



# **International Journal of Advanced Research in Computer Science and Software Engineering**

# **Research Paper**

Available online at: www.ijarcsse.com

# Improving Efficiency of Learning FIDs in Fuzzy Relational Databases

Gargee Singh<sup>1</sup>, Priya Arora<sup>2</sup>, A.K. Sharma<sup>3</sup>

Department of Computer Science & Engineering Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India

Abstract—In this paper we propose efficient SEARCH Algorithm procedure based on parallel processing and use of Boolean Algebra to find  $FID_{\alpha}s$  in fuzzy relational databases. Fuzzy inclusion dependencies  $FID_{\alpha}s$ ,  $\alpha \in [0,1]$  express subset-relationships between fuzzy databases and are thus important indicators for redundancies between fuzzy databases. In general, the discovery of  $FID_{\alpha}s$  will be beneficial in any effort to integrate unknown fuzzy databases. The problem of searching  $FID_{\alpha}s$  between two fuzzy relations is NP-hard. By using the concept of parallel processing and use of Boolean algebra there is a significant improvement in the SEARCH algorithm rather than the naïve algorithm.

Keywords—fuzzy inclusion dependency, fuzzy databases

#### I. Introduction

Fuzzy databases are developed to provide users with the ability to store data that can be used in deriving information satisfying their needs. They accommodate a wider range of real-world requirements and provide a friendlier environment for man-machine interaction. The information obtained from them can be used in decision making and problem solving in certain environments which involve uncertainty and imprecise, incomplete, or vague information can be dealt with approximate reasoning. For example, if the temperature of a person suffering from fever is 101.8°F, then one may specify it to be around 102°F.

A functional dependency is a constraint on a set of attributes  $(A_1,A_2,...,A_k,X)$  in a relation R, specifying that for any two tuples  $t_1$  and  $t_2$  from R, the following condition holds:

 $t_1$  ( $A_1$ ,  $A_2$ ,..., Ak) =  $t_2$  ( $A_1$ ,  $A_2$ ,..., Ak)  $\Rightarrow t_1$  (X) =  $t_2$  (X). Fuzzy inclusion dependencies FID $_{\alpha}$ s,  $\alpha \in [0,1]$  express subsetrelationships between fuzzy databases and are thus important indicators for redundancies between fuzzy databases. These FID $_{\alpha}$ s may arise while putting efforts to integrate fuzzy relational databases into a fuzzy relational multidatabase. In general, the discovery of FID $_{\alpha}$ s will be beneficial in any effort to integrate unknown fuzzy databases. Functional dependencies and inclusion dependencies (INDs) are related but have some important differences. In particular, functional dependencies generally are defined only within one relation, whereas the natural purpose of inclusion dependencies is to define relationships across two different relations. Mitchell (1983) also considers inclusion dependencies within one relation. Functional and inclusion dependencies are related in the sense that they both constrain possible valid database states and are thus helpful in database design. However, for the purpose of discovering information about relationships across unknown fuzzy relational databases, the case of fuzzy inclusion dependencies (FIDs) is more useful.

The problem of searching  $FID_{\alpha}s$  between two fuzzy relations is NP-hard. By using the concept of parallel processing and use of Boolean algebra there is a significant improvement in the SEARCH algorithm rather than the naïve algorithm. The efficiency is a concern as a brute force method exploiting evaluating every step amongst the exponential number of steps is being evaluated.

The motivation behind addressing this problem is that the SEARCH algorithm given does the searching of the  $FID_{\alpha}s$  in a Brute-Force manner and hence we present a method that performs parallel computation and reduces the time taken by the SEARCH algorithm.

This paper is organized as follows. Fuzzy Inclusion Dependency (*FID*), related definitions and inference rules on *FIDs* are given in Section II. Algorithm for the discovery of *FIDs* and its complexity analysis of the problem is given in section III. Problem definition and parallel version of the algorithm and its complexity are given in section IV and section V.

