Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 268063, 8 pages
http://dx.doi.org/10.1155/2013/268063

Research Article

Feature Selection in Decision Systems: A Mean-Variance Approach

Chengdong Yang,[1] Wenyin Zhang,[1] Jilin Zou,[2] Shunbo Hu,[1] and Jianlong Qiu[3]

[1] *School of Informatics, Linyi University, Linyi 276005, China*
[2] *Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 211189, China*
[3] *School of Science, Linyi University, Linyi 276005, China*

Correspondence should be addressed to Jianlong Qiu; qiujianlong@lyu.edu.cn

Received 6 April 2013; Accepted 24 April 2013

Academic Editor: Guanghui Wen

Uncertainty measure is an important implement for characterizing the degree of uncertainty. It has been extensively applied in pattern recognition and data clustering. Because of instability of traditional uncertainty measures, mean-variance measure (MVM) is utilized to perform feature selection, which could depress disturbances and noises effectively. Thereby, a novel evaluation function based on MVM is designed. The forward greedy search algorithm (FGSA) with the proposed evaluation function is exploited to perform feature selection. Experiment analysis shows the validity and effectiveness of MVM.

1. Introduction

Rough sets, originated by Pawlak [1] in 1980s, is a powerful mathematical tool to deal with inexact, uncertain, and vague knowledge in information systems. It has been drawing extensive attention in theory and applications in artificial intelligence, pattern recognition, data mining, intelligence information processing, decision support, image processing, feature selection, neural computing, conflict analysis, and knowledge discovery [2–10].

Uncertainty measure is an important implement for characterizing the degree of uncertainty in rough set theory. It has been extensively applied in pattern recognition and data clustering. However, this paper reveals the issue that classical uncertainty measures are sensitive to disturbances or noises. Therefore, a novel uncertainty measure, called mean-variance measure (MVM), was proposed to characterize the degree of uncertainty of rough sets in paper [11]. Since it takes fully information in the boundary region into account, MVM is more robust and effective than classical uncertainty measures in depressing disturbances and noises.

As an important application of rough sets in artificial intelligence and machine learning, feature selection or attribute reduction in information systems has been drawing wide attention. due to the fact that excessive features or attributes usually confuse learning algorithms, cause significant slowdowns in learning processes, and increase risks of learned classifiers to over-fit the training data [4, 12].

Unfortunately, it has been proved that finding all reducts or finding an optimal reduct (a reduct with the least number of attributes) is an NP-complete problem [13]. Many researchers devote themselves to finding an efficient reduct by optimization techniques. The forward greedy search algorithm (FGSA), also called hill-climbing algorithm or greedy algorithm, is such an optimal technique for finding one reduct quickly and has been extensively investigated [14–16]. A key ingredient of FGSA lies in establishing an evaluation function to examine importance of each feature or attribute in databases. The evaluation function induced by the classical uncertainty measure, that is, the Pawlak's roughness or dependency, has been successfully applied in rough sets based feature selection [17, 18]. Along with the development of rough sets, attribute reduction has been studied extensively in the past decade, such as fuzzy rough sets based attribute reduction [19–23], neighborhood rough sets based attribute reduction [24], cross-entropy based attribute reduction [25], tolerance rough sets based attribute reduction [26], cost based attribute reduction [27, 28], and dynamic attribute reduction

[29, 30], and extended rough set based attribute reduction [31], cover rough sets based attribute reduction [14, 32], covering generalized rough sets based attribute reduction [33], variable precision rough sets based attribute reduction [34]. Nevertheless, the classical uncertainty measure is not robust and maybe fluctuates largely only with minor disturbances. Even a little change in information systems may produce an unpredictable fluctuation of this uncertainty measure.

The mean value and variance in probability theory, able to be used to analyze preciously data, have been widely discussed in portfolio optimization and portfolio selection [35–38]. They are considered as an arbitrator which is used to determine whether a group of data is robust and stable. For example, two shooters obtain the same score (mean value). If one has to be chosen to take part in a tournament, which one should be chosen reasonably? Apparently, the one with a less variance score would like to be chosen. In this paper, the notions of mean value and variance are introduced into information systems as an arbitrator to evaluate the uncertainty degree. A novel uncertainty measure, called mean-variance measure (MVM), is proposed. MVM firstly calculates the mean of every object, and then all objects' variances are taken into account. The effect caused by disturbances of data in decision systems on MVM will decrease, since a tiny alteration of values will not result in a large change of variance.

Based on the new notion of MVM, an evaluation function called D-MVM in decision systems is further designed. The designed evaluation function takes full information in positive region and boundary region into account.

