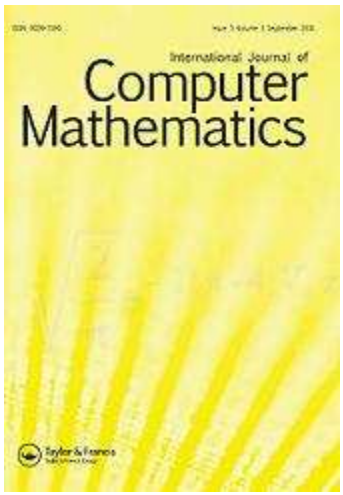
Interpretability of Fuzzy Association Rules as means of Discovering Threaths to Privacy (CMMSE 2010)

Luigi Troiano, Luis J Rodriguez-Muñiz, José Ranilla, Irene Díaz



Interpretability of Fuzzy Association Rules as means of Discovering Threaths to Privacy (CMMSE 2010)

Journal:	*International Journal of Computer Mathematics*
Manuscript ID:	GCOM-2010-0795-B.R1
Manuscript Type:	Review
Date Submitted by the Author:	15-Apr-2011
Complete List of Authors:	Troiano, Luigi; University of Sannio, Department of Engineering Rodriguez-Muñiz, Luis; University of oviedo, Statistics and Operation Research Ranilla, José; Universidad de Oviedo Díaz, Irene; University of Oviedo, Computer Science
Keywords:	disclosure control, fuzzy rules, data privacy, artificial intelligence, anonimity
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
cmmse.zip	

Proceedings of the 10th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2010 27–30 June 2010.

Interpretability of Fuzzy Association Rules as means of Discovering Threaths to Privacy (CMMSE 2010)*

Luigi Troiano[1], Luis J. Rodríguez-Muñiz[2], José Ranilla[3] and Irene Díaz[3]

[1] *Department of Engineering, University of Sannio*

[2] *Department of Statistics and O.R., University of Oviedo*

[3] *Department of Computer Science, University of Oviedo*

emails: troiano@unisannio.it, luisj@uniovi.es, ranilla@uniovi.es, sirene@uniovi.es

Abstract

This paper is focused on studying how data privacy could be preserved with fuzzy rule bases as interpretable as possible. These fuzzy rule bases are obtained from a data mining strategy based on building a decision tree. The antecedents of each rule produced by these systems contain information about the released variables (quasiidentifier) whereas the consequent contains information only about the protected variable. Experimental results show that fuzzy rules are generally simpler and easier to interpret than other approaches but the risk of disclosing does not increase.

Key words: disclosure control, fuzzy rules, data privacy

1 Introduction

Statistical disclosure control (SDC) aims at releasing statistical records while protecting confidentiality of information at the same time. Among the different threats, it is possible that some sensitive information can be disclosed by other information made available. In this case, the risk yields on the hidden information that can be inferred from public information as premise. Discovering a link between hidden and public information can help SDC to prevent such a risk.

*Authors acknowledge financial support by Grants MTM2008-01519 and TIN2010-14971 from Ministry of Science and Innovation and Grant TIN2007-61273 and from Ministry of Education and Science, Government of Spain

It becomes very important to identify paths that, if owned by an intruder, would be able to disclose sensitive information. A possible way to make this identification is to produce a rule base as a mean to prevent disclosure of sensitive information and to search for inferential paths from released attributes to sensitive information [9]. Such a path can be due to background knowledge. Mining rules, able to reconstruct the hidden linkage from given patterns of the other attributes, can put into evidence that if some knowledge is discovered by intruders, this can be used to break privacy protections.

Looking at data, such associations could not come into prominence, as individual data points hardly can be consistently linked to other points. If we look at data within equivalence classes, similar point can be more easily associated to others. Similarity is inherently a fuzzy concept. Therefore, one can think that fuzzy association rules could be able to provide inferential paths that differently would not be discovered. Previous work in this sense is developed in [8].