# II. DEFINITIONS, CONCEPTS AND BACKGROUND

**Definition 1** Fuzzy Value Equivalent: Let A and B be two fuzzy sets with their membership functions  $\mu_A$  and  $\mu_B$  respectively. A fuzzy value a  $\epsilon$  A is said to be equivalent to another fuzzy value b  $\epsilon$  B iff b  $\epsilon$   $\mu_B$  (x)for some x  $\epsilon$  S, where S is the set of crisp values that are returned by  $\mu_A^{-1}$  (a), where  $\mu_A^{-1}$  is the inverse of the membership function of fuzzy set A.

### A. Notations used for Fuzzy Relational Databases

We have considered the fuzzy relational databases (FRDBs) of Buckles and Petry [1], however, throughout this work we will use notations similar to that in [2]. We will denote set variables by capital letters and variables that denote elements of a set by small letters. By "k-subset of X" we mean a subset of X with cardinality k, while a "k set" is simply a set with

cardinality k. A fuzzy value is an element of data that is stored in a fuzzy relation's extent. Examples include .6/good, .5/old, or .8/high etc. A domain D is a finite set of fuzzy values. A fuzzy attribute is a bag (multiset) of fuzzy values. A fuzzy relational schema is a pair (Rel, U) where Rel is name of the fuzzy relation and U =  $(a_b \ldots , a_n)$  is a finite ordered n-tuple of labels, that is known to be fuzzy attribute names. A fuzzy relation is a 3-tuple R = (Rel, U, E) with Rel and U as above and E  $D_1 X D_2 X \ldots X D_n X D_{n+l}$  the fuzzy relation extent. The sets  $D_1, \ldots, D_n$ , are called the domains of R's fuzzy attributes.  $D_{n+l}$  is the domain of the membership function  $\mu_{Rel} = min(\mu_{al}, \ldots, \mu_{an})$  that determines the tuple membership to the fuzzy relation Rel. A fuzzy tuple in fuzzy relation R is an element of E. An operator t  $[a_1, a_2, \ldots, a_k]$  returns the projection of t on the fuzzy attributes named  $a_1, a_2, \ldots, a_k$ .

#### B. Fuzzy Inclusion Dependencies (FIDs)

Fagin [5] introduced and formally defined the inclusion dependency (*IND*) that can be derived across two relations. Similarly we have introduced and formally defined fuzzy inclusion dependency (*FID*) that can be derived across two fuzzy relations as given below.

**Definition2** (*FID*) Let  $R[a_1, a_2, ..., a_n]$  and  $S[b_1, b_2, ..., b_m]$  be (projections on) two fuzzy relations. Let X be a sequence of k distinct fuzzy attribute names from R, and Y be a sequence of k distinct fuzzy attribute names from R, with  $1 < k < \min(n,m)$ . Then, fuzzy inclusion dependency FID is an assertion of the form  $R[X] \subseteq S[Y]$ , where all the Fuzzy Values under all the attribute names in R[X] names in R[X], however; the vice versa may not hold. are Fuzzy Value Equivalent to some Fuzzy Values under respective attribute names in R[X], however; the vice versa may not hold.

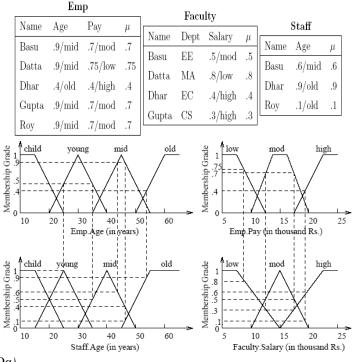
**Definition 3 (valid)** A FID  $p = (R[a_{i1}, \ldots, a_{ik}] \subseteq S[b_{i1}, \ldots, b_{ik}]$  is valid between two relations  $R = (r, (a_i, \ldots, a_n), ER)$  and  $S = (s, (b_1, \ldots, b_m), Es)$  if the sets of fuzzy tuples in ER and Es satisfy the assertion given by p. Otherwise, the FID is called invalid for R and S.

