AMCIS 2008 Proceedings

Americas Conference on Information Systems (AMCIS)

2008

Attribute Noise-Sensitivity Impact: Model Performance and Feature Ranking

Mohamed Sidahmed
University of Colorado Denver, mohamed.sidahmed@ucdenver.edu

Attribute Noise-Sensitivity Impact: Model Performance and Feature Ranking

Mohamed Sidahmed
University of Colorado Denver
mohamed.sidahmed@ucdenver.edu

ABSTRACT

Developing robust and less complex models capable of coping with environment volatility is the quest of every data mining project. This study attempts to establish heuristics for investigating the impact of noise in instance attributes data on learning model volatility. In addition, an alternative method for determining attribute importance and feature ranking, based on attribute sensitivity to noise is introduced. We present empirical analysis of the effect of attribute noise on model performance and how it impacts the overall learning process. Datasets drawn from different domains including Medicine, CRM, and security are employed by the study. Using proposed technique has practical implications by supporting building low volatile, high performance predictive models prior to production deployment. Also the study has implications for research by filling the gap in attribute noise research and its impact.

KEYWORDS

Attribute noise, model performance, feature ranking, ROC Curve, decision trees, classification, data mining, knowledge discovery.

INTRODUCTION

As the field of data mining secures its position as an influential cross-industry modus operandi, more emphasis is placed on exercising appropriate learning techniques that drive non-trivial discoveries. Classification algorithms and learning methods became one of the main important tools in decision making process. It has been recognized as an efficient tool in various disciplines such as medicine, business, marketing, statistics, and many others. The importance of these algorithms comes from its competence to predict outcome and uncover association more efficiently than human experts, also its capability to reduce cost and error level in decision making. Yet accuracy of decisions made based on trained models depends to large extent on underlying data and its quality.

The goal of this research is to investigate the impact of attribute noise-sensitivity on a classifier performance, and determine attribute importance. Despite the fact that noise is context related, this study identifies noise in instance attributes as those external errors introduced to attributes values. Examples of such errors include missing or unknown values, invalid values, and wrong values. Some sources of attribute noise can be ascribed to measurement, recording, or data entry errors to name only a few.

This study examines attribute noise-sensitivity as a proxy for attribute importance. Earlier work has identified robustness as one of model's performance characteristics. Models that demonstrate high volatility to changes in attribute or class prevalence are considered less favorable. As more data mining techniques continue to be incorporated for advanced decision making, the accuracy and robustness of trained models should be immune for real life noisy environments. To take remedy of real-life noisy data, we propose an alternative approach to determine attribute ranking based on noise-sensitivity. This method utilizes attribute noise-sensitivity and its impact on classifier performance as an indicator of attribute importance.

RELATED WORK

Prior research in noise detection placed special emphasis on class noise more than attribute noise because it is easier to handle (Zhu and Wu, 2004). This finding is confirmed by Khoshgoftaar and Van Hulse (2005), who argued that different

techniques were developed and used to identify the instances with class noise with high degree of accuracy, where very few techniques are offered to detect attribute noise that is due to task complexity.
One study that looked at attribute noise attempt to predict learning with random noise in the attributes (Goldman and Sloan, 1995). Both random classification noise (Angluin and Laird, 1988) and arbitrary, malicious attribute noise (Kearns and Li, 1993) have been studied within algorithm performance domain.

A different approach by Quinlan (1986) assumed that the same random noise was present in both the training and testing phases of the learning. This approach virtually guarantees that classification noise will seem to be the most harmful, since each percentage point increase in classification noise for the data for the performance phase should lead to roughly a full percentage point decrease in the learner's performance. Research has also shown that noise in important attributes have more impact on model performance (Laired, 1988; Goldman and Sloan, 1995)

Zhu and Wu, (2004) study put forward a systematic evaluation of the impacts of noise on learning, through testing the class and attribute noise. The focal point of the study was on attribute noise as it has been less discussed by other researchers. Most of datasets used in the study don't contain noise; therefore a manual method was used to insert a noise level to these datasets, for class and attributes. The error values that are introduced to the attributes are done at a level x.100%. The results of the study show that existence of class noise lead to decrease classification accuracy, they also find that removing the noise will eventually increase the classification accuracy. Regarding the attribute noise the results show that the highest classification accuracy is always from classifier trained from clean training set, where the lowest classification accuracy resulted from the classifier trained from corrupted training set in classifying a corrupted test set. Also in case of no attribute noise in the datasets, adopting cleaning attribute noise process will improve the classification accuracy. In general the results show that attribute noise is less harmful than class noise.

