RESULT

To experiment with the above mentioned generic architecture for the evolution of effective rules in a domain, the problem of consumer shopping behaviour is selected. After the collection of data and deduction of domain semantics as discussed in sections 3.1 and 3.2, respectively, fuzzy rules need to be encoded. Some of these encoded rules are randomly selected to form an initial population for further evolution processes, as described in sections 3.3 and 3.4. Here, the basic idea is to define the initial population of rules and evolve stronger rules using operators such as selection, crossover, and mutation. Better rules are evolved through competitions and filtered through proper fitness functions. The following sub-sections discuss the encoding and evolution process.

4.1 Encoding of Fuzzy Rules

Encoding of rules will help represent the logic in a compressed way and provides means of evaluating the strengths of rules as well as carrying out operations such as mutation and crossover. In this case, each rule is considered as an individual. A fuzzy rule can be viewed as linguistic if-then constructions with general form as follows.

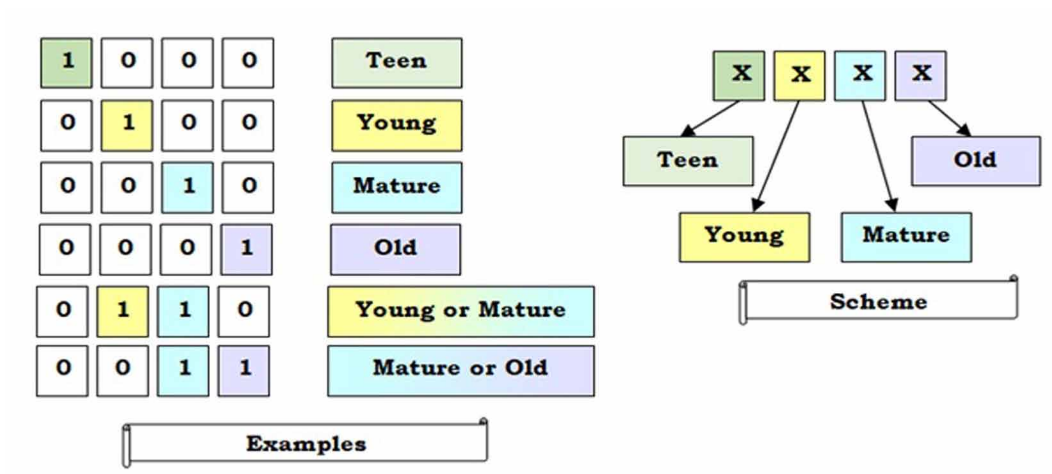
If X then Y; where X (premise) and Y (consequence) are propositions encompassing linguistic variables.

Hence, if there are *I* rules, each with *K* premises in a system, the format of the i^{th} rule is as follows.

If (a_1 is $A_{i,1}$), (a_2 is $A_{i,2}$), (a_3 is $A_{i,3}$), (a_4 is $A_{i,4}$),..... then Bi

It is considered that a fuzzy variable used in a rule has not more than four linguistic terms. For example, Figure 2 showing a fuzzy variable Age in years, has four linguistic values such as 'Teen', 'Young', 'Mature', and 'Old'. For each possibility, there is a value 1 in the corresponding position. Examples are shown in Figure 3.

Figure 3. Genetic Encoding of Fuzzy Variable



Every fuzzy or crisp variable in a rule can be encoded using the above-mentioned scheme. Here, it is considered that each fuzzy variable can take 4 digited values. Consider a rule given as follows.

R1® If the user is 'Teen' or 'Young' and the profession is 'Study' with economic background is 'Rich', then the user is interested in a sports car with value 'High'.

The rule can be encoded as follows.

Encoded R1® 1100 0010 0010

4.2 Evolution of Rules

After efficient encoding, genetic operations such as mutation and crossover would be easy to apply to the encoded rules. Mutation to generate a new rule is considered as a bit flipping operation as the encoding system is binary. Application of mutation is necessary to introduce required diversity in the application; without which, similar individuals keep popping up in the next generations. To automate the genetic mutation operation on a selected rule, the help of a random function is taken. If a random number with seeds 0 and 1 (between 0 and 1) generates a value greater than 0.5 value, then mutation on that digit is performed as a bit flipping operation on the encoded rule; else next bit of the encoded rule is verified.