A fuzzy inclusion dependency is merely a statement about two fuzzy relations that may be true or false. A valid FID describes the fact that a fuzzy projection of one fuzzy relation R forms a fuzzy subset of another fuzzy projection(of the same number of fuzzy attributes) of a fuzzy relation S. Note that FIDs are defined over sequences of attributes, not sets, since the order of attributes is important (FIDs are not invariant under permutation of the attributes of only one side) and concept of Fuzzy Value Equivalent (definition1 given at page 2) is used to measure the equality of two fuzzy values or two fuzzy tuples.

**Definition 4** (arity of a *FID*) Let X and Y be sequences of k fuzzy attributes, respectively and  $\rho = R[X] \subseteq S[Y]$  be a FID. Then k is the arity of  $\rho$ , denoted by  $|\rho|$ , and  $\rho$  is called a k-ary FID.

C. Definition Partial Fuzzy Inclusion Dependency (FIDa)

Let R [ $a_1$ ,  $a_2$ ,....,  $a_n$ ] and  $S[b_1,b_2,...,b_m]$  be (projections on) two fuzzy relations. Let X be a sequence of k distinct fuzzy attribute names from R, and Y be a sequence of k distinct fuzzy attribute names from S, with  $1 \le k \le \min(n,m)$ . Then, a partial fuzzy inclusion dependency  $FID\alpha$  is an assertion of the form  $R[X] \subseteq S[Y]$ , such that the fuzzy subset hood  $S(R[X], S[Y]) = |R[X] \cap S[Y]/|R[X]| \ge \alpha$  where  $\alpha$  is specified in the interval [0,1] and most of the Fuzzy Values under all the attribute names in R[X] are Fuzzy Value Equivalent (FVEQ) to some Fuzzy Values under respective attribute names in S[Y], however, the vice versa may not hold.



D. Inference rules for (FIDα)

# Singh et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(4), April - 2014, pp. 468-473

Set of inference rules from the view point of FIDs repeated application of which will generate valid FIDs are given below.

**Axiom 1** (reflexivity)  $R[X] \subseteq R[X]$ , if X is a sequence of

distinct fuzzy attributes from R.

**Axiom 2 (projection and permutation)** If  $R[A_1, ..., A_m] \subseteq S[B_1, ..., B_m]$  is valid (by Def.3), then  $R[A_{il}, ..., A_{ik}] \subseteq S[B_{i1}, ..., B_{ik}]$  is valid for any sequence  $(i_1, ..., i_k)$  of distinct integers from  $\{1, ..., m\}$ . Note that permutation refers to "synchronous" reordering of attributes on both sides, i.e.,

 $R[X,Y] \subseteq S[X,Y] \Rightarrow R[Y,X] \subseteq S[Y,X]$  but

 $\sim R[X,Y] \subseteq S[X,Y] \Rightarrow R[Y,X] \subseteq S[X,Y]$ 

**Axiom 3** (transitivity) If  $R[X] \subseteq S[Y]$  and  $S[Y] \subseteq T[Z]$ 

are both valid (by Def 3) then  $R[X] \subseteq T[Z]$  is valid.

Proof: It is sufficient to show that if a Fuzzy Values x

is Fuzzy Value Equivalent to some Fuzzy Values y and the

Fuzzy Value y is Fuzzy Value Equivalent to a Fuzzy Value

z then x is Fuzzy Value Equivalent to z.