This paper is organized as follows. Some elementary concepts on rough sets and MVM are reviewed in Section 2. Section 3 investigates the issue on feature selection in decision systems by MVM. Experimental results and analysis are given in Section 4, and Section 5 concludes this paper.

2. Preliminaries

2.1. Rough Sets. This section briefly outlines some basic notions on rough sets.

Definition 1. An information system is a pair $S = (U, \mathrm{AT})$ satisfying

(1) U is a nonempty finite set of objects;

(2) AT is a nonempty finite set of attributes;

(3) for every $a \in \mathrm{AT}$, there is a mapping $a : U \to V_a$, where V_a is the set of values.

Definition 2. Given an information system $S = (U, \mathrm{AT})$ and $P \subseteq \mathrm{AT}$, an indiscernibility relation on U is defined by

$$R_P = \{(x, y) \in U \times U \mid a(x) = a(y),\ \forall a \in P\}. \quad (1)$$

Obviously, R_P is an equivalent relation induced by the attribute set P. $[x]_{R_P} = \{y \in U \mid (x, y) \in R_P\}$ is referred to as the equivalence class of x with respect to R_P. A partition of U induced by the equivalent relation R_P can be denoted by

$$\frac{U}{R_P} = \{P_1, P_2, \ldots, P_n\}, \quad (2)$$

where P_i is some equivalence class of R_P in U, $i = 1, 2, \ldots, n$. U/R_P and $[x]_{R_P}$ are, respectively, denoted by U/P and $[x]_P$, for short, when no ambiguity arises in this paper.

Definition 3. Given an information system $S = (U, \mathrm{AT})$, $P \subseteq \mathrm{AT}$, and $X \subseteq U$, the lower approximation and the upper approximation of X with respect to P are defined, respectively, by

$$\begin{aligned} \underline{P}(X) &= \bigcup \left\{P_i \in \frac{U}{P} \mid P_i \subseteq X\right\}, \\ \overline{P}(X) &= \bigcup \left\{P_i \in \frac{U}{P} \mid P_i \cap X \neq \emptyset\right\}. \end{aligned} \quad (3)$$

Definition 4. Given an information system $S = (U, \mathrm{AT})$, a partial ordering relation $\preceq$ in the family $\{U/B \mid B \subseteq \mathrm{AT}\}$ is defined as

$$\frac{U}{P} \preceq \frac{U}{Q}, \quad (4)$$

if and only if for any $P_i \in U/P$, there exists a $Q_j \in U/Q$ such that $P_i \subseteq Q_j$, where $U/P = \{P_1, P_2, \ldots, P_n\}$ and $U/Q = \{Q_1, Q_2, \ldots, Q_m\}$ are partitions induced by $P, Q \subseteq A$, respectively.

Q is said to be coarser than P, or P is finer than Q, if $U/P \preceq U/Q$. P is said to be strictly finer than Q, denoted by $U/P \prec U/Q$, if $U/P \preceq U/Q$ but $U/P \neq U/Q$.

Proposition 5. *Given an information system $S = (U, \mathrm{AT})$, $P, Q \subseteq AT$, and if $P \subseteq Q$, then $U/Q \preceq U/P$.*

From Proposition 5, the more attributes an information system contains, the finer the corresponding partition is. Therefore, U/AT is the finest one among partitions induced by all subsets of AT.

The classical uncertainty measure is defined as follows.

Definition 6. Given an information system $S = (U, \mathrm{AT})$ or an incomplete information system $IS = (U, \mathrm{AT})$, $P \subseteq \mathrm{AT}$, and $X \subseteq U$, the roughness of X is defined as

$$\rho_P(X) = 1 - \frac{|\underline{P}(X)|}{|\overline{P}(X)|}. \quad (5)$$

The quantity $\rho_P(X)$ characterizes the uncertainty degree of X with respect to P. When $\rho_P(X) = 0$, X is said to be definable; otherwise, it is said to be rough.

When AT is divided into two nonempty sets C and D such that $C \cap D = \emptyset$, then $S = (U, \mathrm{AT})$, denoted by $S = (U, C \cup D)$, is called a decision system, C is called the conditional attribute set, and D is called the decision attribute set.

x_1	x_2	x_5	x_6	x_9	x_{10}
x_3	x_4	x_7	x_8	x_{11}	x_{12}
x_{13}	x_{14}	x_{17}	x_{18}	x_{21}	x_{22}
x_{15}	x_{16}	x_{19}	x_{20}	x_{23}	x_{24}
x_{25}	x_{26}	x_{29}	x_{30}	x_{33}	x_{34}
x_{27}	x_{28}	x_{31}	x_{32}	x_{35}	x_{36}

X

(a) X

0	0	0	0	0	0
0	1	1	1	1	0
0	1	1	1	1	1
0	1	1	1	1	0
0	1	1	1	1	0
0	0	0	0	0	0

X

(b) $A(X)$

1/4	1/4	1/2	1/2	1/4	1/4
1/4	1/4	1/2	1/2	1/4	1/4
1/2	1/2	1	1	3/4	3/4
1/2	1/2	1	1	3/4	3/4
1/4	1/4	1/2	1/2	1/4	1/4
1/4	1/4	1/2	1/2	1/4	1/4

X

(c) $A_R(X)$

Figure 1: X, $A(X)$, and $A_R(X)$.