A crossover operator is also required in the application to combine some really strong individuals in the population, either as it is or partially. A crossover takes substrings from two individuals and recombines them to generate a new individual. Crossover operation helps in combining some of the strong features in two different individuals to evolve into a better individual. Consider two encoded rules as follows.

Encoded R1® 1100 0010 0010

Encoded R2® 0110 0010 1000

A mutation on R1 at position 1 yields the following rule named R3. The encoding of the rule R3 is given as follows.

Encoded R3® 0100 0010 0010

R3 is described as follows.

R3® If the user is ‘Mature’ and profession is ‘Study’ with economic background is ‘Rich’, then the user is interested in a sports car with value ‘High’.

After crossover operation on the 9^{th} and 10^{th} positions on both the rules, the resulting new rule can be given as follows.

Encoded R4® 0110 0010 1010

The new rule, R4 is described as follows.

R4® If user is ‘Young’ or ‘Mature’ and profession is ‘Study’ with economic background is ‘Rich’, then the user is interested in sports car with value ‘High’ or ‘Medium’.

It can be observed that from the given two rules R1 and R2, two new rules have evolved. In a similar manner, many more rules can be evolved automatically. However, all the evolved rules are not worthy of being accommodated into the system without checking its creditability. This makes testing the rules for their fitness is most essential. It has to be noted that the creditability and other qualitative aspects of the genetic algorithm-based system depend on its fitness functions. However, with a more strict fitness function, the system offers good quality of individual but slow evolution.

4.3 Fitness of the Evolved Rules

Since the crossover and mutation operations are carried out in a random manner, not all the rules are strong and meaningful. Sometimes it generates invalid or insignificant results too. To avoid such invalid and insignificant rules and to determine the strength of the newly evolved rule, a fitness function is used. Two major attributes about rules, namely support and confidence, are considered here to evaluate the strength of the rules. Support is an indication of the frequency of existence. It shows how many situations are directly affected by the rule. If support is high, the applicability of the rule is also high. Confidence is the indication of the situation about how often the rule is found to be true. Both support and confidence of a rule need to be high. The fitness function for the case simply uses these two parameters (support and confidence) about a given rule and evaluates the strength of the rule. The higher the strength, the rule is better. For support and confidence following functions are used.

The support of {\displaystyle X}X with respect to {\displaystyle T}T is defined as the proportion of transactions *{\displaystyle t}t* in the dataset which contains the item set {\displaystyle X}X. For example, X = {Teen(age), Rich(Economic), High(Sports Car)) has a support 10%. That is one out of ten cases that happened to be a rich young fellow who wants a sports car.

Confidence (X®Y) is defined as support of (XUY)/ Support of (X). For example, rule {Young, Rich} ® {Sports Car} has full confidence showing all the incidents, where young and rich people want to buy a vehicle, they go for a sports car!

Here, the typical support function is modified to accommodate a number of fuzzy linguistic variables used in the rule. To calculate the support of a given rule, the following function is used.

Support (R1) = (1/ N) {åμA(x^e). μB(y^e). μC(z^e)….}

Where μA, μB, μC, etc., are various fuzzy functions defined for the fuzzy linguistic variables and n is the sample size.

Here is the sample calculation for a rule.