#### III. ALGORITHM FOR DISCOVERY OF FIDS

Fuzzy relational databases consist of fuzzy relations, which in turn consist of fuzzy attributes. In order to discover FIDs in such fuzzy databases, our approach requires that no fuzzy attribute occurs more than once in any one fuzzy data object. The existence of such layers suggests a three-layered strategy to discover relationships between fuzzy databases: compare fuzzy attributes, compare fuzzy relations, and, finally,compare fuzzy databases. It is clear that two fuzzy relations whose fuzzy attributes are not related cannot in turn be related, and likewise a relationship between fuzzy databases requires relationships between their fuzzy relations. There are three necessary stages in the proposed algorithm to solve the abo\e problem:

- 1. SEARCH<sub>n</sub>: finding valid FIDs between a set of given fuzzy relations in a set of fuzzy databases (the general problem),
- 2. SEARCH<sub>2</sub>: finding valid FIDs between a pair of given fuzzy relations,
- 3. VERIFY: Determining whether a given FID is valid. All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. Searching FIDs Between Two Fuzzy Relations

Consider two fuzzy relations R and S. we have to search a

f tuzzy relations R and S. we have to search a

generating set  $G(\sum)$  of valid fuzzy inclusion dependencies  $\sum = \{\rho_1, \rho_2, \ldots, \rho_n\}$  of the form  $R[A_R] \subseteq S[A_s]$  with  $A_R$  a subsequence of fuzzy attributes from R and As a subsequence of fuzzy attributes from S. It can be resolved from Axiom 2 that a k-ary valid FID implies certain i-ary valid FIDs, for i < k. The set of FIDs obtained from an FID  $\rho_0$  through projection only is equivalent to the set  $\{\rho / \rho_0 \models \rho\}$ . An observation about the number of FIDs that can be derived as projections, (i.e. subset of other FIDs) can also be made in terms of the following lemma,

**Lemma 1** A k-ary valid FID implies  $\binom{k}{m}$  m-ary valid FIDs,

For any  $1 \le m \le k$ .

Algorithm 1 SEARCH<sub>2</sub>

This algorithm searches for FIDs between two fuzzy relations each with k attributes.

Step 1: Set m=1. Generate all unary FIDs. Verify and retain valid FIDs in a set  $\sum_{i}$ 

Step 2: Increment m by 1.

Step 3: Generate only those m-ary FIDs whose implied (m-l)-ary FIDs are all members of  $\sum_{m-1}$ .

Step 4: Verify and retain valid m-ary FIDs in a set  $\sum_{m}$ .

Step 5: Repeat from step 2 until m=k.

#### B. Complexity

We will consider the worst-case complexity of the problem as a function of the number of fuzzy attributes in both fuzzy relations. The maximum number of distinct FIDs between two fuzzy relations can be computed as follows. For two relations R and S with  $k_R$  and  $k_s$  fuzzy attributes, respectively, one can form  $k_R.k_s$  unary FIDs. It is possible for all such FIDs to be valid at the same time, namely when all fuzzy attributes in both fuzzy relations have Fuzzy Value Equivalent fuzzy data (which is not likely to occur in practice). The number of k-ary FIDs in general is determined by the number of pairs of k-ary subsets of the fuzzy attributes in each relation. For each such pair, there are k! FIDs, since each permutation of one side of a FID, while keeping the other side unchanged, gives a new FID (i.e., a FID not equal to any previously generated FID). Permutations of both sides do not lead to new FIDs, as this process would generate FIDs equal to previously generated ones. Therefore, the number of k-ary FIDs, denoted by  $F_k$ , is

$$F_k(k_R, ks) = {k \choose n} {k \choose s} .k!$$

# Singh et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(4), April - 2014, pp. 468-473

Assuming without loss of generality that ks < k<sub>R</sub> the total number of FIDs between R and S, denoted by F, is

$$F_{k}(k_{R}, ks) = \sum_{i=1}^{k_{s}} {k \choose i} \cdot {k \choose i} \cdot i! =$$

$$\sum_{i=1}^{k_s} k_R! . \ ks! / (k_{R}-i)! . (ks-i)! . i!$$

A naive brute-force FIDs-finding algorithm for a pair of fuzzy relations could first generate all possible FIDs and then test each of them for validity. It would thus have a complexity in at least  $O(F(k_R, ks))$ , even if FID-testing could be done in constant time. Clearly, since this number grows extremely fast, it is not meaningful to test for or even generate all FID candidates for a pair of relations, even if those relations have a small number (say < 9) of attributes.

