Discussion

Jerzy W. Grzymala-Busse

University of Kansas, Lawrence, KS 66045, USA

and

Institute of Computer Science, Polish Academy of Sciences, 01–237 Warsaw, Poland

Arbitrary Binary Relations

Let R be an arbitrary binary relation, $R \subseteq U \times U$.

Let x be a member of U.

The R-*successor* set of x, denoted by $R_s(x)$, is defined as follows

$$R_s(x) = \{y \mid xRy\}.$$

Let X be a subset of U.

Singleton Parameterized Approximations

The singleton parameterized approximation of X with the threshold α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{singleton}(X)$, is defined as follows

$$\{x \mid x \in U,\ Pr(X|R_s(x)) \geq \alpha\},$$

where $Pr(X|R_s(x)) = \frac{|X \cap R_s(x)|}{|R_s(x)|}$ is the conditional probability of X given $R_s(x)$.

Subset Parameterized Approximations

A subset parameterized approximation of the set X with the threshold α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{subset}(X)$, is defined as follows

$$\cup\{R_s(x) \mid x \in U,\ Pr(X|R_s(x)) \geq \alpha\}.$$

Concept Parameterized Approximations

A concept parameterized approximation of the set X with the threshold α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{concept}(X)$, is defined as follows

$$\cup\{R_s(x) \mid x \in X,\ Pr(X|R_s(x)) \geq \alpha\}.$$

Incomplete Data Sets

Two types of missing attribute values:

lost, denoted by "?", (e.g., the value was erased)

and

"do not care" conditions, denoted by "*", (such a value may be any value of the attribute).

An Incomplete Data Set

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	normal	no	*	no
2	?	no	no	no
3	normal	*	no	yes
4	normal	no	?	no
5	high	yes	*	yes
6	high	yes	yes	no
7	high	?	yes	yes
8	high	yes	yes	yes

