Soft Computing for Scalability in Context Aware Location based Services

Priti Jagwani[1]

Dept. of Computer Science
Aryabhatta College, University of Delhi
New Delhi, India

Prof. Saroj Kaushik[2]

Dept of Computer Science
Shiv Nadar University
Noida, Uttar Pradesh, India

Abstract—Ubiquitous computing blended with context awareness gives user the facility of "anywhere anytime" computing. Location based services represents a class of context aware computing. Involvement of location as the primary input in location based services triggered concerns for user's privacy. Most of the privacy work in domain of location based services relies on obfuscation strategy along with K anonymity. The proposed work acknowledges the idea of calculating value of K for K anonymity using context factors in fuzzy format. However, with increasing number of these fuzzy context factors resulting in more fuzzy rules, the system will tend to get slower. In order to address this issue, requirement is to reduce the size of rule base without hampering the performance much. Goal of the proposed work is to attain scalability and high performance for the above said system. Towards this, reduction of number of rules in the rule base, of fuzzy inference system has been done using Fuzzy C Means and Genetic Algorithm. Results of reduced rule base have been compared with the results of exhaustive rule base. It has been identified that number of rules can be reduced up to considerable extent with comparable performances and acceptable level of error.

Keywords—Context aware LBS; Fuzzy C Means; Genetic Algorithm; location privacy; K anonymity; scalability

I. INTRODUCTION

The explosive growth of mobile technology and internet development has facilitated users with many context aware services. Context aware services are adaptive and automatically acclimatize to the environment of the user. There can be various elements of a context like temperature, time, location, density of surroundings, status information of devices present around or behavior of user and many others. Inclusion of context will enable system to provide more personalized and relevant information to a user. Location Based Services (LBS) are considered as representative of context aware services. Use of LBS has brought convenience to the user, but it has raised many concerns like privacy, pricing, data availability, accurate positioning, and accuracy in dealing with spatial information etc. Among all these important issues of usage of LBS, security and privacy of user is among the most prominent ones.

Privacy is considered as a relative term whose perception changes for every individual under different situations. For example, privacy requirement for a user will be different in daytime of working days while it may be different for weekends especially for night time. Other issues attached with location privacy preservation could be who should have access to what location information (in terms of granularity) and under what circumstances. All these constraints can be addressed using a context aware privacy mechanism. Context aware privacy is a rapidly growing idea in the domain of location privacy.

K anonymity has acquired a place as an established mechanism to protect location privacy. For LBSs, location K-anonymity refers to K-anonymous usage of location information. A user is considered location K-anonymous if and only if the location information of that mobile client is indistinguishable from the location information of at least K-1 other mobile clients. K-anonymity is achieved with respect to a specific area which is obtained through spatial cloaking. Using this technique, a user's exact location is blurred into a spatial region in order to preserve the location privacy. The blurred spatial region must satisfy the user's specified privacy requirement which includes K-anonymity and sometimes minimum area of spatial region. Thus K-anonymity guarantees in-distinguishability of a user's location among the location of K users present in a specific area. There is a close coupling between location privacy and location K-anonymity. A larger value of K in location anonymity implies higher guarantees for location privacy.

The general implementation strategy of pull location based services goes like this - user requests a service which is received by middleware. That request, after stripping out his actual identity (because of privacy concerns) is forwarded to location server (who is actually the service provider and is adversary). Identity stripping is the task of middleware, a trusted component. Apart from identity stripping/modification, middleware gives the location of the client in the perturbed/obfuscated form. The results of user's query are determined by location server and returned to middleware. These results corresponds to perturbed location (given to location server by middleware). So, these results are filtered and given to user according to his exact location which is known by middleware. Point of interest (POI) applications which is also called as proximity services or near me services is an example of pull based LBS. It is an important subclass of location-based services concerned with querying a spatial database in order to find information about features of interest that are nearest to an individual's location. Examples of such queries include, "With reference to my current location,

  – "what is the address of the closest Chinese restaurant?"

– "where is the nearest hospital?"

When a user's exact location is blurred into a region, user is made K anonymous within that region. For this an appropriate value of K is required. A suitable value of K for location anonymity can be safely derived from location disclosure by using the equation (1) from the work addressed in [1].

$$K = round\_to\_integer \ \{(1 - Kmax) \ Lp + Kmax)\} \qquad (1)$$

Where Lp represents the value of location disclosure. Kmax is the maximum value of K in a cloaking region for a particular application.

Location disclosure represents acquiescence to disclose location in the form of a number. This number can be derived from various factors. These factors represents context and are in fuzzy form. Safer the context higher is the value of location disclosure and vice versa. High value of location disclosure indicates safety in which a user's location can be disclosed without much restraint and thus indicates a safe context.

Further, the value of K is inversely proportional to location disclosure, which means that higher values of location disclosure will lead to lower K-anonymity. Similarly a higher value of K- anonymity can be obtained from lower location disclosures. So safer the context, higher location disclosure and finally lower will be the K value for anonymity. This value of K is based on current spatial temporal context through location disclosure and is valid, and personalized for all users present in that context.