### IV. PROBLEM DEFINITION

Given a set of fuzzy relations  $\mathbf{R}^* = \{\mathbf{R}_1, \mathbf{R}_2, \dots \mathbf{R}_m\}$  stored in one or more fuzzy relational database management systems, search for a set of fuzzy inclusion dependencies(i.e. FID) between the two relations  $\{\mathbf{R},\mathbf{S}\}$  in  $\mathbf{R}^*$  efficiently.

- A. The key concepts and notations being used are as follows:
  - 1. Fuzzy Inclusion Dependencies
  - 2. Valid Fuzzy Inclusion Dependencies
  - 3. Fuzzy Value Equivalence
  - 4. Lemma: A k-ary valid  $FID_{\alpha}$  s implies  $\binom{k}{m}$  m-ary valid FID'S for any 1 <= m <= k.
  - 5. R: Here R represents a fuzzy relation
  - 6. S: Here S represents another fuzzy relation
  - 7. k: represents the attributes of the fuzzy relation R and S
- *B. Motivation behind the proposed concept:*
- 1. No of permutations of attributes to be checked for FID $\alpha$  s learning could be **large**. We notice that each FID is *independent* of other hence this can be performed in *parallel*. If there are k attributes each in relation  $\mathbf{R}$  and  $\mathbf{S}$  then there are  $k^2$  possible *unary FID'S*. If there are p processors then we can distribute the p tasks amongst the p processors.
- 2. To compute the *k-ary* FIDs only permutations of **valid** unary FID's are to be used. If there are **m** valid unary FID's then there will be  ${}^{\mathbf{m}}\mathbf{C_k}$  permutations (*Validity checks*) for every k-ary FID, these  ${}^{\mathbf{m}}\mathbf{C_k}$  checks would be performed in *parallel*.
- 3. Given one element of  ${}^{\mathbf{m}}\mathbf{C}_{\mathbf{k}}$  permutation, if any proper subset is invalid through FVEQ check in the prior steps then ,we don't need to perform k-ary FVEQ check. However this property is not exploited in this algorithm.

V.ALGORITHM FOR COMPUTATION OF VALID FID'S

#### A. Algorithm for Computation of Valid Unary FID's

Generating all valid unary FID's: Generates all the unary FID's by running in parallel the routine ComputeFidUnary( $\mathbf{R}$ , $\mathbf{S}$ , $\mathbf{k}$ , $\mathbf{p}$ ). There are  $k^2$  permutation of attributes from relation  $\mathbf{R}$ ,  $\mathbf{S}$  to be verified. These verifications are not interdependent and hence can be performed in parallel using p processors. The verification is done using concept of Fuzzy Value Equivalence (i.e. Validity Test).

## ComputeFidUnary(R,S,k,p): This routine would take inputs:

R, S: Two fuzzy relations

k: Attributes in each relation

p: no of processors

**k\*k** FID's *validity test* would be scheduled on p processes in parallel

- 1. T=k\*k FID tasks with *validity test* to be computed.
- 2. Do while T > 0
  - a. T = T p
  - b. Fork p tasks in parallel
  - c. Collect task which return Valid FIDs
- 3. Maintain a **count** of all valid *unary* FIDS.
- 4. Return the valid FID's collected in a link list

T=k\*k FID tasks with validity test to be computed

## B. Algorithm for Computation of Valid m-ary FID'S

Generating all valid m-ary FID's: For generation of FID at the  $\mathbf{m}^{th}$  level where m>=2, at every  $\mathbf{m}^{th}$  level  $^{\mathbf{m}}\mathbf{C}_{\mathbf{k}}$  FIDs are to be validated. These will be scheduled on p processor in parallel, each processor have  $^{\mathbf{m}}\mathbf{C}_{\mathbf{k}}$  /p tasks. All the valid FID's are collected in a link list.