Definition 7. Given a decision system $S = (U, C \cup D)$, the dependency degree of D on C is defined by

$$\gamma(C, D) = \frac{|\mathrm{POS}_C(D)|}{|U|}, \tag{6}$$

where $\mathrm{POS}_C(D) = \cup_{X \in U/D} \underline{C}(X)$ is the positive region of D with respect to C and $|*|$ denotes the cardinality of $*$.

Definition 8. Given a decision system $S = (U, C \cup D)$ and $B \subseteq C$, B is independent if

$$\gamma(B - a, D) < \gamma(B, D), \quad \forall a \in B. \tag{7}$$

Attribute reduction in decision systems is defined as follows.

Definition 9. Given a decision system $S = (U, C \cup D)$ and $B \subseteq C$, B is called a reduct if

(1) $\gamma(B, D) = \gamma(C, D)$,

(2) $\gamma(B - \{a\}, D) < \gamma(B, D)$ for any $a \in B$.

2.2. A Novel Uncertainty Measure of Rough Sets. Given an information system $S = (U, \mathrm{AT})$ and $X \subseteq U$, the characteristic function of X on U can be denoted by

$$A(x) = \begin{cases} 1, & \text{if } x \in X, \\ 0, & \text{if } x \notin X, \end{cases} \tag{8}$$

where $x \in U$.

Let $A(X) = \{A(x)/x \mid x \in U\}$, then $A(X)$ can be considered as a special fuzzy set derived from X on U.

In rough set theory, objects in the same equivalent class cannot be distinguished for each other, since they have the same characteristic. However, in the boundary region of a rough set, objects in the same class have different characteristics. In this case, their mean value of objects in a class is generally used to characterize each object.

Definition 10. Given an information system $S = (U, \mathrm{AT})$, $X \subseteq U$, $P \subseteq \mathrm{AT}$, and $x \in U$, the mean value of x in X, denoted by $A_P(x)$, is defined by

$$A_P(x) = \frac{|[x]_P \cap X|}{|[x]_P|}. \tag{9}$$

We denote $\{A_R(x)/x \mid x \in U\}$ by $A_P(X)$. It is evident that $A_P(X)$ is a fuzzy set on U. As an example, given $U = \{x_1, x_2, \ldots, x_{36}\}$, $U/P = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6, x_7, x_8\}, \ldots, \{x_{33}, x_{34}, x_{35}, x_{36}\}\}$, and $X = \{u_8, \ldots, x_{11}, u_{14}, \ldots, x_{17}, u_{20}, \ldots, x_{24}, u_{26}, \ldots, x_{29}\}$, seen in Figure 1(a), then $A(x)$ and $A_P(x)$ are calculated by (8) and (9), respectively, as shown in Figures 1(b) and 1(c).

As mentioned above, when an object x is not in X, its mean value is non-zero if and only if its equivalent class has non-empty intersection with X; when x is in X, its mean value is 1 if and only if its its equivalent class is contained in X. From Definition 10, it is easy to verify that the mean value $A_P(x)$ is an inclusion degree $D(X/[x]_P)$ of $[x]_P$ being included in X.

Proposition 11. *Given an information system $S = (U, \mathrm{AT})$, $X \subseteq U$, $P \subseteq AT$, and $x \in U$, the following conclusions hold:*

(1) *if $x \in \underline{P}(X)$, then $A_P(x) = 1$;*

(2) *if $x \notin \overline{P}(X)$, then $A_P(x) = 0$;*

(3) *if $x \in \overline{P}(X) - \underline{P}(X)$, then $0 < A_P(x) < 1$.*

Note that $A_P(x) = A(x)$ when x is in the positive region and the negative region. It is obvious that $A_P(x) \neq A(x)$ only when x is in the boundary region.