The above mentioned system was implemented in the work [2]. This system contain fuzzy inference system (FIS) as one of its key components because factors representing context like sensitivity of location, density of location etc. are in fuzzy format. FIS system will have rules through which location disclosure can be determined on the basis of different values of fuzzy factor. If all the values of various fuzzy factors are considered for defining the rules, number of rules will be huge. Keeping all the rules corresponding to all the values and combinations of various factors may blow up the size of rule base of FIS. This will definitely affect the scalability and performance of the system. So, for scalability and optimization of the proposed system, number of rules in the rule base of fuzzy inference system are needed to be reduced. In the proposed work this task is achieved with the help of soft computing techniques namely Fuzzy C Means Clustering (FCM) and Genetic Algorithms (GA). Firstly it is done using FCM and outcome of the reduced rule base is evaluated in terms of Root Mean Square Error (RMSE). Moreover the reduction of rule base is also done by using GA technique to compare and strengthen the results obtained through FCM.

Upcoming sections of the papers are arranged as follows: Section 2 contains related work, Section 3 presents problem definition and challenges identified; Section 4 contains proposed solution in detail. Section 5 focuses on context modelling and validation followed by implementation details in Section 6. Section 7 describes optimization using the techniques of Fuzzy C means (FCM) clustering and Genetic Algorithm (GA) and along with their evaluation. Finally the Section 8 concludes the proposed work.

II. RELATED WORK

Several research studies concentrated on use of K-anonymity in which location of client is made anonymous among K users.

The idea of location disclosure implemented (as discussed in equation 1) is inspired from [1]. That work is based on event driven model (when a user enters a specific area or any other event occurs) and this paper is focusing on pay per request model (POI search applications). For a pull based application when a user requests a service there can be various factors for considerations like timing of requests, sensitivity/safety of the location, usage duration for the service (like POI) and density of the location. These fuzzy factors are taken care by a Fuzzy inference engine (FIS) which operates on the basis of a set of if then else rules. For the research work done in this domain, value of Kmax (refer equation 1) as 7 has been taken for the implementation purpose as given in [3].

Research described in [3] also proposed a mechanism based on locality-sensitive hashing (LSH) to partition user locations into groups each containing at least K users (called spatial cloaks). The mechanism is shown to preserve both locality and K-anonymity.

Authors in the work [4] investigated the use of location semantics together with K-anonymity. They first learned location semantics from location data. Then, the trusted anonymization server performs the anonymization using the location semantic information by cloaking with semantically heterogeneous locations. This paper proposes algorithms for learning location semantics and achieving semantically secure cloaking areas.

Work in [5] has shown that given a cloaked region including user location, finding the nearest POI to the user location cannot be achieved by range search with a fixed region. They tried to explore unstructured shaped for cloaking areas in the form of Voronoi diagram. Cloaking regions based on K order Voronoi diagrams have been generated. In the old researches, K was kept fixed but later on application started asking the value of K from user itself. This value is customizable according to the privacy needs and application.

Problem of privacy preservation is addressed via anonymization in the [6]. A person may still be identified based on his/her profile if the profiles of all k people in the generalized region are not the same. Notion of k-anonymity has been extended by proposing a profile based k-anonymization model that guarantees anonymity even when profiles of mobile users are revealed to untrusted entities. Specifically, the anonymization methods generalize both location and profiles to the extent specified by the user. A novel unified index structure, called the PTPR-tree to enhance the performance during anonymization has been proposed. PTPR-tree is an extension of the TPR-tree.

In the research work [7], a k-anonymity algorithm based on locality-sensitive hashing is proposed to solve the problem of location-privacy preservation in the subspace. In the proposed algorithm, higher efficiency and higher quality of service are achieved by applying a bottom-up grid-search method.

Authors in [8] proposed a clustering algorithm based on the k-anonymity location privacy preserving model, which is used to realize the establishment of anonymous group in the anonymous model. User's location query is replaced by the center of the anonymous group to improve QoS.

Another model to avoid attacks on location privacy from the leaked information in a continuous query with the user's background knowledge is given by [9]. It depends on the technology of one dimensional coding of Geohash geographic information. It also has a preferable performance in time cost of system process.

Use of soft computing techniques to protect location privacy is also researched extensively.

Authors in the work [10] introducing the concept of fuzzy location which may be desirable to reduce computational overhead and/or to preserve location privacy.

Work described in [11] proposes a combined anonymizing algorithm based on K-member Fuzzy Clustering and Firefly Algorithm (KFCFA) to protect the anonymized database against identity disclosure, attribute disclosure, link disclosure, and similarity attacks, and significantly minimize the information loss.

III. PROBLEM DEFINITION AND CHALLENGES

For a pull based service, various fuzzy parameters representing the context information of the current scenario and query have been identified in the literature [1]. The system is considering fuzzy values of these parameters. On the basis of these fuzzy values various rules have been coined based on which K value for k anonymity has been calculated (through location disclosure) using equation 1. Validation of factors of the context is also very important research goal to attain.