Research in noise in learning setting, have focused on the case where the training data is corrupted by noise, but the data used for evaluating the learning in the performance phase is noise-free (Goldman and Sloan,1995). Although this might be reasonable choice for malicious noise, random noise scenarios should account for corrupted data in both training and evaluating sets.

Recent study by Mannino, Yang and Ryu (2007) presented an empirical comparison of different classification algorithms about noise level between training and field data. The experiment used repeated measures design to account for algorithm, noise level, and training set size as factors that affect individual algorithm performance. The study results show that it might not be worthwhile to achieve very low noise level. Although, over representative training noise should be avoided. The results of the interaction among algorithms, noise level and training set size, show that these general results may not apply to particular practice situations.

New approach for noise detection, Partitioning- and Rule-based Filter (PRBF) was proposed for detecting class noise (Melville et al. 2004). The proposed approach combined four different components to achieve high-accuracy noise detection mechanism. These components are repeated data partitioning, inclusive evaluation, un-weighted voting and dual-two-class-classifiers. The results of case study experiments on real-world datasets show that PRBF works well as a noise detection tool. Also, this tool proved to be useful with datasets containing nominal and numerical attributes. The tests show that the proposed filter outperforms the partitioning Filter with respect to noise detection accuracy.

RESEARCH METHODOLOGY

This study aims to investigate the impact of attribute noise on leaning model performance. Also we look at the determining attribute importance in terms of its influence on model predictive power. In this section a description of the research hypothesis and experimental methodology used to test the research hypothesis will be presented. Also a description of the problem domain, experimental design, and statistical tests will be offered.

In particular we designed the study to investigate the following four research questions/hypotheses:

- In general, an attribute that is highly sensitive to noise, is likely to have larger impact on classifier's performance compared to a less noise-sensitive one.

- How attribute importance could be determined based on its sensitivity to change of noise level?
- What's the level of agreement/disagreement between attribute sensitivity deterministic method for attribute importance and other feature ranking techniques
- We argue that attribute noise-sensitivity could be employed as a measure of attribute importance (i.e. A higher noise-sensitive attribute imply superior rank)

Experiment Design

Three datasets have been utilized by the study. Our goal was to validate results across various degrees of complexity. These include relatively small, medium and high dimensionality datasets. In terms of size, the datasets are not considered trivial, as each has practical number of instances. The smallest, *Pima Indians Diabetes,* is relatively simple dataset with eight attributes. The second dataset, *Customer Churn*, has moderate complexity, with nineteen attributes plus the class. *Spam Email* is a complex, high dimensionality dataset. It consists of fifty seven attributes, in addition to, the class attribute. Specifications of the datasets used in this study are presented in Table 1.

Table 1: Study Datasets Specifications

Dimensionality	Dataset	Attributes	Classes	Class Prevalence	Instances	Missing Values	Source
Small	Diabetes	8	2	Class 1: 34.9%	768	None	UCI Repository (Sigillito, 1990)
Medium	Churn	20	2	Churn: 14.5%	3333	None	Duke Center for CRM (Fuqua, 2003)
Large	Spam Email	57	2	Spam: 39.4%	4601	Yes	UCI Repository (Forman, 1999)

We manually corrupt each attribute in the dataset with noise level η .100%. The value of attribute A_i is assigned a random number from attribute space.

To perturb each dataset attributes, a random sample of 30% from each dataset was obtained using SPSS. Noise level was controlled across all attributes. Noise introduced to each attribute according to the pseudo-technique presented next:

1. Descriptive statistics were calculated for the datasets attributes to define the maximum and minimum values for each attribute. Table 2 shows the ranges of values for each attribute of churn dataset as an example.
2. For numeric attributes, we used Excel Rand function to generate a random value between [min-max] for a random 30% subset of the dataset. Formula used: =RAND ()*(MIN-MAX) +MAX.
3. For binary attributes we selected the opposite
4. Class attribute is excluded and it was not corrupted.
5. Some string value attribute just altered slightly
6. We repeat these steps for every attribute and save as separate datasets, resulting in *i* different samples for each dataset. Where *i* is the number of attributes in a specific dataset.