However, the existence of these paths is only in potential. The possibility of linking some information to other does not entail the disclosure of sensitive information. Often paths are somehow writhed, more due to statistical linkage than due to shared background knowledge. Association rules should be known in order to portray a real threat. A measure of rule interpretability can provide a proxy for such a risk. The more interpretable a set of rules is, the more likely this set is part of some background knowledge, and the higher the risk of disclosing sensitive information is. Thus, pursuing the goal of identifying potential paths able to break privacy protections, we are interested in discovering highly interpretable association rules.

This paper is aimed at outlining how threats to privacy can be identified by interpretability of fuzzy association rules. The goal is twofold: to check if there exist a set of rules able to break privacy protection, to assess model interpretability and to identify attributes that are more able to reveal sensitive information by inference. To do that and considering that decision trees (DT) can provide an effective mean for mining association rules and Fuzzy logic can help to deal with lower granularity of data and to better express semantics of rules by means of linguistic terms, we will assume models based on decision trees in which rules cope with fuzzy information in the premises and intervals in the consequences. The algorithms used in this paper were *FArni-rules* [28], *FId3* [35] and *FPrism* [33]. The proposed method has been tested on CENSUS data set, as provided by the CASC project[1]. The experiments were focussed on searching those rules able to link combination of released data (assessed by fuzzy partitions) to protected one.

As in many cases of practical interest it is not important that precise values of sensitive attributes are disclosed but rather the equivalence class they belong to, association rules can be regarded as a mapping between publicly available information and unreleased data. Thus, the discovery of threats to disclose of sensitive information can be regarded as a classification problem, where an intruder could be able to associate non-sensitive data to a class of sensitive information. For this purpose we will assume 1-consequent rules, whose premises are considered at some level of similarity

[1]http://factfinder.census.gov/servlet/DatasetMainPageServlet

and possible consequences are quantitative and partitioned in ranges of values.

The reminder is organized as follows: Section 2 is devoted to some preliminaries regarding information disclosure; Section 3 presents data mining in the context of SDC problem as well as some issues about interpretability of fuzzy systems; Section 4 describes the approach followed in the paper; Section 5 shows experimental results; Section 6 draws conclusions and future directions worth being investigated.

2 Information Disclosure

Information disclosure has place when an entity (i.e. a person or an organization) is able to learn something regarding another entity by released microdata sets. For example, illness regarding patients could be released via medical databases, or competitors' financial figures by business databases.

Microdata attributes of interest for statistical disclosure control can refer to respondent identity (key attributes), or to relevant information (sensitive attributes). In order to preserve the respondent's privacy, the direct linkage between key and sensitive attributes is hidden by SDC. This process is known as data anonymization. However, an intruder can still attack data anonymization by reconstructing the original link with respect to some records.

In particular, there are two types of disclosure associated to microdata [34]: (i) identity disclosure when the entity is (re-)associated to some sensitive data in an anonymized database; (2) prediction disclosure when some sensitive data is inferred by the other attributes for some known entity. The first is also known as *re-identification*, the second as *attribute disclosure*. In this paper we will focus on the second.

Different metrics for measuring the level of privacy guaranteed by SDC have been proposed over the time. Among them, k-anonymity [31], l-diversity [17], p-sensitiveness [32] and t-closeness [16]. Each of these metrics is able to drive data anonymization with respect to same aspect, but all of them share the common idea that having more records within a group associable to an entity enforce privacy protection.

However privacy should be related to the extent some information can be considered sensitive. For instance, disclosing that incomes are within a given range, can be considered as much as sensitive than more precise information. This case is known in literature as similarity attack.

Therefore diversification, obtaining by altering the initial information, does not necessarily lead to a stronger privacy protection. Even masking or removing a sensitive attribute could be not enough to avoid attribute disclosure.