Definition 12. Given an information system $S = (U, \mathrm{AT})$, $P \subseteq \mathrm{AT}$, and $X \subseteq U$, the mean-variance uncertainty measure (MVM) of X with respective to P, denoted by $\sigma_P(X)$, is defined as

$$\sigma_P(X) = \sqrt{\frac{\sum_{x \in U} (A(x) - A_P(x))^2}{|U|}}. \tag{10}$$

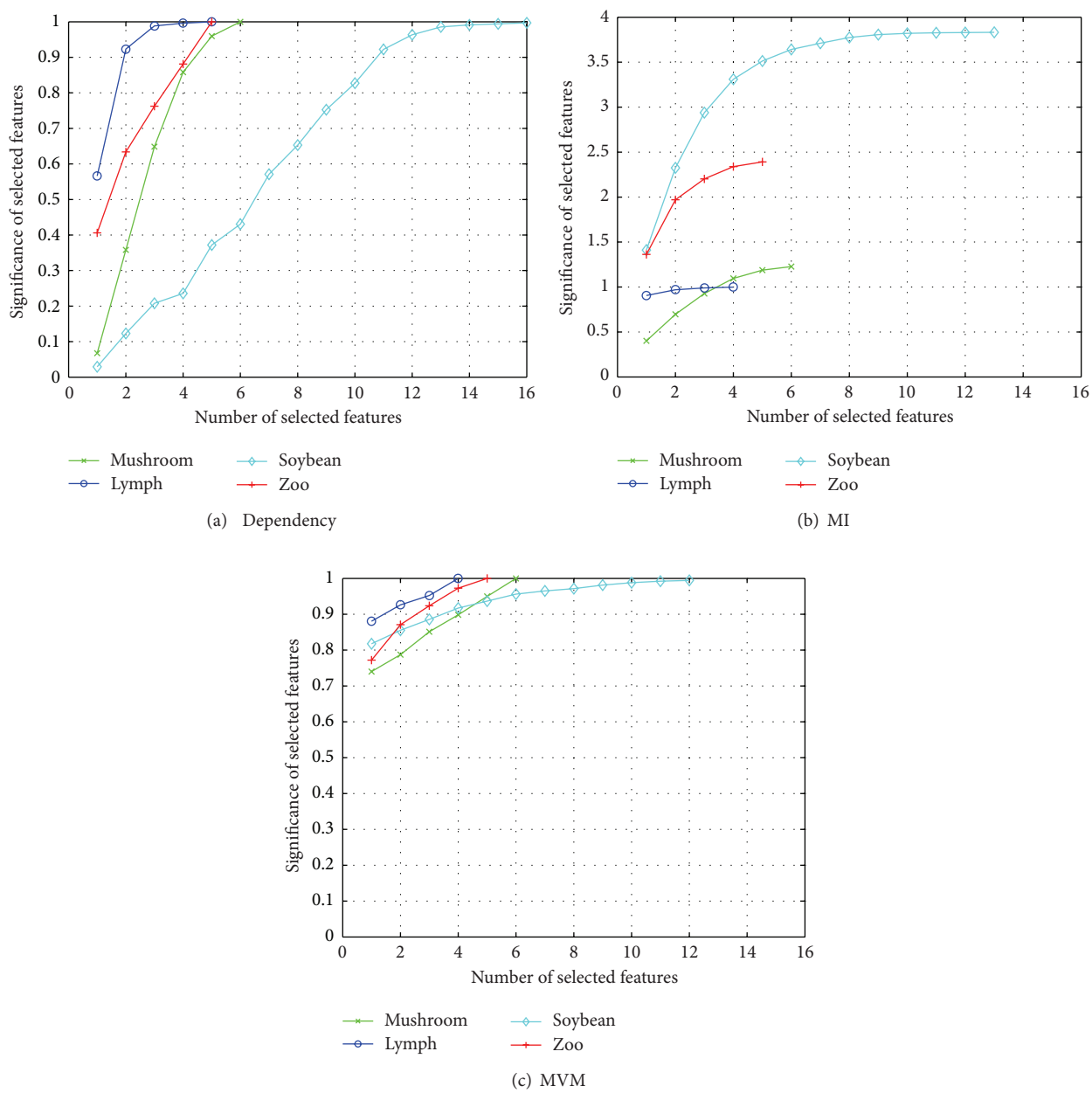

Figure 2: Comparison of convergence among three methods.

It is clear that

$$\sigma_P(X) = \sqrt{\frac{\sum_{x\in X}(1 - A_P(x))^2 + \sum_{x\in U-X}(A_P(x))^2}{|U|}}. \tag{11}$$

Assume $\sigma_P(X) = 0$ when $X = \emptyset$ or $P = \emptyset$. From Definition 12 one can see that only objects in the boundary region of X contribute to the value of $\sigma_P(X)$. In this sense, $\sigma_P(X)$ takes fully information in the boundary region into account. Therefore, it is a proper measure to evaluate the uncertainty of X.