Table 2: Churn Attributes Descriptive Statistics

Churn Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
AccLen	3333	1	243	101.06	39.822	.097	.042
IntlPlan	3333	0	1	.10	.296	2.726	.042
VMailPlan	3333	0	1	.28	.447	.999	.042
VMailMessage	3333	0	51	8.10	13.688	1.265	.042
DayMins	3333	.0	350.8	179.775	54.4674	-.029	.042
DayCalls	3333	0	165	100.44	20.069	-.112	.042
DayCharge	3333	.00	59.64	30.5623	9.25943	-.029	.042
EveMins	3333	.0	363.7	200.980	50.7138	-.024	.042
EveCalls	3333	0	170	100.11	19.923	-.056	.042
EveCharge	3333	.00	30.91	17.0835	4.31067	-.024	.042
NightMins	3333	23.2	395.0	200.872	50.5738	.009	.042
NightCalls	3333	33	175	100.11	19.569	.032	.042
NightCharge	3333	1.04	17.77	9.0393	2.27587	.009	.042
IntlMins	3333	.0	20.0	10.237	2.7918	-.245	.042
IntlCalls	3333	0	20	4.48	2.461	1.321	.042
IntlCharge	3333	.00	5.40	2.7646	.75377	-.245	.042
CustServCalls	3333	0	9	1.56	1.315	1.091	.042
Churn	3333	0	1	.14	.352	2.018	.042
Valid N (listwise)	3333						

Classification Algorithm

To overcome commensurability issues that hinder comparison of dissimilar units of analysis, all of the experiments described in this research use C4.5, a program for inducing decision tree from pre-classified training examples introduced by Quinlan (1993). The importance of this program comes from its popular use as disjunctive concept and the ability to easily modifying it (Weiss and Hirsh, 1998). Latest survey revealed that 79% of data mining professionals use decision trees (Rexer et al, 2007).

Weka 3.4, open source data mining tool, implementation of the algorithm (J48) has been utilized in the experiment.

Feature Selection Methods

Proposed technique of determining attribute importance based on its noise-sensitivity has been compared to other conventional feature selection methods. Several feature selection taxonomies are suggested in the literature. Though a relevant categorization divided ranking algorithms into those which evaluate individual attributes and those which evaluate subsets of attributes (Hall and Holmes, 2003).

Our basis of comparison is attribute ranking. Four feature ranking methods are employed by the study as benchmarks; each evaluated one attribute at a time:

- Information Gain Attribute Evaluation
- Gain Ratio Attribute Evaluation
- OneR Attribute Evaluation
- Chi-Squared Attribute Evaluation

STUDY RESULTS

This section introduces the results of the study and provides details of the analysis. To account for external validity, experiment was replicated across multi-dataset with various domains, sizes, and complexity. For each perturbed attribute in a dataset, a learning classifier has been built using 10-fold cross validation (10Fcv) training/testing iterations. Results of single-perturbed attribute learned classifiers for each dataset are presented next. Model performance was captured by the area under the curve (AUC) measure.

Table 3 present AUC values for each trained model for diabetes dataset, based on its perturbed attribute. Greater AUC value indicates better model performance. Figure 1 shows the receiver operator characteristic (ROC) curves for Diabetes perturbed attributes models. Curves further to the top left corner have larger AUC areas and represent better classifiers. ROC is considered superior measure than widely reported accuracy rate, because it's class prevalence independent.

AUC values and ROC curves for Churn dataset models are presented in tables 4 and figure 2, while those for Spam dataset are included in table 5 and figure 3 respectively.

Table 3: Diabetes Perturbed Attributes Models Performance

Attr Order	Model	AUC
2	PlasmaGlucoseConcen	0.7215
6	BodyMassIndex	0.7223
5	SerumInsulin	0.7290
1	NumTimesPregnant	0.7340
7	PedigreeFunction	0.7468
3	DiastolicBloodPres	0.7513
	NoPerturbedAttr	*0.7514*
4	TricepsSkinThickness	0.7520
8	Age	0.7757