Also, for studying the performance of the system a prototype has been implemented which is having a fuzzy inference system (FIS). This FIS contain a rule base. However, depending on the number of factors and the number of possible values which they can take the size of rule base containing all possible rules with all combination of values, can be huge. Initially, rule base in FIS has been populated with exhaustive rules. This rule base of exhaustive rules contain all possible rules. The inference engine applies an exhaustive search through all the rules during each cycle. With a large set of rules the whole system can be slow [12]. Further, in the current system itself, if more number of relevant factors are to be taken into account, for this scenario or for any other request, exhaustive rule base will have a large number of rules. This will cause slower performance of the system, resulting in scalability and performance issues. So the problem is to achieve scalability and optimization of performance for pull based location services which are using K anonymity based on fuzzy factors. Also context validation of the factors identified is a necessary task.

Based on the challenges highlighted above, the formally coined problem statement is:

Given a scenario of pull based services (POI request as a case study); study its context validation, scalability and overall system performance. Keeping in view the performance of fuzzy systems, to devise the mechanism in order to attain scalability avenues and performance.

IV. PROPOSED SOLUTION

In order to handle the extensibility issues of the system discussed above Fuzzy C means clustering has been used. Using FCM, representative rules with low errors have been identified. These representative rules are improvised centers of clusters of rules obtained through various iteration of clustering. So instead of using an exhaustive rule base, the rules selected through clustering will be used. The number of rules in rule base is reduced for optimization purpose and Root Mean Square Error (RMSE) for every reduced set is computed. It is found that size of rule base can be reduced around 50% with acceptable RMSE. This strategy helped to reduce the size of rule base which in turn directly affects the scalability and performance.

Further, the reduced rule base is also computed using Genetic Algorithm–a paradigm in soft computing. Here representative rules are better chromosomes of GA setup of our problem with high fitness values. Error in the reduced rule base obtained through GA is also calculated. It has been discovered that error calculated using both the techniques; GA and FCM agrees and are in consonance with each other. This discovery established the fact that size of rule base for a fuzzy system can successfully be reduced by having small tolerable errors in the system which are acceptable. This greatly helped in achieving high system performance by handling scalability issues of large systems.

Hence, this work provides the contribution of , evolving the reduced set of rules and optimizing the rule base for scalability and system performance, This reduction is done firstly done through FCM technique and the result obtained are verified through GA.

V. CONTEXT MODELING AND VALIDATION

An indispensable part of developing context based applications is to analyze, select and conceptualize the elements of a specific context for the particular application. This activity refers to as context modeling. In the domain of ubiquitous computing, context can be classified as:

Linguistic context: linguistic context refers to words in texts. It represents the pieces of text that are connected with the particular word of interest. It contains all the words which are relevant to a specific word under observation.

Situational context: This includes any information which can characterize the state of entity or location.

Relational context: It refers to the information which is used in characterizing the relation of entity under observation to other entities.

According to the definition of location based services, determining a context to get any location based service fits into the class of situational context which includes information used to describe entity or location.

For the application under consideration (POI request) sensitivity of the location, usage duration of the requested POI / location, time of the day and density of the area are identified as important factors of context.

In the domain of ubiquitous computing, context aware application framework should always be able to answer "when", "where" and "what" related to the service requested [13]. In the example of pull based application scenarios (POI services particularly) , "where" denotes context characteristics of location which is addressed by the factors "sensitivity" and "density" for the proposed system; "when" denotes temporal factors represented by "time of the day" and, "what" denotes characteristics of the service requested, which is satisfied by name/type of POI and its usage duration. All these parameters of the context validation are satisfied by the choice of context adopted in this work. So, according to the definitions given by literature context modeling identified (in the form of factors identified in the above sections) for the proposed problem is validated/satisfied.

VI. EXPERIMENTS AND IMPLEMENTATION DETAILS

This section gives an insight about the technical implementation details of the system. FIS is presented as a key component of the middleware of the system as shown in Fig. 1.

Client request arriving at middleware invokes FIS. Value of location disclosure has been determined by FIS. This value is computed on the basis of rules in the rule base and then the value of K is determined. The relationship between location disclosure and K values is derived on the basis of equation 1. Value of K will be different for different context scenarios. The context is based on sensitivity, density of the place which represents spatial factors and time, usage duration of POI (temporal factors) thus we have achieved context based location privacy. Complete step by step process of the above procedure is given in listing 1.

A. Experiments

For deciding upon the values of location disclosure (output) for the exhaustive rule base a survey has been done and according to the responses values of location disclosure is assigned to various rules. A set of input values and corresponding output (location disclosure) values are shown in Table I followed by some example rules of FIS.
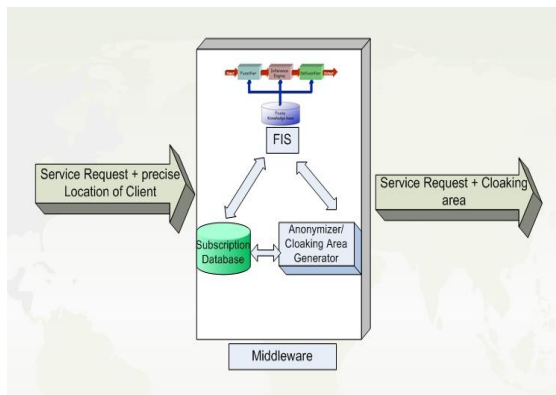

Fig. 1.   Middleware Architecture.