Definition 13. Given an information system $S = (U, \mathrm{AT})$, $P, Q \subseteq \mathrm{AT}$, and $X, Y \subseteq U$,

(1) X is said to be σ-definable if $\sigma_P(X) = 0$;

(2) X is said to be σ-rough if $\sigma_P(X) \neq 0$;

(3) X is said to be coarser with respect to P than Y with respect to Q if $\sigma_P(X) < \sigma_Q(Y)$, in which case, Y is called finer with respect to Q than X with respect to P.

Next, we investigate properties of $\sigma_P(X)$ and show its efficiencies in evaluating uncertainty of a set in information systems.

Proposition 14. *Given an information system $S = (U, \mathrm{AT})$, $P \subseteq AT$, and $X \subseteq U$, the following conclusions hold:*

(1) $\sigma_P(X) = 0 \Leftrightarrow \rho_P(X) = 0$;

(2) *if* $U/P = \omega$, *then* $\sigma_P(X) = 0$, *where* $\omega = \{\{x\} \mid x \in U\}$ *is the finest partition of* U;

(3) *if* $U/P = \delta$, *then* $\sigma_P(X) = \sqrt{|X|/|U| - (|X|/|U|)^2}$, *where* $\delta = \{U\}$ *is the coarsest partition of* U.

3. Feature Selection in Decision Systems

In this section, the proposed uncertainty measure is further investigated to perform feature selection in decision systems.

Definition 15. Given a decision system $S = (U, C \cup D)$ and $B \subseteq C$, MVM of the decision attribute set D with respect to the conditional attribute subset B, called D-MVM, or an evaluation function, is defined by

$$\sigma(B, D) = \frac{1}{N}\left(\sigma_B(D_1) + \sigma_B(D_2) + \cdots + \sigma_B(D_N)\right), \quad (12)$$

where N is the number of the decision classes induced by the decision attribute set D, $\sigma_B(D_i)$, $i = 1, 2, \ldots, N$, reflect the uncertainty measure of each decision class, and $\sigma(B, D)$ describes the integrated uncertainty degree of blocks $D_1, D_2, \ldots, D_N$.

In the following, some properties of $\sigma(B, D)$ are studied.

Proposition 16. *Given a decision system $S = (U, C \cup D)$ and $B \subseteq C$, the following conclusions hold:*

(1) $\sigma(B, D) = 0 \Leftrightarrow \rho(B, D) = 0$;

(2) *if* $U/B = \omega$, *then* $\sigma(B, D) = 0$;

(3) *if* $U/B = \delta$, *then* $\sigma(B, D) = (1/N)\sum_{i=1}^{N}\sqrt{|D_i|/|U| - (|D_i|/|U|)^2}$.

Proof. The proof is analogous to that of ([11], property 2.14). □

Definition 17. Given a decision system $S = (U, C \cup D)$ and $B \subseteq C$, B is independent if

$$\sigma(B, D) < \sigma(B - a, D), \quad \forall a \in B. \quad (13)$$

By D-MVM, a relative reduct can be defined as follows.

Definition 18. Given a decision system $S = (U, C \cup D)$ and $B \subseteq C$, B is a relative reduct of C with respective to D if and only if

(1) $\sigma(B, D) = \sigma(C, D)$,

(2) $\sigma(B, D) < \sigma(B - a, D)$ for any $a \in B$.

A relative reduct is a minimal subset which has the same discriminating power as the raw decision systems.

Table 1: Data sets.

	Data sets	Abbreviation	Samples	Features	Class
1	Lymphography	Lymph	148	19	4
2	Mushroom	Mush	8124	23	2
3	Soybean	Soybean	683	36	19
4	Zoo	Zoo	101	17	7

Definition 19 (significance based on D-MVM). Given a decision system $S = (U, C \cup D)$, $B \subseteq C$, and a feature $a \in C - B$, the significance of a is defined as

$$\mathrm{Sig}_\sigma(a, B, D) = \sigma(B \cup a, D) - \sigma(B, D). \quad (14)$$

Notice that if B is an empty set, $\sigma(B, D) = 0$, and $\mathrm{Sig}_\sigma(a, B, D)$ is a nonnegative real number; otherwise, $\mathrm{Sig}_\sigma(a, B, D) \leq 0$.

With the proposed evaluation function, a forward greedy search algorithm for feature selection can be designed as follows.

In the first iteration, we start with an empty set specified with $\sigma(B, D) = 0$. The quantity $\mathrm{Sig}_\sigma(a, B, D)$ is negative in every iteration except the first one. The rest features in each iteration are all evaluated, and the one with the minimal significance will be chosen. The algorithm does not stop until adding any of the rest features to selected feature set will not bring a change larger than threshold ε in Algorithm 1, where ε controls the precision of the algorithm.