The aim of this paper is to show evidence that, even if there is no correlation between data, it is still possible to find a link, although approximated, between public and sensitive variables. The simpler this link is, the most likely it can be discovered or known by intruder, representing thus a threat to no-disclosure of sensitive information.

3 Data Mining and Statistical Disclosure Control

The relationship between data mining (DM) and statistical disclosure control (SDC) has been firstly outlined in [9]. The problem in attribute disclosure is basically to find an inferential path from released attributes to sensitive information. Such a path can be due to background knowledge.

DM [10] searches for the relationships that exist in large databases, but hidden due to the large amount of data. DM models the behavior of a given variable in terms of the others, finding non-trivial relationships among the attributes involved [14]. These relationships may provide valuable knowledge regarding the individuals the data are related to. As rule mining is aimed at reconstructing the hidden linkage of given patterns between attributes, it is able to put into evidence that if some knowledge is discovered by intruders, this can be used to break privacy protections. In this sense, DM may infer relationships that can lead to sensitive information disclosure.

The problem of mining association rules have been widely investigated in literature, and several search algorithms have been proposed. Among them, the most prominent is Apriori [1]. This algorithm and its variants perform an exhaustive search of rules with high support and confidence. Sometimes this approach is not feasible with large databases due to its computational cost and, more important, it does not take into account the distortion (and similarity) of data.

There exist a large number of machine learning approaches to overcome this limitation. Some of them, such as Neural Networks or Support Vector Machines (SVM) [13] are very effective and computationally efficient, but the models they provide are generally black boxes and thus not informative enough for the purpose of mining interpretable rules, as outlined in this paper. With regard to rule learning algorithms, there are also several approaches. For example CN2 [6], Swap-1 [11] and RIPPER [7] use separate-and-conquer to learn multiple rules, which increases rule dependence and decreases comprehensibility. CN2 conducts a general-to specific search through a space of rules by adding conditions, Swap-1 builds its rules greedily by removing as well as adding literals during the rule induction phase and RIPPER prunes rules incrementally but literals may only be added to the end of a rule, not deleted or swapped out. On the other hand, other methods as those based on decision tress, such as ID3-based systems [21, 24], OC1 [19] or CART [3] algorithms, are able to provide more informative models. Rough Set Theory [22] also gives methods of drawing conclusions from data, without referring to prior and posterior probabilities intrinsically associated with Bayesian reasoning.

If in data mining it is enough to infer models from training data sets in order to overcome data distortion, in SDC we are interested in models able to reveal and explain relationships in presence of data distortion, as generally introduced by the anonymization process. Indeed some information, although hidden or even removed, could be still linked to identities at some extent considering a lower level of data granularity, at which different point-wise information are assimilated. This problem has been raised in [15].

Fuzzy logic (FL) [36] can help to deal with lower granularity of data and to better

express semantics of rules by means of linguistic terms [18]. This helps to obtain interpretability of models. This is the case of fuzzy decision trees [23].

After years spent on how to build models as much accurate as possible, research in fuzzy logic is focusing on how to obtain accurate but also interpretable models [2], although finding a trade-off between the accuracy and interpretability is known to be difficult [4]. Interpretability of fuzzy models can be regarded by different perspectives. However, it depends on understandability of resulting knowledge base and inference relationships. Roughly, this depends on structural complexity of rule base and fuzzy partition of data. Several metrics aimed at measuring interpretability have been proposed [12] , among them:

- Number of Rules (NOR)
- Total Rule Length (TRL), the overall number of premises entailed by rules
- Average rule length (ARL), the total rule length divided by the number of rules
- Nauck Index, $I_{Nauck} = Comp \times Part \times Cov$ ([20]), where
  - $Comp$ is the complexity of a classifier measured as the number of classes divided by the total number of premises
  - $Part$ is computed as the inverse of the number of labels minus 1
  - Cov is the average normalized coverage degree of the fuzzy partition; it is equal to one in strong fuzzy partitions [30].