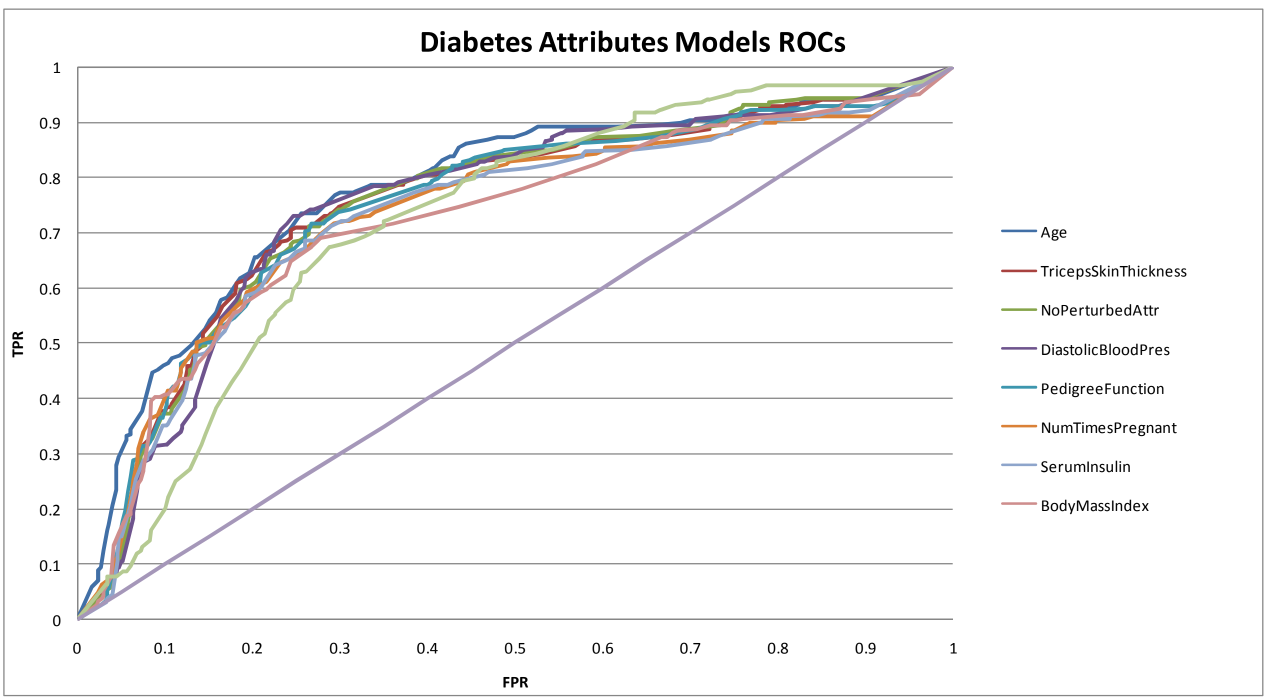

Figure 1: Diabetes Perturbed-Attribute Classifiers Performance

Table 4: Churn Perturbed Attribute Models Performance

Attr Order	Model	AUC
20	CustServCalls	0.7469
5	IntlPlan	0.7795
18	IntlCalls	0.8656
17	IntlMins	0.8684
12	EveCalls	0.8696
14	NightMins	0.8697
7	VMailMessage	0.8697
6	VMailPlan	0.8701
13	EveCharge	0.8714
2	AccLen	0.8716
9	DayCalls	0.8716
8	DayMins	0.8716
19	IntlCharge	0.8716
15	NightCalls	0.8716
16	NightCharge	0.8716
	NoPerturbedAttr	*0.8716*
11	EveMins	0.8717
10	DayCharge	0.8734
1	State	0.8761
3	AreaCode	0.8779