There is no doubt that FGSA-MVM is for the sake of searching a subset of conditional attributes with minimal positive real D-MVM. We obverse step 7 of FGSA-MVM. In the first iteration, we choose the minimal significance because $\mathrm{Sig}_\sigma(a, \emptyset, D)$ is a positive number. In the rest iterations, we also select the minimal significance with the biggest step length since $\mathrm{Sig}_\sigma(a, B, D)$ is nonpositive for any $B \neq \emptyset$.

4. Experiments and Analysis

In order to test the validity of the proposed method for feature selection, comparative experiments have been implemented in efficiency and convergence of proposed algorithm with two of the most important methods, feature selection based on dependence [39] and mutual information [40].

As shown in Table 1, four standard data sets, cited from the machine learning data repository, University of California, Irvine, CA, USA [41], are employed in our experiments.

CART and RBF-support vector machine (SVM) learning algorithms are introduced to test the classification performances of feature selection for raw sets and for selected feature sets. As a widely used technique to evaluate classification performances in machine learning, 10-fold cross-validation [42] is carried out in our experiments by dividing the samples into 10 subsets. Nine of them are used as training set, and the rest one is used as the test set. After 10 rounds, the average value and variation are computed as the final classification performance.

Classification performances are evaluated by CART in Table 2 and by RBF-SVM in Table 3. "Hold" marks the highest classification performances among these obtained

Forward Greedy Search Algorithm of Feature Selection based on Mean-Variable in Decision Systems (FGSA-MVM):
Input: $(U, C \cup D, V, f), \varepsilon$
Output: red
(1) $\emptyset \rightarrow red$
(2) while $C - red \neq \emptyset$
(3) for each $a_i \in C - red$
(4) compute $\mathrm{Sig}_\sigma(a_i, B, D) = \sigma(B \cup a_i, D) - \sigma(B, D)$
(5) end for
(6) select the attribute a_k such that
(7) $\mathrm{Sig}_\sigma(a_k, red, D) = \min_i \mathrm{Sig}_\sigma(a_i, red, D)$
(8) if $|\mathrm{Sig}_\sigma(a_k, red, D)| < \varepsilon$
(9) $red \cup a_k \rightarrow red$
(10) else
(11) break
(12) end if
(13) end while
(14) return red

Algorithm 1

Table 2: Comparison of classification performance of reducts based on different uncertainty measures with CART.

Data sets	Raw data	Dependency	MI	MVM
Lymph	0.6994 ± 0.2195	**0.6825 ± 0.1822**	**0.6825 ± 0.1822**	**0.6825 ± 0.1822**
Mush	0.9637 ± 0.0990	0.9637 ± 0.0990	**0.9685 ± 0.0996**	**0.9685 ± 0.0996**
Soybean	0.9174 ± 0.0507	0.8485 ± 0.0700	0.8780 ± 0.0629	**0.9192 ± 0.0490**
Zoo	0.9065 ± 0.0913	0.8329 ± 0.0676	**0.9276 ± 0.0987**	**0.9276 ± 0.0987**
Aver.	0.8567	0.8439	0.8672	0.8676

Table 3: Comparison of classification performance of reducts based on different uncertainty measures with SVM.

Data sets	Raw data	Dependency	MI	MVM
Lymph	0.5623 ± 0.0583	**0.8448 ± 0.0940**	**0.8448 ± 0.0940**	**0.8448 ± 0.0940**
Mush	0.9587 ± 0.0984	**0.9587 ± 0.0984**	**0.9587 ± 0.0984**	**0.9587 ± 0.0984**
Soybean	0.5445 ± 0.0649	0.6041 ± 0.0723	0.6239 ± 0.0683	**0.6566 ± 0.0817**
Zoo	0.8615 ± 0.0901	**0.9239 ± 0.0924**	**0.9239 ± 0.0924**	**0.9239 ± 0.0924**
Aver.	0.7181	0.8435	0.8523	0.8549

by the methods based on three uncertainty measures. The number of selected features with the highest classification performance by the new measure is larger than that by dependency and by MI. It is 12, 4, and 12, respectively, via CART algorithm, whereas it is 14, 8, and 11 via SVM algorithm.

From the experiments one can see that the proposed measure outperforms not only in the smallest average number of selected features in reducts but also in the highest classification performance in feature selection.