LISTING I.   ALGORITHM FOR WORKING OF THE SYSTEM

1. Mobile user chooses required POI, value of sensitivity and density attribute of location through a user interface.
2. Usage duration of the chosen POI and time of request is taken by system.
3. Above values of contextual parameters and requested POI (service request) is sent to middleware.
5. Contextual parameters will serve as input values for FIS. Fuzzy inference system residing inside middleware is executed.
6. This fuzzy system determines the value of location disclosure according to the rules written in rule base.
7. The rules having different degree of association for the given input are executed.
8. Determined value of location disclosure is then used to calculate K using equation (1)

TABLE I.   LOCATION DISCLOSURE AND VALUE OF K FOR DIFFERENT INPUTS

Sensitivity	Density	Duration	Time	Disclosure level (Lp)	K
Not_sensitive	High	Till_30	Day	High (.83)	2
Less_sensitive	High	Till_250	Day	Normal (.66)	3
Very_sensitive	Moderate	Till_60	Eve	Low (.33)	5
Moderate	Sparse	Less than 3 hrs.	Night	Low(.5)	4
Very very sensitive	Deserted	Less than 6 hrs.	Eve	Very low (.16)	6

B. Example Rules

- If (locationsensitivity is less) and (usageduration is till 150) and (density is deserted) and (requesttime is night) then (disclosurelevel is low)

- If (locationsensitivity is very_less) and (usageduration is till_30) and (density is moderate) and (requesttime is evening) then (disclosurelevel is normal)

- If (locationsensitivity is very_less) and (usageduration is till_30) and (density is high) and (requesttime is day) then (disclosurelevel is high)

- If (locationsensitivity is high) and (usageduration is till_60) and (density is high) and (requesttime is day) then (disclosurelevel is low)

- If (locationsensitivity is very_high) and (usageduration is 250 onwards) and (density is deserted) and (requesttime is night) then (disclosurelevel is very_low).

VII. OPTIMIZATION OF RULE BASE

Now with exhaustive rules in the rule base performance issues may crop up. To tackle this, optimization of rule base is the requirement. So, for the purpose of optimization and to make the system more scalable and fast while running online,

number of rules is reduced. Firstly some sets of input values (fuzzy factors) are given to FIS with exhaustive rules and the location disclosure determined is recorded and designated as the bench mark. These sets are having 50 location inputs (cardinality of the set =50).

Further reduction of rule base is done and the performance is compared with the bench mark results. In order to perform the experiments, numbers of rules are reduced gradually like 650, then 600, and so on (with a set of 700+ exhaustive rules). Same set of data which was processed earlier, with exhaustive rule base is processed with reduced number of rules. RMSE (root mean square error) as a performance metric has been recorded for the lesser number of rules (reduced rule base) and the bench mark set. Now the focal step is the selection of rules to be populated in the reduced rule base. The technique adopted for the selecting the rules for reduced rule base is described in the next subsection.

For the purpose of optimization and reduction of Rule base, two approaches have been applied.1) Fuzzy C means (FCM) technique; 2) Genetic Algorithm based approach.

A. FCM

Fuzzy clustering (also referred to as soft clustering) as the name suggests is a form of clustering in which each data point can belong to more than one cluster. One of the most widely used fuzzy clustering algorithms is the Fuzzy C-means clustering (FCM) Algorithm. It is a data clustering technique consisting of n clusters of data. Every data point in the dataset will belong to the every cluster to a certain degree (membership grade) In FCM data are bound to each cluster by means of a membership function, which represents the fuzzy behaviour of this technique.

Technically, FCM starts with an initial guess for the cluster centers. These cluster centers are intended to mark the mean location of the designated cluster. The initial guess being the random one is most likely to be incorrect. Next, FCM assigns every data point a membership grade for each cluster. By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to the right location within a data set. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade.

FCM is based on the minimization of the following objective function

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\| \qquad (2)$$

Where,

C is the number of data points.

N is the number of clusters.

m is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with $m > 1$. Fuzzy overlap refers to how fuzzy the boundaries between clusters are, that is the number of data points that have significant membership in more than one cluster.

x_i is the ith data point.

c_j is the center of the jth cluster.

μ_{ij} is the degree of membership of x_i in the jth cluster. For a given data point, x_i, the sum of the membership values for all clusters is one.

B. FCM for Current Rule Base

This section presents the FCM technique applied for rule base of the current system. The reason behind choosing FCM as our method of choice for clustering is simple. In the current scenario rules are to be clustered and all rules contain values of linguistic variables which are in the fuzzy form. So it's an appropriate choice to opt for FCM which assigns the data point (rules) to the clusters with a membership grade rather than taking a binary decision for assignment. The FCM clustering techniques is performed in Matlab using Fuzzy Logic Toolbox™.