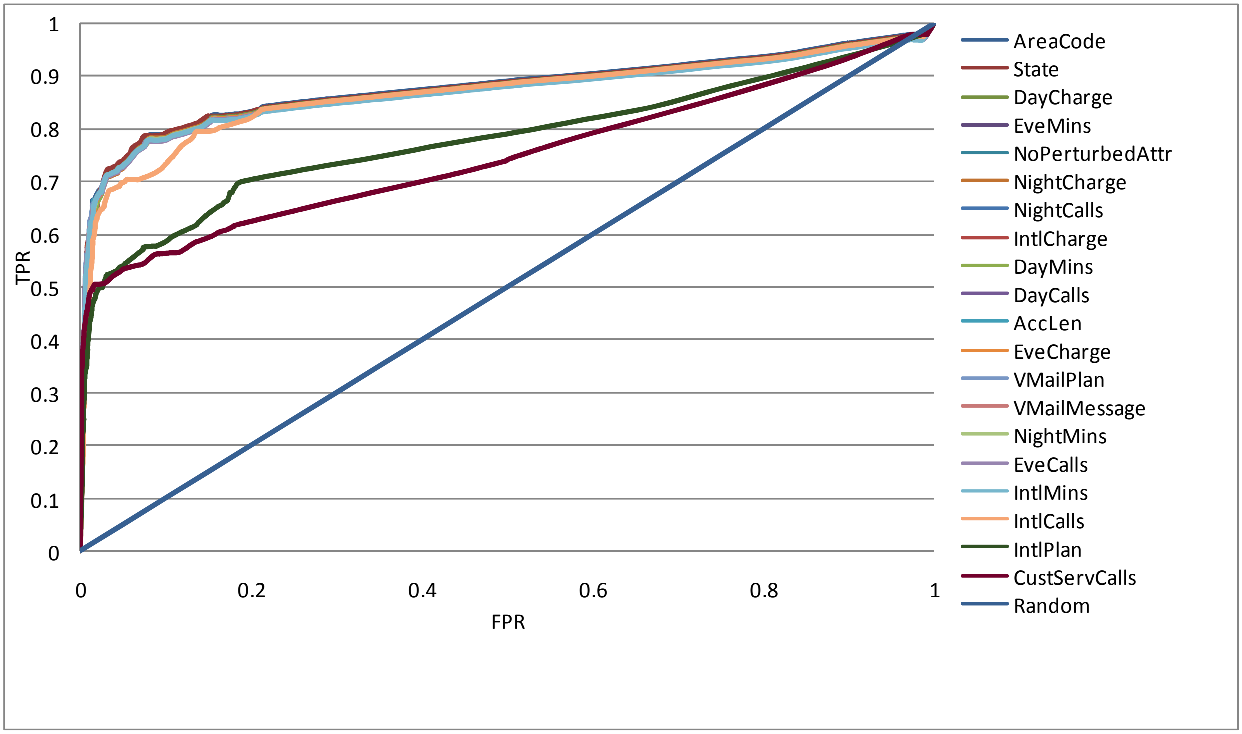

Figure 2: Churn Perturbed-Attribute Classifiers Performance

Table 5: Spam Perturbed Attribute Models Performance

Rank	Model	AUC	Rank	Model	AUC
1	char_freq_lparanthesis	0.9080	30	word_freq_cs	0.9394
2	word_freq_hp	0.9319	31	word_freq_mail	0.9396
3	word_freq_edu	0.9344	32	word_freq_data	0.9396
4	word_freq_george	0.9360	33	word_freq_original	0.9397
5	word_freq_you	0.9362	34	word_freq_labs	0.9399
6	char_freq_pang	0.9362	35	word_freq_remove	0.9400
7	word_freq_meeting	0.9363	36	word_freq_addresses	0.9401
8	word_freq_order	0.9367	37	word_freq_lab	0.9402
9	word_freq_85	0.9368	38	word_freq_technology	0.9402
10	word_freq_our	0.9369	39	word_freq_1999	0.9403
11	word_freq_re	0.9372	40	word_freq_address	0.9404
12	char_freq_lsquareparan	0.9377	41	word_freq_credit	0.9405
13	word_freq_parts	0.9378	42	word_freq_project	0.9405
14	char_freq_dollar	0.9379	43	*NoPerturbedAttr*	*0.9407*
15	char_freq_semicolon	0.9380	44	word_freq_receive	0.9407
16	word_freq_email	0.9381	45	word_freq_business	0.9407
17	capital_run_length_average	0.9383	46	word_freq_font	0.9407
18	word_freq_money	0.9384	47	word_freq_direct	0.9407
19	word_freq_3d	0.9385	48	word_freq_report	0.9408
20	word_freq_make	0.9386	49	word_freq_650	0.9408
21	char_freq_pound	0.9387	50	capital_run_length_total	0.9409
22	word_freq_table	0.9388	51	word_freq_over	0.9411
23	word_freq_will	0.9391	52	word_freq_pm	0.9413
24	word_freq_free	0.9391	53	word_freq_internet	0.9421
25	word_freq_857	0.9391	54	capital_run_length_longes	0.9421
26	word_freq_conference	0.9391	55	word_freq_all	0.9422
27	word_freq_hpl	0.9392	56	word_freq_telnet	0.9426
28	word_freq_415	0.9393	57	word_freq_people	0.9433
29	word_freq_your	0.9394	58	word_freq_000	0.9434