In particular, the Nauck Index would score 1 if the model is made of one rule per class, which makes use of one variable and strong partition. The Nauck Index can give us a criteria to compare two fuzzy rule systems against the same problem. It is less meaningful in comparing model interpretability of systems applied to different problems.

4 Our approach

In this paper we are interested to verify if sensitive (hidden) information can be inferred from available data at some extent. The goal is twofold: (i) to check if there exist a set of rules able to break privacy protection and to assess model interpretability; (ii) to identify attributes that are more able to reveal sensitive information by inference.

As in many cases of practical interest it is not important that precise values of sensitive attributes are disclosed but rather the equivalence class they belong to, association rules can be regarded as a mapping between publicly available information and unreleased data. Thus, the discovery of threats to disclose of sensitive information can be regarded as a classification problem, where an intruder could be able to associate non-sensitive data to a class of sensitive information. For this purpose we will assume 1-consequent rules, whose premises are considered at some level of similarity and possible consequences are quantitative and partitioned in ranges of values.

As argued above, decision trees (DT) can provide an effective mean for mining association rules. We will assume models based on decision trees in which rules cope with fuzzy information in the premises and intervals in the consequences. There exist several algorithms able to build association rules by means of fuzzy decision trees. In this paper we will consider the following:

- *FArni-rules* is a fuzzy extension of *Arni-rules*, which is a crisp classifier based on *C*4.5 [25], but despite of using *Information gain* as in *C*4.5, *FArni-rules* uses a measure called *Imputity Level* (*IL*) for determining the quality of the rules induced from examples [28]. *IL* [26] explicitly takes into account not only the probability of success p, but also the difficulty of attaining that amount of examples of class C. Later, once the fuzzy decision tree is induced, *FArni-rules* returns compact fuzzy rule sets after applying a pruning process inherited from *Arni* and *Fan* [27]. *FArni-rules* tries to generate a leaf node. So, it can obtain a tree consisting of just one leaf. *FArni* is presented in detail in [28].

- *FId3* [35, 29] is a variant of ID3, which measures the amount of information the training base is able to transmit, thus enabling to select the most informative attributes. It uses classification ambiguity as a tool to select the most relevant test when constructing the fuzzy tree and it always generates a root node, so it has at least a minimum of three rules. It also copes with typical problems of induction such as binaritation of attribute values, noisy domains, dependences among the input attributes, and incrementality.

- *FPrism* [33] is a fuzzy inductive learning algorithm based on the PRISM learning strategy [5]. It handles vagueness and skips irrelevant tests occurring in each rule by maximizing fuzzy information gain. Its focus is on finding relevant attribute-value pairs, rather than only attributes. During induction, the actual amount of information contributed by each attribute-value pair (selector) is evaluated for a specific classification, and the one with the maximum fuzzy information gain is then selected and added to the induced rule.

In order to compare the fuzzy approach to the conventional one, we will consider the *Arni-rules* system. This algorithm is based on C4.5, but instead of using *Information gain*, it makes use of a measure called *Imputity Level* (*IL*) for determining the quality of the rules induced by examples [28]. *IL* [26] explicitly takes into account the probability of success p and the difficulty of attaining that amount of examples within a class ([27]).

The antecedents of each rule produced by these systems contain information about the released variables (quasi-identifiers among them) whereas the consequent contains information regarding one of the protected variables.

5 An Illustrative Example

This section outlines an example in order to better illustrate the idea. The dataset chosen for our experiments is CENSUS, as provided by the CASC project. This dataset

entails data provided by the U.S. Census Bureau[2] regarding business and financial figures of 1080 companies. These figures are namely Final weight (AFNLWGT), Adjusted gross income (AGI), Employer contribution for health insurance (EMCONTRB), Business or Farm net earnings in 19. (ERNVAL), Federal income tax liability (FEDTAX), Social security retirement payroll deduction (FICA), Amount of interest income (INTVAL), Total person earnings (PEARNVAL), Total other persons income (POTHVAL), Total person income (PTOTVAL), State income tax liability (STATETAX), Taxable income amount (TAXINC), and Total wage and salary (WSALVAL). We selected AGI as the variable to protect. The other variables are assumed to be released and thus as possible rule premises.