Further, the goal behind applying FCM is to reduce the number of rules in rule base without sacrificing much on the accuracy of the system. For this the input data which is to be clustered is the exhaustive rule base set. As a result of FCM technique centers of clusters are obtained. These centers of clusters are the rules representing the cluster which indicates that instead of using all the rules of a cluster, center data point (rule) can be used safely as the representative of corresponding cluster. This idea of using only the center of a cluster (a single rule) instead of the whole cluster (some rules) will helps in reducing the size of rule base. Initially the number of clusters to be made was taken as 650. As a result 650 clusters have been formed whose centers are given as output. These centers of clusters are formed after applying FCM on the set of exhaustive rules (700+). These centers represent the rules to be populated in the FIS instead of exhaustive rules. FIS has been populated with these 650 rules (center of 650 clusters) and same set of 50 inputs is again run on this concise, small FIS. Root mean square error (RMSE) has been calculated between the output of these 50 datasets run on exhaustive rule base and on 650 rules. RMSE is calculated using the formula listed in (3).

$$RMSE^k = SQRT \ \{\Sigma(\ Ai - ai^k)^2\ \}/\ 50 \qquad (3)$$

Where Ai is the output (value of location disclosure) of a particular input set i with exhaustive rules and aik is the output of location disclosure for Kth cluster with same set of input i. The complete process of FCM implementation is shown in listing 2.

Further the similar experiment has been performed with different number of rules varying from 650 to 300. Sizes of the clusters have been reduced taking it as 650,600,550 and so on. Corresponding to those clusters, representative (center) rules are extracted. FIS has been populated and system is executed with those reduced set of rules and Root Mean Square Error (RMSE) has been calculated for each reduced set with respect to exhaustive rules which have been taken as a standard/bench mark for the purpose. For more accuracy this experiment has been performed 10 times for every n and average has been taken. Fig. 2 is showing the values of RMSE for different size of rule base.

LISTING II.   FCM IMPLEMENTATION

1. Encode the exhaustive rules in the matrix form. This can easily be done using indexed form of various rules present in the rule base. All these are stored in the initial data matrix D.
2. Taking the above matrix (D) as input set and number of desired clusters as 650(Nc) initially, FCM has been applied.   { [centers,U] = fcm(D,Nc) }
3. As an output of the above step vector of 650 cluster centers (centers) and vector corresponding to their membership degrees (U) has been generated.
4. Rules corresponding to the centers in the above step are extracted.
5. FIS is then populated with these 650 rules which are corresponding to the centers of 650 clusters generated through the technique of FCM.
6. FIS is again executed by using input data (set whose results are recorded as benchmark) used with exhaustive rule base and result values (location disclosure Ai) are calculated.
7. RMSE of the output values with 650 rules has been calculated taking output values of exhaustive rule base as benchmark.
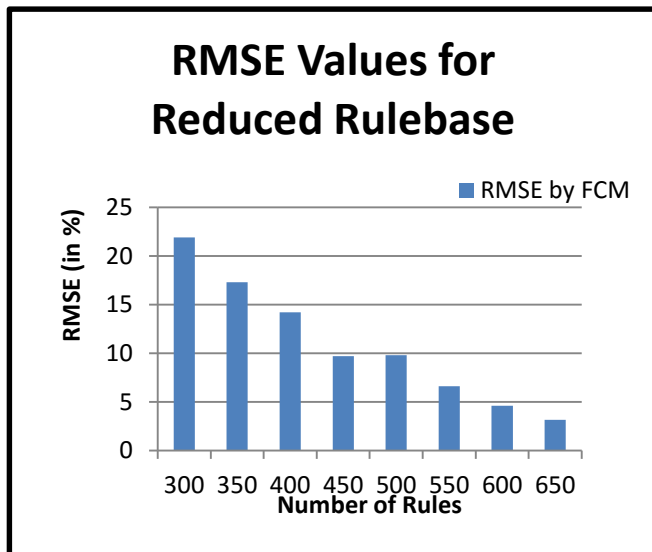

Fig. 2.   RMSE for different Number of Rules Extracted using FCM.

From the above figure it is clear that average RMSE is increasing with decreasing number of rules. The RMSE varies from 2-3% for 650 rules to 22% for 300 rules. Looking at these statistics one can opt for a reduced number of rules with acceptable level of RMSE.

Now, in order to strengthen and verify the claim that system can work with reduced number rules with acceptable errors, GA techniques has also been used to reduce the number of rules. Following section presents the details of GA implementation.

C. Genetic Algorithm (GA) Technique

Genetic algorithm is one of the most widely used nature inspired computing technique. GA evolves nearly optimal solution from given set of potential solutions. Therefore, GA is suitable for searching the solution of underlying problem. The basic concept behind GA is that the 'strong' have a tendency to adapt and survive whereas the 'inferior' tend to die out (survival of fittest).

In GAs, a pool or a population of possible solutions corresponding to a given problem is specified. Various recombination and mutation operations will be performed over these solutions. These operations will result in producing new children and this process is repeated over various generations. Individuals (or candidate solution) are assigned fitness values. This fitness value is derived from its objective function. In consonance with the Darwinian theory of "survival of the fittest", fitter individuals are given a higher chance to mate and generate other "fitter" individuals. In this way, "evolving" better individuals or solutions over generations is continued, till a stopping criterion has been reached.