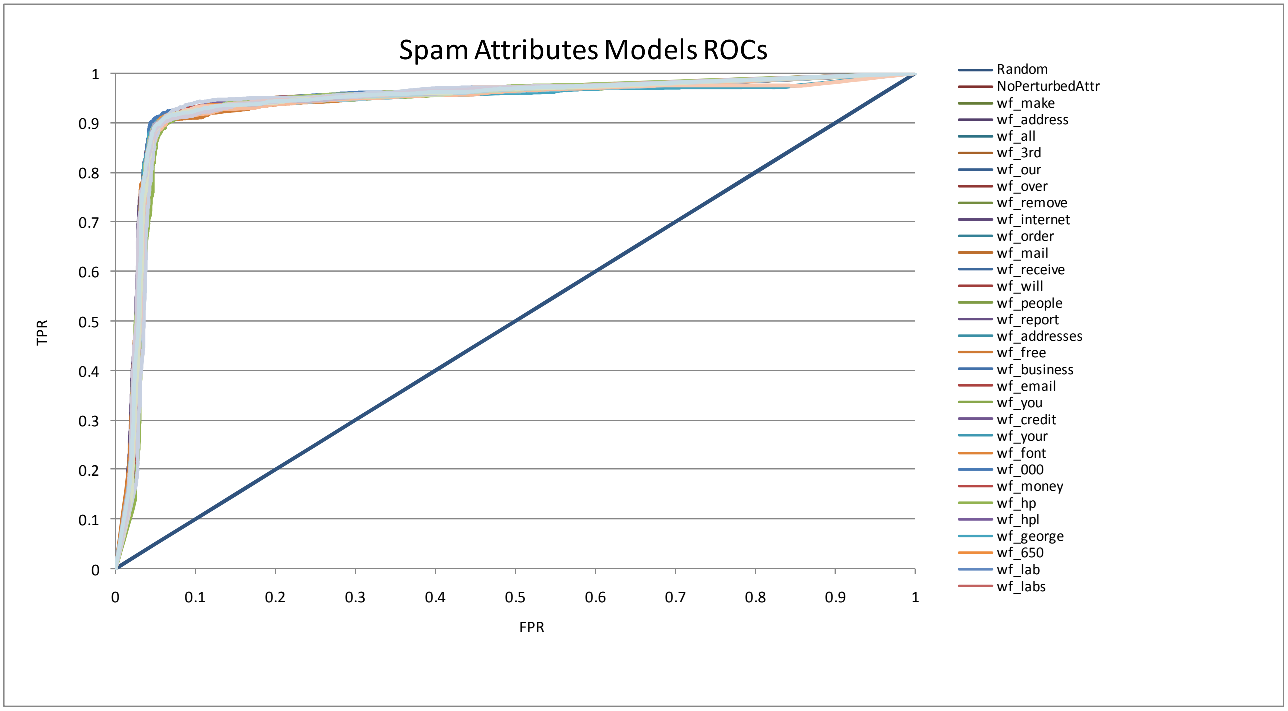

Figure 3: Spam Perturbed-Attributed Models Performance

Rater Agreement Tests

Agreement among ratings of multiple judges has been a common interest in numerous fields. Landis and Koch's (1977) unified model of evaluating observer agreement reflect the extent to which the observers agree among themselves as functions of observed proportions obtained from underlying multidimensional contingency tables. In this study we are concerned with judges' agreement on numeric rankings. However, there is no general consensus regarding appropriate statistical methods to analyze raters' agreement (Mackinnon, 2000).

Nonparametric Statistical Tests

Two main non parametric statistical evaluation techniques have been used to compare our method achievement to other attribute ranking techniques. We found that the nonparametric tests for multiple related samples are useful alternatives to a repeated measures analysis of variance. We used Friedman and Kendall's W (Robson, 1993; Siegel, 1956).

Friedman test is an equivalent of one sample repeated measures. It's used to determine if the attributes are of equal importance based on each method. It ranks the scores in each row of the data file independently of every other row.

Kendall's Coefficient of Concordance (W) was used as an index of agreement among n rankings (4 feature ranking raters in addition to the proposed one). The test requires that all members be ranked, Kendall's W measures the agreement between φ (φ =5) observers who have ranked the same objects or individuals. Numerical ratings or scores can be reduced to ranks so as to make Kendall's coefficient more widely applicable. W assumes a range of values between 0 and 1, where zero indicates total rankings disagreement, and one represent identical rankings.