In each experiment, we have followed the same fuzzyfication scheme for all variables, depending on the case, the variables are covered by 3 and 5 triangular fuzzy set uniform partition (according to [35]),respectively labelled as *low*, *medium*, *high* and *very low*, *low*, *medium*, *high* and *very high*. Besides, the values of the protected variable have been clustered in 3, 5 and 7 classes (categories). Therefore, what we want to check is not only if the fuzzy systems perform better than the crisp one, but also which level of granularity is more appropriate for this task.

We aim at searching those rules able to link combination of released data (assessed by fuzzy partitions) to AGI categories. This can be regarded as a multi-category problem from a classification point of view. We have a total of 6 problems, each for a combination of fuzzy partition and number of categories. Each problem will be solved using the systems described above. As we are interested in assessing the robustness of system conclusions, we adopted a cross-validation process with 5 folders and 10 repetitions. In other terms, 20% of data have been randomly selected in order to assess the quality of rules obtained by the remaining 80% of data. This procedure has been repeated 10 times.

Accuracy

Table 1 reports the highest Accuracy (in percentage) that systems reached during the cross-validation process, with respect to the different AGI categorization and fuzzy partition coverage of premises.

#(AGI categories)	3-category		5-category		7-category	
#(premises partition)	3	5	3	5	3	5
Farni-rules	60,65%	93,52%	37,04%	70,37%	31,94%	39,35%
FId3	75,46%	93,06%	64,81%	72,22%	48,61%	56,48%
Fprism	50,93%	64,35%	8,80%	23,61%	3,70%	15,28%

Table 1: Accuracy of the different fuzzy systems

From results we can note how accuracy of rules, thus their meaning, stands as far as the number of categories does not become too large. For instance, with 7 AGI

[2]http://factfinder.census.gov/servlet/DatasetMainPageServlet

#(classes)	3	5	7
Arni-rules	99,31	86,57	81,94

Table 2: Accuracy of the benchmark crisp system

categories the level of accuracy provided by rules decreases at a level by which they do not stand any more. This means that we can be confident on linkages between variables only with lower level of granularity of data to protect. By contrary, we can note the opposite for premises. Accuracy improves by increasing the number of fuzzy sets. This is due to the fact that rules become more precise but also more specific, thus loosing their generality. This is confirmed by the experiments with Arni, as depicted in Table 2. In this case, we obtain very accurate, but too specific rules.

Interpretability

Specificity is correlated to intepretability of rules. Indeed, looking at Table 3, we can observe how higher complexity in terms of NOR, TRL and ARL indexes (described in Section 4), thus lower interpretability of rules, is obtained at higher granularity of information. In particular Table 3 provides results for 3 and 5 categories when a 5-set partition is employed.

	Farni	FId3	Farni	FId3	Farni	FId3
3-category problem	5	5	5	5	1	1
5-category problem	7	20	10	51	1,42	2,55

Table 3: NOR(left), TRL (center) and ARL (right) indexes for the fuzzy systems

This aspect becomes even more evident in the case of crisp rule bases, as depicted in Table 4.

	NOR	TRL	ARL
3-category problem	6	17	2,8
5-category problem	15	51	3,4

Table 4: NOR(left), TRL (center) and ARL (right) indexes for the benchmark crisp system

However, NOR, TRL and ARL do not take into account the coverage provided by fuzzy partition. Table 5 shows the Nauck index for Farni and Fid3 (this index is not applicable to Arni). As $I_{Nauck} = Comp \times Part \times Cov$, $Cov = 1$ since we work with strong fuzzy partitions and $Part = 1/4$ since we use 5 fuzzy set partition, we get $I_{Nauck} = Comp/4$. The closer to 1 is I_{Nauck}, the more interpretable the system is. In this case we can note how interpretability improves with less granular information.