GA has been used in the current problem to establish the claim that even with reduced number of rules; FIS will provide tolerable errors and satisfactory output with less error. For this we need to select the rules for reduced set. Here, with reference to GA, set of those rules are our fittest candidate solutions. To find these fittest rules, all the rules in an encoded form are taken as initial population. After modifying the chromosomes (which is group of rules) through crossover and mutation operations new population of rules which is fitter than the previous one has been generated. RMSE has been taken as the fitness function to select the population for next iterations. RMSE of the new population as well as RMSE of the initial population of exhaustive rules has been measured with designated input sets (sets referred earlier with cardinality as 50). Finally the set of rules with the lowest RMSE as compared to the exhaustive set will be the final candidate solution.

In order to use GA for the current problem, we need to define encoding scheme for chromosome, fitness function, selection operator, crossover operator and mutation operator. GA formulation for the current problem is inspired from [14]. The details of architecture of our GA based approach are discussed as follows.

D. Chromosome Encoding

To devise a scheme for encoding the individuals of a population is a primary requirement of GA based approach. The individuals are the candidate or potential solution for the underlying problem and the blueprint of any individual is chromosome. The individuals may be encoded as string or real numbers or binary bit string or any other problem specific format. For solving the problem of rule base optimization, GA chromosome is chosen to be a matrix M with number of rows as n. It is the matrix formed by putting n encoded rules together. Each rule is represented by the number string in the indexed format of FIS. For example if there is a rule like.

If (locationsensitivity is very_low) and (usageduration is till_60) and (density is deserted) and (requesttime is day) then (disclosurelevel is low).

Its indexed format is

1 2 4 1, 2;

where initial four values represent antecedents while last value represents the consequent. This form of a rule is easier to be encoded in the form of matrix. So a matrix can have n such encoded rules, as every row in this matrix represents one rule.

Each cell value $P_{i,j}$ indicates the value of jth fuzzy variable for the ith rule. Experiments are performed on varying sizes of rule base. Part of example chromosome that is generated using indexed values is shown in Fig. 3, where R1, R2 (rows) etc. are rules, VA_1, VA_2, VA_3, VA_4 are values of fuzzy antecedents while VC is the value of fuzzy consequent.

E. Initial Population

The initial population at the beginning of algorithm, consists of N chromosomes (each chromosome is matrix having n rows)). The value of N is selected intuitively.

F. Fitness Function

The fitness function computes fitness value of the individual. For the current problem, the root mean squared error (RMSE) is used to model the fitness function. An individual or a chromosome is considered fitter as compared to other, if it has comparatively lower value of RMSE. In order to compute the RMSE for kth candidate solution the equation (3) is used.

Fitness of the Kth candidate solution can be determined by using the following equation.

$$f^k = 1 - RMSE^k \qquad (4)$$

G. Selection Operator

For applying the selection operator, the fitness values of all the chromosomes in the selection pool are computed and they are sorted on the basis of their relative fitness values. Every individual chromosome in population is assigned with a subjective fitness based on their rank within the population defined as follows by equation.

$$sf_i = \frac{(N-r_i)(max-min)}{(N-1)} + min \qquad (5)$$

Where ri represents the rank of ith chromosome in current population. N is the size of population and fitness of best and worst individuals are represented by max and min respectively.

The selection probability for an individual is computed using the equation.

$$p_i = \frac{sf_i}{\Sigma_j sf_j} \qquad (6)$$

	VA₁	VA₂	VA₃	VA₄	VC
R₁	1	2	4	1	1
R₂	1	3	4	3	2
R₃	1	4	4	1	3
R₄	2	1	3	1	1
R₅	3	2	3	2	3

Fig. 3.   Example Chromosome.

Here, the probability of ith individual to be selected is represented by pi and the subjective fitness of ith individual is represented by sfi. The sum of subjective fitness of all the individuals in the current population is represented by Σsfj.

H. Crossover and mutation

Genetic algorithm generally uses two types of procreation operators namely crossover and mutation. Crossover plays primary role in reproduction as it is used to generate offspring whereas mutation is used just for introducing the diversity in the population.

The example of block uniform crossover is shown in Fig. 4. The fitness of the new offspring chromosomes produced using the crossover operation are again evaluated using fitness function given in equation 4. After crossover, the individuals from the old population are killed and replaced.

For mutation, the two-dimensional single-point swapping mutation operator is used. Fig. 5 show the operation of mutation.

I. Termination Criteria

In the current problem, a solution is considered as optimal solution if and only if the overall difference between RMSE of the generated solution and bench mark candidate solution (one with exhaustive rules) is minimized.
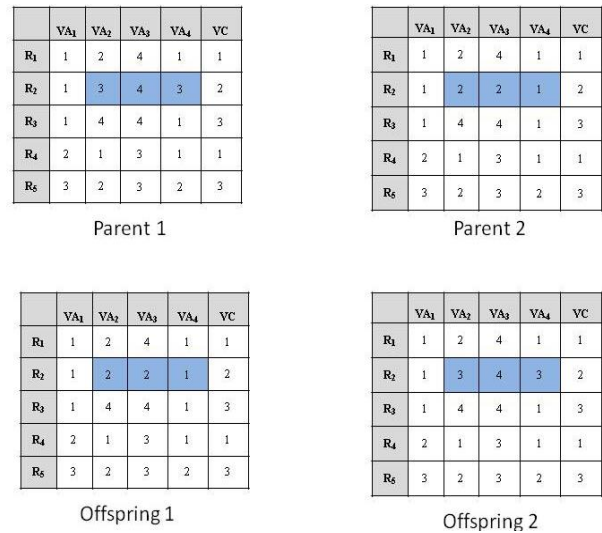