Rule ®*If the user is 'Young' and economic background is 'Rich', then the user is interested in a sports car with value 'High'.*

$A_{young}\ (x_1^{(1)}) = 0.66 \wedge B_{rich}\ (y_1^{(1)}) = 0.72$
$A_{young}\ (x_1^{(2)}) = 0.36 \wedge B_{rich}\ (y_1^{(2)}) = 0.25$
$A_{young}\ (x_1^{(3)}) = 0.0 \wedge B_{rich}\ (y_1^{(3)}) = 0.50$
$A_{young}\ (x_1^{(4)}) = 0.27 \wedge B_{rich}\ (y_1^{(4)}) = 0.36$
Support = {min (0.66, 0.72)+ min(0.36, 0.25) + min(0.0, 0.50) + min(0.27, 36)}/4
= {0.66+0.25+0.0+0.27}/4
= 0.295
Confidence = {min(0.66,0.72). max {1- (min(0.66,0.72)), (min(0.66,0.72))} +
min(0.36,0.25). max {1-(min(0.36,0.25)), (min(0.36,0.25))} +
min(0, 0.50). max {1-(min(0, 0.50)),(min(0, 0.50))} +
min(0.27, 0.36). max {1-(min(0.27, 0.36)), (min(0.27, 0.36))}}
/{0.66+0.36+0+0.27} = 0.8631/1.29 = 0.66
= {0.66. max {1-0.66, 0.66} + 0.25. max {1-0.25, 0.25} + 0. max {1-0, 0} + 0.27. max
{1-0.27, 0.27}}/ {0.66+0.25+0+0.27}
= 0.4356 + 0.1875 + 0+ 0.1971 / 1.18
= 0.8202/1.18
= 0.695085

Table 2 presents sample rules used in the experimental system in symbolic form along with their respective support and confidence values. Later, based on these values, creditability, applicability, and strength of rules are determined. Rules with strong fitness can be shifted either directly to the next generation or chosen for further genetic operation such as mutation and crossover.

Table 2. Sample Rules along with Support and Confidence: Initial Generation

Rule	Age	Gender	Profession	Economic Bkground	History	Price	Services/ Product	Support	Confidence
R_1	Teen	M	Study	Poor	NA	High	Fashion	0.037	0.085
R_2	Mature	M	Job	Medium	High	Medium	Mobile	0.145	0.390
R_3	**Young**	**F**	**Job**	**Rich**	**Good**	**Medium**	**Fashion**	**0.274**	**0.827**
R_4	Old	F	NA	Poor	Low	Low	Décor	0.126	0.411
R_5	**Young**	**M**	**Study**	**Medium**	**High**	**High**	**Automobile**	**0.295**	**0.695**
R_6	Teen	F	Job	Poor	High	Low	NA	0.057	0.232
R_7	**Mature**	**M**	**Job**	**Rich**	**Low**	**Medium**	**FMCG***	**0.462**	**0.546**
R_8	Young	F	Job	Medium	High	Medium	NA	0.223	0.379
R_9	**Old**	**M**	**NA**	**Rich**	**Good**	**High**	**Medicine**	**0.401**	**0.807**
..	…	…	…	…	…	…	…	…	…

*FMCG: Fast Moving Consumer Goods such as Soap, Shampoo, Biscuits, Plastic boxes, etc.

The process of evolution should be continued till required rules are identified with the necessary strength or there is no significant improvement in the strength of available rules. This depends on the domain and volume/nature of data. Some of the rules from Table 2 shown above are better (with bold fonts) in terms of support and confidence. These rules can be directly copied to the next generation. Later, some genetic mutation and crossover can be done. Table 3 shows the next generation with support and confidence values. It can be observed from these tables that support and confidence factors have improved in just a few evolutionary cycles. A similar process is continued for many generations until the desired results are achieved.

Table 3. Sample Rules along with Support and Confidence: Next Generation

Rule	Age	Gender	Profession	Economic Bkground	History	Price	Services/ Product	Support	Confidence
R_1	Teen	M	Study	Rich	Good	High	Fashion	0.211	0.485
R_2	Teen	M	Study	Medium	High	Medium	Mobile	0.159	0.410
R_3	Young	F	Job	Medium	Good	Medium	Fashion	0.274	0.827
R_4	Young	F	Study	Rich	Low	Low	Décor	0.261	0.521
R_5	Young	M	Study	Rich	High	High	Automobile	0.295	0.695
R_6	Mature	F	Job	Poor	V Low	Low	NA	0.157	0.332
R_7	Mature	M	Job	Medium	Low	Medium	FMCG	0.462	0.546
R_8	Old	F	Job	Medium	Low	Medium	NA	0.241	0.401
R_9	Old	M	NA	Rich	Good	High	Medicine	0.401	0.807
..	…	…	…	…	…	…	…	…	…

The average support of the rules shown in Table 2 is 0.224, and the average confidence is 0.485 as returned by the experimental system. The average support and confidence values of the rules enlisted in Table 3 are 0.235 and 0.652. After many such generations, these values can be compared to verify progress in the evolution process. The generation-wise progress is shown in Figure 4 for the first 6 generations.