Characteristic Relation

$$K(1) = K(4) = \{1, 3, 4\},$$

$$K(2) = \{1, 2, 3\},$$

$$K(3) = \{1, 3\},$$

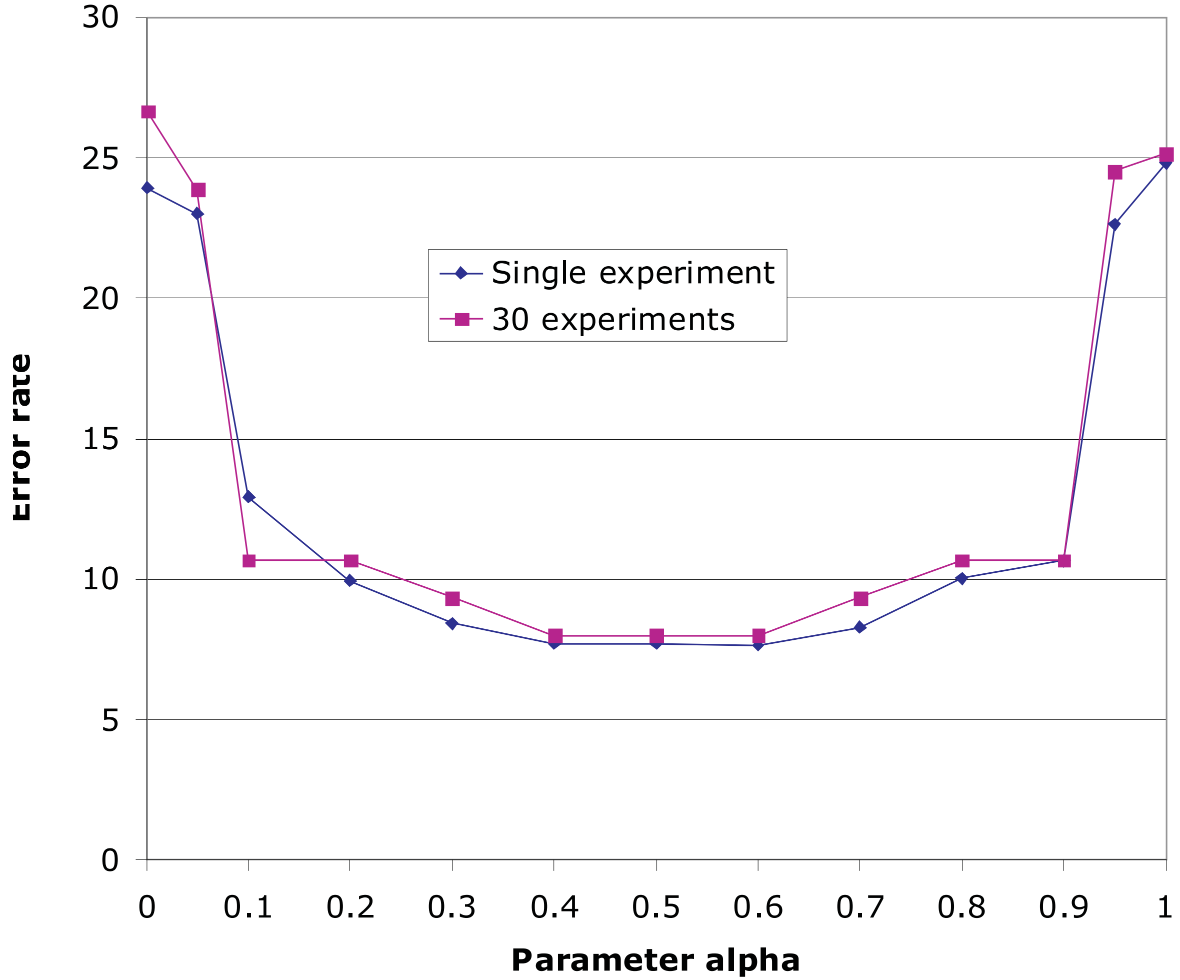
$$K(5) = K(6) = K(8) = \{5, 6, 8\},$$

$$K(7) = \{5, 6, 7, 8\}.$$

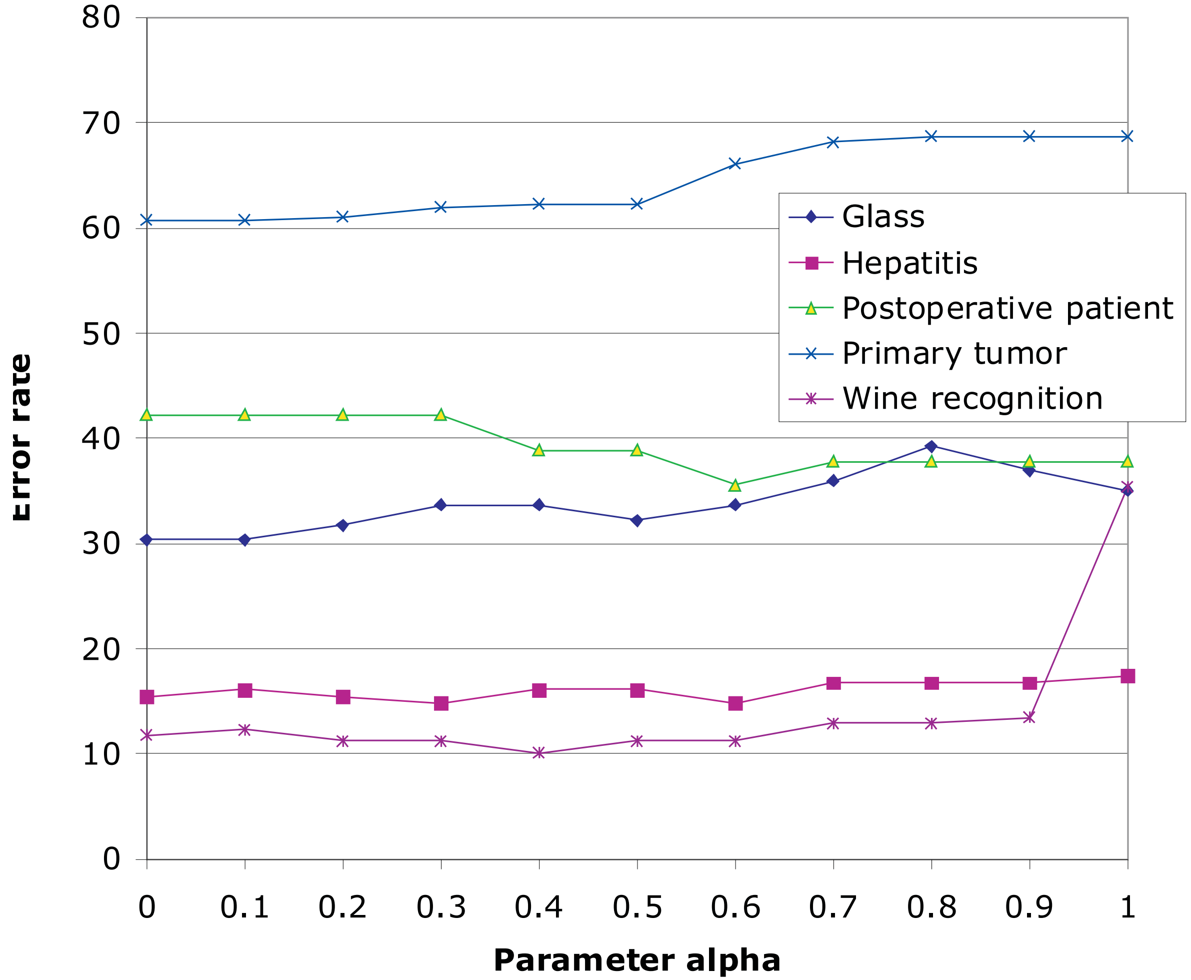
R = {1, 1), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 3), (4, 1), (4, 3), (4, 4), (5, 5), (5, 6), (5, 8), (6, 5), (6, 6), (6, 8), (7, 5), (7, 6), (7, 7), (7, 8), (8, 5), (8, 6), (8, 8)}

R is not symmetric (since (5, 7)) and not transitive (since (2, 1) and (1, 4)).

Iris Data Set



Five Other Data Sets



Iris Data Set (56% consistent)

Concept	Number of distinct parameterized approximations
Class, Iris-setosa	2
Class, Iris-versicolor	6
Class, Iris-viginica	5