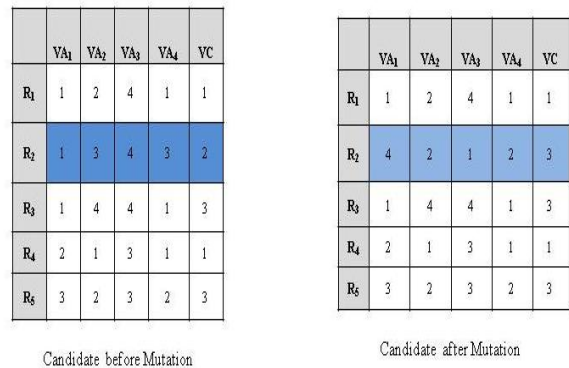Fig. 4.   Crossover Operation on Chromosomes (Encoded Rules).

Fig. 5.   Mutation over Chromosome.

LISTING III.   EVOLVING THE MOST APPROPRIATE RULES FOR FIS

Input:
1. Matrix M of encoded chromosomes
2. Initial population of such matrices (total N). The values of cells of matrix are taken from rules of FIS encoded in 'indexed' form.
3. Number of iterations δ
4. X fraction of population to be replaced in each iteration
5. μ is the mutation rate.

Output: solution containing rules which provide output close to the exhaustive rule base.

1. Pop_size = N
2. Assign current population POP = N matrices of size m x n
    /*loop until the convergence criteria is met */
3. for i = 1 to δ
    /*compute fitness of all individuals */
        3.1 for K =1 to N
        /* use listing 4 to compute fitness */
        3.1.1 f^k = call *findFitness(POP$_k$, M)*
        3.1.2 end / end for of line 3.1 */*
    3.2 Select $(1 - \chi) \times N$ individuals of POP and add into NEW_POP;
    3.3 Select $\chi \times N$ chromosomes of POP using linear rank selection
    3.4 pair selected chromosomes
    3.5 produce offspring using block uniform crossover
    3.6 insert the offspring generated into NEW_POP
    3.7 Select $\mu \times N$ members of NEW_POP and invert a randomly- selected bit in each;
    3.8 POP=NEW_POP
4. End

LISTING IV.   FITNESS FUNCTION

findFitness: It computes the fitness value of a chromosomes

Input:
1. Chromosome POP$_k$
2. Matrix M containing n encoded rules

Output: fitness f^k

1. For i = 1 to m /* where m is number of columns in encoded chromosome */

2. For j = 1 to n
    a. RMSE^k = 0
    b. $RMSE^k = RMSE^k + $ **SQRT** $\{\Sigma(Ai - ai^k)^2 \}/$ **50**

3. End /* end of line 2*/

4. End /* end of line 1 */
5. $f^k = 1 - RMSE^k$
6. Return f^k

The implementation details for the genetic algorithm used is presented in listing 3 and listing 4.

As the result of GA, various rule base with different sizes (as 650,600,550...) have been determined which are fittest among their own population size as compared to others. In other words FIS when applied with these rule base, results in the minimum RMSE as compared with bench mark of exhaustive rule base. Value of RMSE resulted after the process of GA is presented in Fig. 6 along with the size of rules base.

Fig. 6 shows average values of RMSE for different size of rule base. The statistics plotted is showing that with increasing number of rules, average RMSE is evidently decreasing. One can opt for reduced number of rules to optimize the system based on the acceptable level of RMSE.

These results clearly depicts that the limitation of scalability and slow system performance which is introduced with FIS and fuzzy parameters can be handled by reduced number of rules. FIS with suitable number of rules according to the system scalability and error tolerance can be chosen.

The aim of the proposed work was to address the scalability and performance issues of fuzzy context aware pull LBS. Towards this two techniques FCM and GA have been presented and results in terms of RMSE has been calculated.
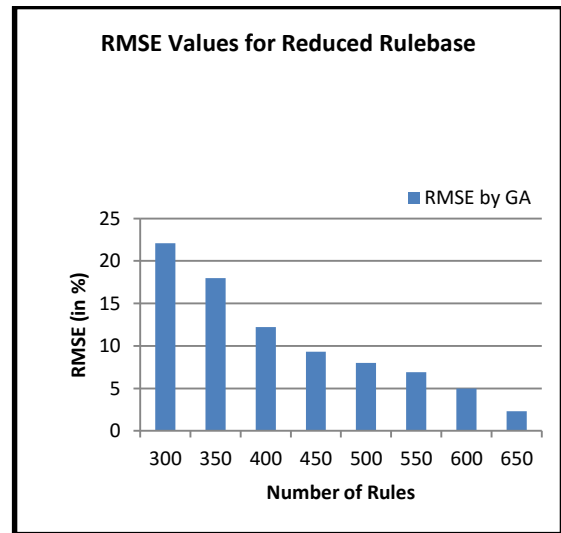

Fig. 6.   RMSE for different Number of Rules Extracted using GA.