Figure 4. Improvement in Support as well as Confidence of Rules

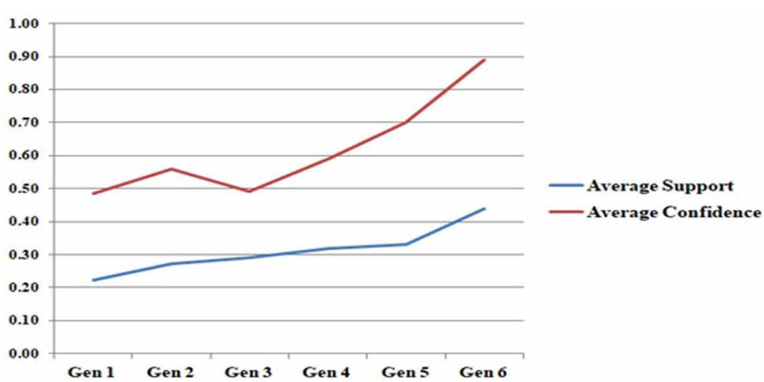


Figure 4 describes the progress in values of support and confidence, and hence it can be concluded that in a few more genetic cycles quality of rules can be improved, and the discovery of the consumer behaviour and pattern can be made automatically from the really large set of data without deducing the generalised logic. Since the use of a genetic algorithm enables effective handling of big of consumer-related data without manual deducing of domain knowledge, the process can be easily automated. Further, the use of fuzzy logic offers human-like interaction, processing, and approximate reasoning. It is to be noted that after some cycles, the average support and average confidence would become steady. This is the time to stop genetic evolution. Another alternative is to set a fixed number of generations or the desired precision of the resulting values of support and confidence.

5. CONCLUSION

Effective modelling of consumer behaviour enables discovery of the underlying pattern of consumer behaviour, their requirements, feedback about products and/or services, and many other types of knowledge in the related domain. Customers have always remained the core component of any business. The proposed method can be used in various domains where there is a need to model consumer behaviour and discover insights about consumers. Knowledge of consumers' behaviour leads to not only the success of the business but also strengthens the business and often leads to other businesses. Besides tracking the possible consumers, one can identify the possible categories of the consumers, future products, improved quality of products and services, and improved brand image of the company. Using fuzzy logic, it is possible to represent domain vagueness, uncertainty, and flexibility in order to get humans like approximate reasoning. With the help of genetic algorithms evolution of better problem-solving logic with strong rules can be achieved, which is considered as discovered knowledge. Such knowledge discovery is otherwise very difficult because of the large amount of data involved in the domain. By combining both technologies, one can get the dual advantages of fuzzy logic as well as evolutionary benefits. The use of fuzzy logic, as stated, offer an ability to handle vagueness in the domain and offers flexibility. Parameters related to the target users (such as age, economic background, history, and shopping behaviour) as well as products can be better represented with the help of fuzzy logic. The evolutionary approach offers evolutionary advantages. The fuzzy rules with good support and confidence can be evolved effectively. Such hybrid systems can model the possible consumer behaviour to help the consumer in facilitating the right product, strengthen the sales and promote the business. It also helps in introducing new and innovative products for a specific group of consumers.

The research work presented in this paper uses a novel and generic architecture to use genetic fuzzy hybridisation that evolves fuzzy rules with good support and confidence values. Further, it is also observed from the results that as data increases, which is inevitable in the business, the confidence and support values of the evolved rules also increases. An increase in such parameters indicates good quality of knowledge discovery.