In the remainder of this section, we pay attention to the convergence of the proposed method. Figure 2 shows the fluctuations of evaluation functions with respect to the number of selected features. The significance of selected features is calculated based on dependency, on MI, and on MVM, respectively. The four data sets are used to show the convergence of different techniques. The selected orders of the four data sets based on different evaluation functions are shown in Table 4, in which the sequences of selected features are different, even the number of selected features in the optimal reducts may be the same. As a whole, significance degrees based on dependency and MI increase, while significance based on MVM decreases. With MVM, all four evaluation functions decrease fast at the beginning of the selection process. The evaluation function of credit data slowly decreases, and this result constitutes a different pattern of behavior compared with the three other data sets. Feature selection algorithms may stop very early if we specify a threshold to stop the search in this case. The convergence and good classification performances are observed in the results.

Table 4: Comparison of selected features by different uncertainty measures.

Data sets	Dependency	Mutual information	MVM
Lymph	18 2 13 14 15 16	13 18 14 2 15 16	13 18 14 2 15 16
Mush	5 20 8 12 3	5 20 22 21	5 20 22 21
Soybean	18 26 11 12 35 29 22 1	29 15 22 1 18 3 7 6	28 26 22 15 1 18 3 7
	3 6 7 10 4 5 9 16	10 4 9 16 5	10 6 14 4
Zoo	4 13 12 6 8	13 4 6 8 3	13 4 6 8 3

5. Conclusion

This contribution studied feature selection based on MVM in decision information systems, which is one of the most important applications of rough set theory. A novel approach to feature selection was proposed by introducing an evaluation function based on MVM. Theoretical analysis and experimental results concluded that the performances of proposed method are outperformed by dependency and by MI not only in the number of selected features but also in the classification precision.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 61273012 and 61102040, and the Shandong Province College Science and Technology Plan Project under Grant J12LN91, and the Shandong Natural Science Foundation of China under Grant ZR2011FL014, and the Shandong Province Science and Technology Development Plan Projects under Grant 2012YD01052.

References

[1] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[2] P. Fortemps, S. Greco, and R. Słowiński, "Multicriteria decision support using rules that represent rough-graded preference relations," *European Journal of Operational Research*, vol. 188, no. 1, pp. 206–223, 2008.

[3] T. Deng, Y. Chen, W. Xu, and Q. Dai, "A novel approach to fuzzy rough sets based on a fuzzy covering," *Information Sciences*, vol. 177, no. 11, pp. 2308–2326, 2007.

[4] W. Z. Wu, "Attribute reduction based on evidence theory in incomplete decision systems," *Information Sciences*, vol. 178, no. 5, pp. 1355–1371, 2008.

[5] M. C. Lee, "Using support vector machine with a hybrid feature selection method to the stock trend prediction," *Expert Systems with Applications*, vol. 36, no. 8, pp. 10896–10904, 2009.

[6] Z. Pawlak, "Some remarks on conflict analysis," *European Journal of Operational Research*, vol. 166, no. 3, pp. 649–654, 2005.

[7] H. Xiumei, F. Haiyan, and S. Kaiquan, "S-rough sets and the discovery of F-hiding knowledge," *Journal of Systems Engineering and Electronics*, vol. 19, no. 6, pp. 1171–1177, 2008.

[8] C. Wang, C. Wu, D. Chen, Q. Hu, and C. Wu, "Communicating between information systems," *Information Sciences*, vol. 178, no. 16, pp. 3228–3239, 2008.

[9] D. Chen, Y. Yang, and X. Zhang, "Parameterized local reduction of decision systems," *Journal of Applied Mathematics*, vol. 2012, Article ID 857590, 14 pages, 2012.

[10] Md. M. Kabir, Md. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," *Neurocomputing*, vol. 74, no. 17, pp. 2914–2928, 2011.

[11] C. Yang, W. Zhang, J. Zou et al., "A novel uncertainty measure on rough sets: a mean-variance approach," in *Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 900–904, 2012.

[12] L. Yu and H. Liu, "Efficiently handling feature redundancy in highdimensional data," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03)*, pp. 685–690, Washington, DC, USA, August 2003.

[13] S. K. M. Wong and W. Ziarko, "On optimal decision rules in decision tables," *Bulletin of the Polish Academy of Sciences*, vol. 33, no. 11-12, pp. 693–696, 1985.

[14] F. Li and Y. Yin, "Approaches to knowledge reduction of covering decision systems based on information theory," *Information Sciences*, vol. 179, no. 11, pp. 1694–1704, 2009.

[15] J. Guan and D. Bell, "Rough computational methods for information systems," *Artificial Intelligence*, vol. 105, pp. 77–103, 1998.

[16] Z. Geng and Q. Zhu, "Rough set-based heuristic hybrid recognizer and its application in fault diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2711–2718, 2009.

[17] Z. Pawlak, "Rough sets and fuzzy sets," *Fuzzy Sets and Systems in Information Science and Engineering*, vol. 17, no. 1, pp. 99–102, 1985.