	Farni	FId3	Farni	FId3	Farni	FId3
3-category problem	0,6	0,6	0,25	0,25	0,15	0,15
5-category problem	0,5	0,1	0,25	0,25	0,12	0,02

Table 5: COMP(left), Part (center) and Nauck (right) indexes for the fuzzy systems

Examples of rules obtained by *FArni-rules* for the 3 and 5-category problems, is given below.

FArni-rules, 3-category, 5 fuzzy sets

TAXINC is *medium* $->$ **AGI** is *medium*
TAXINC is *low* $->$ **AGI** is *low*
TAXINC is *high*$->$ **AGI** is *high*
TAXINC is *very high* $->$ **AGI** is *high*
TAXINC is *very low* $->$ **AGI** is *low*

FArni-rules, 5-category, 5 fuzzy sets

FEDTAX is *medium* **AND TAXINC** is *high* $->$ **AGI** is *high*
TAXINC is *medium* **AND FEDTAX** is *low* $->$ **AGI** is *medium*
TAXINC is *low* **AND FEDTAX** is *low* $->$ **AGI** is *low*
ERNVAL is *very high* $->$ **AGI** is *very high*
FEDTAX is *very low* $->$ **AGI** is *very low*
FEDTAX is *high* $->$ **AGI** is *very high*
TAXINC is *very high* $->$ **AGI** is *very high*

These rule bases provide us many information about the existing relationships among the variable to protect (AGI) and the released information. Results show a strong dependence between some variables and AGI. Thus, the rule bases point out the real threats come from attributes such as *TAXINC* for the 3-category problem and *TAXINC*, *FEDTAX* and *ERNVAL* for the 5-category problem. On these variable we should paid attention in order to avoid a linkage to AGI, able to disclose its value withing a given range.

6 Conclusions and Future Directions

In this paper we investigated the application of fuzzy rules mining as means for discovering conditions able to infer sensitive information, also known as attribute disclosure, although approximated. An illustrative example has been discussed. Fuzzy rules are generally simpler and easier to interpret than other approaches, based on decision trees for example.

Results are still preliminary but encouraging. From the experiments we can conclude that general and interpretable rules can be built on data, revealing possible threats

to privacy. In the future we aim to study at which extent rules can be generalized, and what is the role of background knowledge in determining logical connections between data.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. *In 20th VLDB Conference*, September 1994.

[2] José M. Alonso, Luis Magdalena, and Gil González-Rodríguez. Looking for a good fuzzy system interpretability index: An experimental approach. *Int. J. Approx. Reasoning*, 51(1):115–134, 2009.

[3] Leo Breiman, editor. *Classification and regression trees*. The Wadsworth statistics/probability series. Wadsworth, Belmont, Calif., 1984.

[4] Jorge Casillas. *Accuracy Improvements in Linguistic Fuzzy Modeling*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003.

[5] Jadzia Cendrowska. Prism: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 27(4):349–370, 1987.

[6] Peter Clark and Tim Niblett. The cn2 induction algorithm. *Machine Learning*, 3:261–283, 1989.

[7] William W. Cohen. Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning*, 1995.

[8] Irene Díaz, José Ranilla, Luis J. Rodríguez-Muñiz, and Luigi Troiano. Identifying the risk of attribute disclosure by mining fuzzy rules. In Eyke Hüllermeier, Rudolf Kruse, and Frank Hoffmann, editors, *IPMU (1)*, volume 80 of *Communications in Computer and Information Science*, pages 455–464. Springer, 2010.

[9] Josep Domingo-Ferrer and Vicenç Torra. On the connections between statistical disclosure control for microdata and some artificial intelligence tools. *Inf. Sci. Inf. Comput. Sci.*, 151:153–170, 2003.