ComputeFidM-ary (R,S,k,p): This routine would take input as:

R, S: Two fuzzy relations

k: Attributes in each relation

p: no of processors

<sup>m</sup>C<sub>k</sub> FID's validity test would be scheduled on p processes in parallel

 $1.T = {}^{\mathbf{m}}\mathbf{C_k}$  k-ary FID tasks with **Validity Test** to be computed.

2. Do while T > 0

- a. T = T p
- b. Fork p tasks in parallel
- c. Collect task which return Valid FIDs
- d. Go to step 2.
- 3. Maintain a **count** of all valid *m-ary* FIDS.
- 4. Return the valid FID's collected in a link list.

C.Definitions Used:

### 1.Validity Test:

Performing FVEQ test: Using the concept of Fuzzy Value Equivalence as defined before a FID would be declared as either valid or unvalid.

#### 2.The *ValidFid()* linkedlist:

It is essentially a array of link list of size k, where i<sup>th</sup> position in the array defines the i<sup>th</sup> –ary valid FID list.

### D. Algorithm: $SEARCH_2(R, S, k, p)$

The algorithm takes as input two relations **R** and **S** with **k** attributes each and **p** processors for computation.

- 1. Set a suitable value to  $\alpha$  from the interval [0, 1].
- 2. L = NULL
- 3. L = L U ComputeFidUnary(R,S,k,p)
- 4. Set m=2
- 5. L = L U ComputeFidM-ary(R,S,k,p)
- 6. Increment m
- 7. If m<k goto step 5

### E. Example for the Proposed SEARCH2 algorithm

The following is a pseudo example of Fuzzy Relation:

Emp					Foor	14				
Name	Age	Pay	μ	N	Faculty				Staff	
Basu	.9/mid	.7/mod	.7	Name	-		μ	Name	Age	μ
Datta	,	.75/low	.75	Basu	EE	.5/mod	.5	Basu	.6/mid	.6
Dhar	.4/old	.4/high	.4	Datta Dhar	MA EC	.8/low	.8	Dhar	.9/old	.9
Gupta	$.9/\mathrm{mid}$	$.7/\mathrm{mod}$	.7	Gupta	CS	.4/high .3/high	.4	Roy	$.1/\mathrm{old}$	.1
Roy	.9/mid	.7/mod	.7	Gupta	CO	.5/ mgn	.0			

- 1. Inputs for the algorithm are relations Faculty and Emp
- 2.  $\alpha = 0.75$
- 3. L=null
- 4. **ComputeFidUnary(Faculty,Emp,3,p)** :T=3\*3 FID's tasks with **validity test** to be computed using **p** processors hence task of computing  $\rho_{1.9}$  is distributed amongst **p** processors i.e  $\rho_1$ ,  $\rho_2$ ,  $\rho_3$   $\rho_4$ ,  $\rho_5$ ,  $\rho_6$ ,  $\rho_7$ ,  $\rho_8$ ,  $\rho_9$  are computed parallely(via FVEQ) and only  $\rho_1$  and  $\rho_9$  are found to be valid using FVEQ. These valid unary FID's are collected in a link list(L={  $\rho_1$ ,  $\rho_9$  }) and a count is maintained (count==2).
- 5. m==2
- 6. **ComputeFidM-ary(Faculty,Emp,3,p):**  $T=^{m}C_{k}$  tasks ( $^{2}C_{2}$ ) binary tasks need to be computed in parallel for the *validity test* hence the task of computing one value( as  $^{2}C_{2}=1$ ) of  $\rho$  (using FVEQ)is distributed amongst p processors. The valid binary FID's are collected in a link list( $L=\{\rho_{19}\}$ ) and a count is maintained(count==1)
- 7. m=m+1:m=2+1=3
- 8. m=3==k hence stop

### F. Key Contribution:

The following areas as mentioned below had some scope of improvement in the SEARCH<sub>2</sub> algorithm given by (Sharma et al., 2004) which have been explored in the above mentioned algorithm in the following manner.