In order to obtain W:

- the ranks received by each individual Feature Selection (FS) and attribute noise-sensitivity model are summed and the squared deviations of these sums from the mean are computed
- This quantity is then divided by the sum of the squared deviations which would result from perfect agreement among the rankers

Each of the five methods has already performed this ranking. For each method, these ranks are summed and then divided by the number of methods to yield an average rank for each attribute. Results were obtained using the statistical software, SPSS 13. In addition, Attribute's order in a dataset has been appended to its name for reference convenience. This naming convention is adopted throughout the study for all datasets.

Diabetes Statistical Tests

Analysis yield a Kendall's coefficient value of 0.787 or 79%, indicating a high degree of inter-method concordance. Interesting results showed total agreement (μ =1, σ =0) on '*PlasmaGlucoseConcen*' as the highest ranked attribute. Friedman test's Chi-square value 27.533 with 7 degrees of freedom (*n* attributes - 1), was significant at 0.001; confirming close agreement level between the five different methods. Overall, our Noise-sensitive attribute ranking method performed well. Method results were in agreement with at least one (lower bound) of the other four attributes ranking techniques. The upper bound was full agreement with the four methods on attribute ranking.

Churn Statistical Tests

Similarly, Churn average attribute ranking was measured among five methods. Uniform non-parametric tests for Noise-sensitive attribute ranking method and other four benchmarking approaches were performed on this data set. However, Phone attribute was not included in the trained models since it's used as an identifier.

Both Friedman and Kendall's W tests were significant at 0.001 confidence level. Chi-square value was 114.835 with 18 degrees of freedom. The Kendall's W inter-method concordance was 0.709 or 71%. Results for this dataset found that the five methods tend to rank '*CustServCalls*', '*DayMins*', and '*IntlPlan*' attributes higher than the other attributes.

Spam Statistical Tests

Spam results demonstrated that *Char_freq_pand_52* (Attribute 52) has the highest agreement among raters (i.e. five attribute ranking techniques including our method) with mean rank of 3.80, while *word_freq_report_14* attribute had the highest mean.

Friedman test Chi-square value confirms agreement on highlighted attributes shown in table 5. The Chi-square value 140.2 was significant at 0.001 level. We conclude that the methods do not have equal attribute importance impact on performance. Although the study observed a Kendall's coefficient of 0.5, still it indicates a relatively high degree of inter-method concordance.

DISCUSSION

Key contribution of the study is a proposed alternative method to determine attribute noise impact on model performance and attribute rank determination. Such impact has been captured in different experiments with various perturbed attribute models, across multiple datasets from different domains. While performance degrades for most trained models due to noise introduction, a noticeable improvement in performance for some perturbed attribute has been observed. This could be attributed to either one or both of the following reasons:

1. change of attribute distribution post noise introduction
2. Original dataset may contain prior noise treated by introduced one

In practice, different attribute ranking methods produce different results. Therefore it's unlikely to achieve 100% agreement between proposed method and other ones.

One of the fundamental assumptions of the study is that attributes correlation does not impose significant impact when training single-perturbed attribute models. We introduced noise to each attribute in isolation.

Results confirmed that attributes that are highly sensitive to noise have larger impact on overall classifier performance. These sensitive attributes should be carefully examined during model building for fine tuning of the model and decide about inclusion/ exclusion. The noise-sensitivity method showed wide consensus for attribute ranking with other feature selection methods. Although no two methods will produce 100% agreement all across the feature space, still agreement on key attributes is significant.

CONCLUSION

This study presented an empirical analysis of the effect of attribute noise on model performance and how it impacts the overall learning process. Three datasets drawn from different domains; Medicine, CRM, and fraud; Data quality is an important issue in all fields that rely on the analysis of data to make decisions or confirm hypotheses. This research also presents a method for determining attribute importance based on noise sensitivity, a noise assessment technique that builds on ranking of attributes noise sensitivity.

In related literature, some work has been done to identify instances with either class or attribute noise. These instances can undergo further treatment before building a learner or completing any data mining analysis. The proposed approach was validated with multiple datasets from business, marketing, and medicine. The experiment was carefully designed to insure both internal and external validity. Several statistical techniques have been employed by the study to ensure results validity. However, the experiment could be extended to include different learning algorithms and data sets.

