Information Systems and Soft Sets

Zofia Machnicka and Marek Palasinski

University of Information Technology and Management
Rzeszow, Poland
zmachnicka@wsiz.rzeszow.pl
mpalasinski@wsiz.rzeszow.pl

Abstract. It will be shown that each information system can be considered a soft set and each *finite* soft set can be considered an information system.

1 Basic Information

The notion of an information system is well established in the literature, e.g. [1, 4]. Informally, an information system consists of a finite nonempty set of *objects* and a finite nonempty set of *attributes*. Each attribute a assigns to each object some *value* $a(x)$, which is a members of a specific finite set V_a, called the *range* of the attribute a. Thus an attribute can be thought of as a function $a : V \to V_a$, or as a sequence of elements of V_a. Such a tuple is represented as a column in a matrix (or table) representing the information system. The rows of the matrix are labelled by objects and the columns by attributes. The matrix dimention is $n \times p$, where n is the number of objects and p the number of attribues, see Table 1 below.

Table 1. Information system

	a_0	$\cdots$	a_k	$\cdots$	a_p
x_0	a_{00}	$\cdots$	a_{k0}	$\cdots$	a_{p0}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	a_{0i}	$\cdots$	a_{ki}	$\cdots$	a_{pi}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	a_{0n}	$\cdots$	a_{kn}	$\cdots$	a_{pn}

Here $V = \{x_0, \ldots, x_n\}$, $A = \{a_0, , a_p\}$, and a_{ki} is the value of the attribute a_k on x_i, i.e., $a_{ki} = a_k(x_i)$. Thus for each $a \in A$, we have that $a \in V_a^V$ and $A \subseteq \bigcup\{V_a^V : a \in A\}$, where the union is disjoint. Similarly, as each object $x \in V$ labels a row in the matrix, it can be identified with a tuple of values of attributes from A on this object. Hence the set of objects V can be identified with a subset of the Cartesian product of sets V_a (see [4])

$$V \subseteq \prod_{a \in A} V_a.$$

Definition 1. *An information system is a pair $S = \langle V, A \rangle$, where V and A are nonempty finite sets, such that for each $a \in A$ there exists a finite set V_a such that $a : V \to V_a$.*

The elements of V are called objects of the system S and the elements of A attributes of the system. An element $x \in V$ is identified with a tuple

$$x = \langle a(x) : a \in A \rangle.$$

An information system S is called *two valued* iff for every $a \in A$, the set V_a has two elements, which we denote by 0 and 1. Expressions of the form

$$a_{i_1} = b_{i_1} \wedge \cdots \wedge a_{i_k} = b_{i_k} \longrightarrow a_{i_t} = b_{i_t}$$

or

$$a_{i_1} = b_{i_1} \wedge \cdots \wedge a_{i_k} = b_{i_k} \longrightarrow a_{i_t} \neq b_{i_t}$$

where $a_{i_1} \dots a_{i_k}, a_{i_t}$ are attributes and $b_{i_1}, \dots, b_{i_k}, b_{i_t}$ are their possible values, i.e., $b_{i_j} \in V_{i_j}$, are called *rules*. A rule of the first form is called a *deterministic association rule* while the one of the second form is called an *inhibitory association rule*. Mathematicaly, any rule is an implication and a *true* rule is a rule that is true as an implication for any object from V. A rule is *realizable* if its predecessor is true for at least one object from V.It was shown in [2] that every information system can be equivalently replaced by a two valued information system. We recall the procedure in Section3.

2 Soft Sets

The notion of a *soft set* has been proposed in [3]as a mathematical approach to uncertainty, alternative to that of a fuzzy set. It has subsequently provoked a lot of research. Let U be a set called a *universe* and let E be another set, disjoint with U, called the set of *parameters.*

Definition 2. *(Following [3]) Let U be a set. A pair $\langle F, E \rangle$ is called a soft set over U if and only if F is a mapping from E into the set of all subsets of the set U.*

A soft set then can be seen as a parametrized family $\{F(\varepsilon) : \varepsilon \in E\}$ of subsets of the set U and the elements of E are called parameters. In the terminology of [3], for each $\varepsilon \in E$, the set $F(\varepsilon)$ is called an *ε-approximation* of the soft set. If both U and E are finite then we will call the soft set $\langle F, E \rangle$ *finite*. In the next section we show that there is a natural connection between soft sets and information systems.