[10] Marcel Holsheimer and Arno P.J.M. Siebes. Data mining: the search for knowledge in databases. Technical report, Amsterdam, The Netherlands, The Netherlands, 1994.

[11] Nitin Indurkhya and Sholom M. Weiss. Iterative rule induction methods. *Appl. Intell.*, 1(1):43–54, 1991.

[12] Hisao Ishibuchi and Yusuke Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *Int. J. Approx. Reasoning*, 44(1):4–31, 2007.

[13] T. Joachims. Making large-scale support vector machine learning practical. In A. Smola B. Scholkopf, C. Burges, editor, *Advances in Kernel Methods: Support Vector Machines*. MIT Press, Cambridge, MA, 1998.

[14] Philip D. Laird. *Learning from good and bad data*. Kluwer Academic Publishers, Norwell, MA, USA, 1988.

[15] Ninghui Li and Tiancheng Li. t-closeness: Privacy beyond k-anonymity and ?-diversity. In *In Proceedings of IEEE International Conference on Data Engineering*, 2007.

[16] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE*, pages 106–115. IEEE, 2007.

[17] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. In *22nd IEEE International Conference on Data Engineering*, 2006.

[18] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Trans. Computers*, 26(12):1182–1191, 1977.

[19] Sreerama K. Murthy, Simon Kasif, Steven Salzberg, and Richard Beigel. Oc1: A randomized induction of oblique decision trees. In *AAAI*, pages 322–327, 1993.

[20] Detlef D. Nauck. Measuring interpretability in rule-based classification systems. In *Proc. of IEEE International Conference on Fuzzy Systems, 2002*, pages 196–201.

[21] Nikhil R. Pal and Sukumar Chakraborty. Fuzzy rule extraction from id3-type decision trees for real data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 31(5):745–754, 2001.

[22] Zdzisław Pawlak. *Rough Sets*. 1991.

[23] D. T. Pham, S. Bigot, and S. S. Dimov. Rules-f: a fuzzy inductive learning algorithm. In *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, pages 1433–1444, 2006.

[24] J. R. Quinlan. Constructing decision tree in c4.5. In *Programs of Machine Learning*, pages 17–26. Morgan Kaufman, 1993.

[25] Ross Quinlan. *c4.5: Programs of machine learning*. 1993.

[26] J. Ranilla, O. Luaces, and A. Bahamonde. A heuristic for learning decision trees and pruning them into classification rules. *AICom (Artificial Intelligence Communication)*, 16(2):in press, 2003.

[27] Jose Ranilla and Bahamonde Antonio. Fan: Finding accurate inductions. *International Journal of Human Computer Studies*, 56(4):445–474, 2002.

[28] Jose Ranilla and Luis J. Rodriguez-Muniz. A heuristic approach to learning rules from fuzzy databases. *IEEE Intelligent Systems*, 22(2):62–68, 2007.

[29] J Rives. Fid3: Fuzzy induction decision tree. In *Proc. 1st Int"l Symp. Uncertainty Modeling and Analysis*.

[30] Enrique H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, July 1969.

[31] Pierangela Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13:1010–1027, 2001.

[32] Xiaoxun Sun, Hua Wang, Jiuyong Li, and David Ross. Achieving p-sensitive k-anonymity via anatomy. In *Proceedings of the 2009 IEEE International Conference on e-Business Engineering*, pages 199–205, Washington, DC, USA, 2009. IEEE Computer Society.

[33] Ching-Hung Wang, Jau-Fu Liu, Tzung-Pei Hong, and Shian-Shyong Tseng. A fuzzy inductive learning strategy for modular rules. *Fuzzy Sets and Systems*, 103(1):91 – 105, 1999.

[34] Leon Willenborg and Ton de Waal. *Statistical Disclosure Control in Practice*. Springer Verlag, 1996.

[35] Yufei Yuan and Michael J. Shaw. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69(2):125 – 139, 1995.

[36] Lofti A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.