TABLE II.   RMSE VALUES FOR GA AND FCM

Number of Rules	RMSE (%)	
	GA	*FCM*
650	2.3	3.15
600	5.0	4.6
550	6.9	6.6
500	8.0	9.8
450	9.3	9.7
400	12.2	14.2
350	18.0	17.3
300	22.1	21.9

From Fig. 2 and Fig. 6, it is clear that RMSE values of the rule base extracted using FCM and GA are in line with each other. These values are shown in Table II. Table II compares the RMSE % for different number of rules obtained through both the techniques.

Value of RMSE obtained by the two methods differs only slightly which strengthen the fact that size of rule base can certainly be decreased with acceptable error to maximize the performance of the system proposed.

VIII. CONCLUSION

With the proliferation of mobile devices, context aware computing including location based services is now at the fingertips of users. This paper extended the idea of fuzzy context based location privacy in ubiquitous computing. Previous researches witnessed that, in order to determine value of location disclosure fuzzy value of context determining factors is taken. Location disclosure is computed on the basis of these fuzzy values which in turn used to calculate K values for K anonymity.

In this work for the, purpose of scalability and optimization, reduction of number of rules in the rule base of fuzzy inference system within a tolerable level of error has been done. For rule base reduction two techniques have been implemented FCM and GA. Reduced size of rule base have been determined by the above two techniques and their results have been compared with the results of exhaustive rule base. It has been identified that number of rules can be reduced up to a considerable extent with comparable performances and acceptable level of error. Reduction for FIS based on any type of rules can be done offline so that it will not affect the overall response time. Reduction produces the FIS which is more portable with similar performances. Moreover reduced rule base establishes the scalability avenues of the proposed concept.

Further, comparing the time taken by exhaustive and reduced rule base can be taken up. Also automatic evolution of rule base for location privacy in pull based services and performing experiments for large number of context determining factors can serve as promising future research directions.

REFERENCES

[1] Poolsappasit, Nayot, and Indrakshi Ray. "Towards Achieving Personalized Privacy for Location-Based Services." Trans. Data Privacy 2, no. 1 (2009): 77-99.

[2] Jagwani, Priti, and Saroj Kaushik. "K anonymity based on fuzzy spatio-temporal context." In 2014 IEEE 15th International Conference on Mobile Data Management, vol. 2, pp. 15-18. IEEE, 2014.

[3] Vu, Khuong, Rong Zheng, and Jie Gao. "Efficient algorithms for k-anonymous location privacy in participatory sensing." In 2012 Proceedings IEEE INFOCOM, pp. 2399-2407. IEEE, 2012.

[4] Lee, Byoungyoung, Jinoh Oh, Hwanjo Yu, and Jong Kim. "Protecting location privacy using location semantics." In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1289-1297. 2011.

[5] Kim, Hyoungshick. "A spatial cloaking framework based on range search for nearest neighbor search." In Data Privacy Management and Autonomous Spontaneous Security, pp. 93-105. Springer, Berlin, Heidelberg, 2009.

[6] Shin, Heechang, Jaideep Vaidya, and Vijayalakshmi Atluri. "A profile anonymization model for location-based services." Journal of Computer Security 19, no. 5 (2011): 795-833.

[7] Wang, Xiaohan, Yonglong Luo, Shiyang Liu, Taochun Wang, and Huihui Han. "Subspace k-anonymity algorithm for location-privacy preservation based on locality-sensitive hashing." Intelligent Data Analysis 23, no. 5 (2019): 1167-1185.

[8] Zheng, Lijuan, Huanhuan Yue, Zhaoxuan Li, Xiao Pan, Mei Wu, and Fan Yang. "K-anonymity location privacy algorithm based on clustering." IEEE Access 6 (2017): 28328-28338.

[9] Xiang, Wei. "An Efficient Location Privacy Preserving Model based on Geohash." In 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), pp. 1-5. IEEE, 2019.

[10] Hashemi, Mahdi, and Mohammad Reza Malek. "Protecting location privacy in mobile geoservices using fuzzy inference systems." Computers, environment and urban systems 36, no. 4 (2012): 311-320.

[11] Langari, Rohulla Kosari, Soheila Sardar, Seyed Abdollah Amin Mousavi, and Reza Radfar. "Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks." Expert Systems with Applications 141 (2020): 112968.

[12] Kalbande, Dhananjay R., and G. T. Thampi. "Multiattribute and Multi-criteria Decision Making Model for technology selection using fuzzy logic." International Journal of Computing Science and Communication Technologies 2, no. 1 (2009): 377-383.

[13] Sandkuhl, Kurt, and Ulrike Borchardt. "How to identify the relevant elements of "context" in Context-aware Information Systems?." In International Conference on Business Informatics Research, pp. 290-305. Springer, Cham, 2014.

[14] Tiwari, Sunita, and Saroj Kaushik. "Evolving recommendations from past travel sequences using soft computing techniques." International Journal of Computational Science and Engineering 14, no. 3 (2017): 242-254.