Besides modelling consumer behaviour, the work can be used for many different domains by changing domain parameters and semantics. The encoding and fitness function might remain the same. As the genetic algorithm is proved effective in searching within a high volume of data, applications involving big data in various other domains such as medical health care, Internet of Things/Everything, Planning, Governance, etc., can also be benefited from this approach. In the future, some more parameters in the fitness function can be added, which can be fuzzy. A natural language-based friendly interface can also be helpful for managers and non-computer professionals. Type-2 fuzzy logic along with suitable membership functions and type reducer components can also be introduced for further benefits of handling extreme vagueness.

FUNDING AGENCY

Publisher has waived the Open Access publishing fee.

REFERENCES

Akerkar, R. A., & Sajja, P. S. (2016). *Intelligent techniques for data science*. Springer International Publishing. doi:10.1007/978-3-319-29206-9

Armstrong, S. (1991). Prediction of consumer behavior by experts and novices. *The Journal of Consumer Research*, *18*(2), 251–256. doi:10.1086/209257

Azzini, A., Marrara, S., & Topalovic, A. (2019). New trends of fuzzy systems: Fintech applications. In Lecture Notes in Computer Science. Springer. doi:10.1007/978-3-030-12544-8_27

Bala, R., & Ratnoo, S. (2016). A genetic algorithm approach for discovering tuned fuzzy classification rules with intra and inter class exceptions. *Journal of Intelligent Systems*, *25*(2), 263–282. doi:10.1515/jisys-2015-0136

Berndtsson, M., Lennerholt, C., Svahn, T., & Larsson, P. (2020). Organizations' attempts to become data driven. *International Journal of Business Intelligence Research*, *11*(1), 1–21. doi:10.4018/IJBIR.2020010101

Beynon, M., Curry, B., & Morgan, P. (2001). Knowledge discovery in marketing: An approach through rough set theory. *European Journal of Marketing*, *35*(7-8), 915–935. doi:10.1108/EUM0000000005734

Casillas, J., & Lopez, F. (2009). A knowledge discovery method based on genetic-fuzzy systems for obtaining consumer behaviour patterns. An empirical application to a Web-based trust model. *International Journal of Management and Decision Making*, *10*(5/6), 10. doi:10.1504/IJMDM.2009.026685

Crawford, G., & Melewar, T. (2003). The importance of impulse purchasing behaviour in the international airport environment. *Journal of Consumer Behaviour*, *3*(1), 85–98. doi:10.1002/cb.124

Dalvi, D. S. (2021). Employer brand equity and its impact on the application intent of the prospective employees. *International Journal of Business Intelligence Research*, *12*(1), 1–25. doi:10.4018/IJBIR.20210101.oa1

Dastidar, U. G., Ambekar, S. S., Hudnurkar, M., & Lidbe, A. D. (2021). Experiential retailing leveraged by data analytics. *International Journal of Business Intelligence Research*, *12*(1), 98–113. doi:10.4018/IJBIR.20210101.oa6

Donovan, R. J., & Rossiter, J. R. (1982). Store atmosphere: An environmental psychology approach. *Journal of Retailing*, *58*(2), 34–57.

Gohary, A., & Hamzelu, B. (2016). Modelling customer knowledge management to make value co-creation. *Business Information Review*, *33*(1), 19–27. doi:10.1177/0266382116631850

Hadhoud, R., & Salameh, W. A. (2020). How business intelligence can help you to better understand your customers. *International Journal of Business Intelligence Research*, *11*(1), 50–58. doi:10.4018/IJBIR.2020010104

Hernández-Julio, Y. F., Ferraz, P. F., Junior, T. Y., Ferraz, G. A., Barbari, M., & Nieto-Bernal, W. (2020). Fuzzy-genetic approaches to knowledge discovery and decision making: Estimation of the cloacal temperature of chicks exposed to different thermal conditions. *Biosystems Engineering*, *199*, 109–120. doi:10.1016/j.biosystemseng.2020.02.005

Howells, K., & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Computer Science*, *120*, 664–670. doi:10.1016/j.procs.2017.11.293

Jeon, S., & Kim, M. (2012). The effect of the servicescape on customers' behavioral intentions in an international airport service environment. *Service Business*, *6*(3), 279–295. doi:10.1007/s11628-012-0136-z

Jian, S., Rashidi, T., & Dixit, V. (2017). An analysis of carsharing vehicle choice and utilization patterns using discrete-continuous extreme value models. *Transportation Research*, *103*(C), 362–376.