[18] X. Hu and N. Cercone, "Learning in relational databases: a rough set approach," *Computational Intelligence*, vol. 11, no. 2, pp. 323–338, 1995.

[19] Q. Shen and R. Jensen, "Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring," *Pattern Recognition*, vol. 39, no. 7, pp. 1351–1363, 2004.

[20] Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognition*, vol. 40, no. 12, pp. 3509–3521, 2007.

[21] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, 2009.

[22] Q. He, C. Wu, D. Chen, and S. Zhao, "Fuzzy rough set based attribute reduction for information systems with fuzzy decisions," *Knowledge-Based Systems*, vol. 24, no. 5, pp. 689–696, 2011.

[23] D. Zhang, Y. Wang, and H. Huang, "Rough neural network modeling based on fuzzy rough model and its application to texture classification," *Neurocomputing*, vol. 72, no. 10–12, pp. 2433–2443, 2009.

[24] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[25] J. Zheng and R. Yan, "Attribute reduction based on cross entropy in rough set theory," *Journal of Information & Computational Science*, vol. 9, no. 3, pp. 745–750, 2012.

[26] N. M. Parthaláin and Q. Shen, "Exploring the boundary region of tolerance rough sets for feature selection," *Pattern Recognition*, vol. 42, no. 5, pp. 655–667, 2009.

[27] X. Jia, W. Liao, Z. Tang, and L. Shang, "Minimum cost attribute reduction in decision-theoretic rough set models," *Information Sciences*, vol. 219, pp. 151–167, 2013.

[28] F. Min, H. He, Y. Qian, and W. Zhu, "Test-cost-sensitive attribute reduction," *Information Sciences*, vol. 181, no. 22, pp. 4928–44942, 2011.

[29] F. Wang, J. Liang, and C. Dang, "Attribute reduction for dynamic data sets," *Applied Soft Computing*, vol. 13, no. 1, pp. 676–6689, 2013.

[30] J. Zhang, T. Li, D. Ruan, and D. Liu, "Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems," *International Journal of Approximate Reasoning*, vol. 53, no. 4, pp. 620–635, 2012.

[31] Z. Meng and Z. Shi, "Extended rough set-based attribute reduction in inconsistent incomplete decision systems," *Information Sciences*, vol. 204, pp. 44–69, 2012.

[32] W. Zhu and F. Y. Wang, "Reduction and axiomization of covering generalized rough sets," *Information Sciences*, vol. 152, pp. 217–230, 2003.

[33] E. C. C. Tsang, C. Degang, and D. S. Yeung, "Approximations and reducts with covering generalized rough sets," *Computers & Mathematics with Applications*, vol. 56, no. 1, pp. 279–289, 2008.

[34] J. S. Mi, W. Z. Wu, and W. X. Zhang, "Approaches to knowledge reduction based on variable precision rough set model," *Information Sciences*, vol. 159, no. 3-4, pp. 255–272, 2004.

[35] X. T. Deng, Z. F. Li, and S. Y. Wang, "A minimax portfolio selection strategy with equilibrium," *European Journal of Operational Research*, vol. 166, no. 1, pp. 278–292, 2005.

[36] X. Huang, "Mean-variance model for fuzzy capital budgeting," *Computers and Industrial Engineering*, vol. 55, no. 1, pp. 34–47, 2008.

[37] P. Gupta, M. K. Mehlawat, and A. Saxena, "Asset portfolio optimization using fuzzy mathematical programming," *Information Sciences*, vol. 178, no. 6, pp. 1734–1755, 2008.

[38] X. Y. Zhou and G. Yin, "Markowitz's mean-variance portfolio selection with regime switching: a continuous-time model," *SIAM Journal on Control and Optimization*, vol. 42, no. 4, pp. 1466–1482, 2003.

[39] M. Modrzejewski, "Feature selection using rough sets theory," in *Proceedings of the European Conference on Machine Learning*, pp. 213–226, Vienna, Austria, 1993.

[40] F. F. Xu, D. Q. Miao, and L. Wei, "Fuzzy-rough attribute reduction via mutual information with an application to cancer classification," *Computers and Mathematics with Applications*, vol. 57, no. 6, pp. 1010–1017, 2009.

[41] D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz, "UCI Repository of Machine Learning Databases," University of California, Department of Information and Computer Science, 1998, http://mlearn.ics.uci.edu/MLRepository.html.

[42] M. van der Gaag, T. Hoffman, M. Remijsen et al., "The five-factor model of the positive and negative syndrome scale II: a ten-fold cross-validation of a revised model," *Schizophrenia Research*, vol. 85, no. 1–3, pp. 280–287, 2006.