1. The computation of validity of independent events and hence could be performed parallel.

# Singh et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(4), April - 2014, pp. 468-473

- 2. The FID'S which become invalid at the unary stage need not be considered at the next stage for **ValidityTest** and hence less no of parallel processes need to be forked and made run in parallel in case some become invalid at the unary level. So for e.g. if m are the no of valid unary then  ${}^{m}C_{2}$  tasks are distributed over p processors for binary FID evaluation and then undergo the **Validity Test** however most of them can fail depending on the the FVEQ condition of the unary FID's.
- 3. The processes spawned at every stage is  ${}^{m}C_{k}$  where **m** is the count of the unary FID'S and **k** is the k-ary FID being computed. These  ${}^{m}C_{k}$  tasks are distributed over p processors that run in parallel computing the validity.
- 4. The *Boolean computation* that was being performed before the **FVEQ TEST** (Sharma et. al 2004) need not be performed over here as the generating or spawned processes generated would always being *generated from a valid set of FID'S*. The only thing that determines or need to be taken care of is the validity using the FVEQ concept. But the Lemma mentioned above is being used only for the unary FID, it can be improved for k=2.....and so on.

### G. Complexity Analysis:

- 1. The parallel *version* of  $SEARCH_2$  performs the computation in  $O(k^2/p)$  time where **k** is the number of attributes and **p** is the number of processors.
- 2. If there are **m** valid unary FID's then m-ary search takes order **O(m!/p)**.

This can be explained as follows:

Computation of the k-ary FID's using p processors takes time of the order:  $O(^mC_k/p)$  for 2 <= k <= m. Hence the total time taken is of the order:  $O(^mC_2/p) + O(^mC_3/p) \dots O(^mC_k/p)$  which is equal to O(m!/p) V. CONCLUSION

The concept of Fuzzy Inclusion Dependency is very *interesting* as it helps to establish a relationship across two fuzzy relations but generating and testing validity of FID's is compute intensive. We have used parallel algorithm to speed up the validation of FID's.

#### VI. SCOPE FOR FUTURE WORK

- 1. The lemma needs to be exploited in a better fashion specially the invalid FID'S to improve the number of checks to be performed. Since at each level the number of valid FID's is going to decrease we can reduce the no of <sup>m</sup>C<sub>k</sub> checks by exploiting the computation.
- 2. (Sharma et.al 2004) have proposed a *clique based approach for the N-P hard* problem of discovery of fuzzy inclusion dependencies ,we would investigate use of *SATISFYABILITY FOR THIS PROBLEM*
- 3. Also to establish Fuzzy Value Equivalence between two attributes a mathematical model can be investigated.

#### REFERENCES

- [1] A.K. Sharma et al.: An Algorithm for Discovery of Fuzzy Inclusion Dependencies
- [2] Buckles B-P, Petry E-F (1982) A fuzzy representation of data for relational databases. Fuzzy Sets and Systems,7 (3), 213-226
- [3] Casanova M-A, Fagin R, and Papadimitriou C-H (1982) Inclusion Dependencies and their Interaction with Functional Dependencies. In Proceedings of ACM Conference on Principles of Database Systems (PODS), pages 171-176.
- [4] **W. W.** Cohen. Integration of Heterogeneous Databases Without Common Domains Using Queries Based on TextuaY Similarity. SIGMOD Record, 27(2):201213,1998.
- [5] J. Cho, N. Shivakumar, and H. Garcia-Molina. Finding Replicated Web Collections. SIGMOD Record (ACM Special Interest Group on Management of Data), 29(2):355366,2000.
- [6] Fagin R. (1981) A Normal Form for Relational Databases that is Based on Domains and Keys. ACM Transactions on Database Systems (TODS), 6(3):387-415.