As an implication of the study, feature selection could be achieved with reliable confidence in noisy inflicted environments. This is an enhanced preprocessing step for classifier training that takes into account attribute noise as a factor. Significant project savings could be achieved, since substantial investment on data cleansing and quality should be directed to those important attributes that are highly sensitive to noise. Using proposed technique provide preemptive solution for building low volatile models prior production deployment. The study also has implications for research by filling the gap in attribute noise research and its impact.

REFERENCES

1. Angluin, D. and Laird, P. (1988) Learning from noisy examples. *Machine Learning,* 2, 4, 343-370.
2. Forman, George (1999) UCI Machine Learning Repository, Department of Information and Computer Science, University of California, Irvine, (http://archive.ics.uci.edu/ml/datasets/Spambase)
3. Fuqua (2003) The Center for Customer Relationship Management, Fuqua School of Business, Duke University, (http://www.ccrmatduke.org/centers/ccrm/)
4. Goldman, S. A. and Sloan, R. H. (1995) Can PAC learning algorithms tolerate random attribute noise?, *Algorithmica*,14, 1, 70-84.
5. Hall, Mark A. and Holmes, Geoffrey (2003) Benchmarking Attribute Selection Techniques for Discrete Class Data Mining, *IEEE Transactions on Knowledge and Data Engineering* , 15, 6, 1437-1447.
6. Kearns, M. and Li, Ming (1993) Learning in the presence of malicious errors. *SIAM Journal of Computing,* 22, 4, 807-837.
7. Kendall, Maurice G. (1948) Rank correlation methods, London: Griffin and Co., Chapters 6 and 7
8. Khoshgoftaar, Taghi M. and Van Hulse, Jason (2005) Identifying noisy features with the Pairwise Attribute Noise Detection Algorithm," *Intelligent Data Analysis*, 9, 6, 589-602.
9. Laird, P. *Learning from Good and Bad Data*, Kluwer Academic Publishers, Norwell, MA, 1988
10. Landis, J. Richard and Koch, Gary G. (1977) The Measurement of Observer Agreement for Categorical Data, *Biometrics*, 30, 1, 159-174.
11. Mackinnon, Andrew (2000) A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement, *Computers in Biology and Medicine*, 30, 3, 127-134.

12. Mannino, M., Yang, Y., and Ryu, Y. (2007) Classification Algorithm Sensitivity to Training Data with Non Representative Attribute Nois*e, Proceedings of the Thirteenth Americas Conference on Information Systems*, August 9-12, Keystone, CO, USA*, Colorado State University*

13. Melville, Prem; Shah, Nishit; Mihalkova, Lilyana; and Mooney, Raymond J. (2004) Experiments on Ensembles with Missing and Noisy Data, *Proceedings of the Fifth International Workshop on Multiple Classifier Systems* (MCS-2004), F. Roli, J. Kittler, and T. Windeatt (Eds.), Lecture Notes in Computer Science, Vol. 3077, pp. 293-302, Cagliari, Italy, Springer Verlag.

14. Quinlan, J. R. (1986) The effect of noise on concept learning. *Machine Learning, An Artificial Intelligence Approach,* Volume II, Morgan Kaufmann, Los Altos, CA, 149-166.

15. Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA

16. Rexer, Karl; Gearan, Paul; Allen, Heather N. (2007) Surveying the Field: Current Data Mining Applications, Analytic Tools, and Practical Challenges, Rexer Analytics, Winchester, MA

17. Robson, C. (1993). *Realworld research.* Oxford: Blackwell.

18. Siegel, S. (1956). *Nonparametric statistics.* London: McGraw-Hill Kogakusha, Ltd.

19. Sigillito, Vincent (1990) UCI Machine Learning Repository, Department of Information and Computer Science, University of California, Irvine, (http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes)

20. Weiss, G. M. and Hirsh, H. (1998) The problem with Noise and Small Disjuncts, proceedings of the 15th International Conference on machine Learning, Morgan Kaufmann, 574-578.

21. Xiao, Y., Khoshgoftaar, T.M., and Seliya*,* N. (2005) The partitioning- and rule-based filter for noise detection, *Information Reuse and Integration, Conf, 2005. IRI -2005 IEEE International Conference on*. Aug 2005, 205-210.

22. Zhu, X. and Wu, X. (2004) Class noise vs. attribute noise: a quantitative study of their impacts, *Artificial Intelligence Review*, 22, 3, 177–210.