Khan, N., & Khan, F. (2013). Fuzzy based decision making for promotional marketing campaigns. *International Journal of Fuzzy Logic Systems*, *3*(1), 64–77. doi:10.5121/ijfls.2013.3107

Kotler, P. (1973). Atmospherics as a marketing tool. *Marketing Retail, 49*(3).

Kotler, P. (1977). From sales obsession to marketing effectiveness. *Harvard Business Review*, *55*, 67–75.

Malpica, A., & Frías, P. (2019). Valuation of an American option for the Spanish secondary reserve market using a machine learning model. *IEEE Transactions on Power Systems*, *34*(1), 544–554. doi:10.1109/TPWRS.2018.2859762

Moyce, C. (2018). *AI, big data, digitisation, blockchain: The fintech that will dominate*. Bobs Guide.

Nicosia, F. (1966). *Consumer decision process*. Prentice- Hall.

Rasiwala, F. S., & Kohli, B. (2021). Artificial intelligence in FinTech: Understanding stakeholders perception on innovation, disruption, and transformation in finance. *International Journal of Business Intelligence Research*, *12*(1), 48–65. doi:10.4018/IJBIR.20210101.oa3

Rau, P., & Samiee, S. (1981). Models of consumer behavior: The state of the art. *Journal of the Academy of Marketing Science*, *9*(3), 300–316. doi:10.1007/BF02721990

Sadikoglu, G. (2017). *Modeling of consumer buying behaviour using Z-number concept*. Intelligent Automation & Soft Computing.

Sajja, P. S. (2020). *Illustrated computational intelligence: Examples and applications*. Springer.

Stubseid, S., & Arandelovic, O. (2018). *Machine learning based prediction of consumer purchasing decisions: The evidence and its significance*. North America.

Tan, C. H., Tan, M. S., Chang, S. W., Yap, K. S., Yap, H. J., & Wong, S. Y. (2018). Genetic algorithm fuzzy logic for medical knowledge-based pattern classification. *Journal of Engineering Science and Technology*, *13*, 242–258.

Tymkiv, A. D. (2016). *Emotions involved in shopping at the airport*. The Aquila Digital Community.

Vatamanescu, E., Nistoreanu, B., & Miltan, A. (2017). Competition and consumer behavior in the context of the digital economy. *Amfiteatru Economic*, *19*(45), 354.

Vijayalakshmi, S., Mahalakshmi, V., & Magesh, S. (2013). Knowledge discovery from consumer behavior in electronic home appliances market in Chennai by using data mining techniques. *African Journal of Business Management*, *7*(34), 3332–3342. doi:10.5897/AJBM2013.7038

ENDNOTE

[1] https://www.pewresearch.org/topics/

Priti Srinivas Sajja has been working at the Post Graduate Department of Computer Science, Sardar Patel University, India since 1994 and presently holds the post of Professor. She specialises in Artificial Intelligence, Machine Learning, and Systems Analysis & Design. She is the author of Illustrated Computational Intelligence (Springer, 2020) and Essence of Systems Analysis and Design (Springer, 2017) published in Singapore and co-author of Intelligent Techniques for Data Science (Springer, 2016); Intelligent Technologies for Web Applications (CRC, 2012) and Knowledge-Based Systems (J&B, 2009) published in Switzerland and USA, and four books published in India. She has served as Principal Investigator of a major research project funded by University Grants Commission, India and as a member of the Indo-Russian Joint Research Project. She has produced 219 publications, out of which eight publications have won best research paper awards. She is a recipient of the Sardar Patel Research Award thrice successively.