3 Connection Between Soft Sets and Information Systems

The following procedure of getting a two valued system $S^{(2)}$ from a given information system S was presented at the HSI conference in 2010 ([2]) and we recall it here for completeness. Let $S = \langle V, A \rangle$, where $A = (a_1, \dots, a_n)$ and for each $i = 1, \dots, n$

let the set of values of the attribute a_i be $\{0, \ldots, k_i - 1\}$. We define the associated two-valued system $S^{(2)}$ by setting its attributes set to be

$$\{a_{10}, \ldots, a_{1k_i-1}, \ldots, a_{n0}, \ldots, a_{nk_n-1}\},$$

each one assuming one of the two values $0, 1$. The set of objects remains the same. For any rule r for S we define the corresponding rule $r^{(2)}$ for $S^{(2)}$, replacing each expression of the form $a_i = j$ by $a_{ij} = 1$ and $a_{ij} = 0$ in other case and in case the rule is an inhibitory one expression $a_i \neq j$ by $a_{ij} = 0$. We then have the following lemma.

Lemma 1. *For any rule r, if r is true and realizable in S then $r^{(2)}$ is true and realizable in $S^{(2)}$.*

The proof of the lemma is straightforward, by contradiction. To show the converse, one needs to define new rules (each rule true and realizable in S^2, using the inverse translation, leads to a rule of a new kind).

Remark 1. In case of binary information systems, if two systems S_1 and S_2 are different then the sets of true and realizable implications associated with S_1 and S_2, respectively, are different (see [1, 2]).

From now on let $S^{(2)}$ denote a binary information system. So $S^{(2)}$ can be presented as in Table 1 above, where the value a_{ki} of the attribute a_k on x_i is 0 or 1.

It is now easy to see that each information system $\langle V, A\rangle$ can be considered a soft set. It is enough to take into account that any subset $W \subseteq$V can be uniquely identified with its characteristic function $f : V \to \{0, 1\}$ such that

$$f(x) = \begin{cases} 1, & \text{when } x \in W \\ 0, & \text{otherwise.} \end{cases}$$

For $k = 1, \ldots, p$, the k th column in the table above can be considered a function $a_k : V \to \{0, 1\}$, with $a_k(x_i) = a_{ki}$, for each $j = 1, \ldots, n$, so each column determines a subset of V. Take the function $F : A \to \mathcal{P}(V)$ such that $F(a_k)$ is the subset of V determined by the column in the table corresponding to the attribute $a_k \in A$. Then $\langle F, A\rangle$ is a soft set over U.

On the other hand, consider a finite soft set $\langle F, E\rangle$ over a universe V. Assume that $V = \{x_0, \ldots, x_n\}$ and $E = \{\varepsilon_0, \ldots, \varepsilon_p\}$, for some n and p. To show that this soft set determines the unique information system, take $A = \{F(\varepsilon) : \varepsilon \in E\}$. For each $k = 0, \ldots, p$ and $i = 0, \ldots, n$ let a_k be the characteristic function of the set $F(\varepsilon_i)$. Then the values of a_k form the k-th column in table 1. So the soft set $\langle F, E\rangle$ over V is identified with the information system $\langle V, A\rangle$. Clearly, the two identifications are mutual inverses, so each soft set determines a unique information system and vice versa. This allows us to state the main result of this note.

Theorem 1. *Given a finite set V there is a one-one correspondence between the information systems of the form $\langle V, A\rangle$ and finite soft sets over the universe V.*

As the finite set V in the theorem is arbitrary, the theorem can be restated as

Corollary 1. *Every information system can be regarded as a finite soft set and every finite soft set can be regarded as an information system.*

Remark 2. As the anonymous referee points out, the main result can be alternatively explained in more general terms as follows. Given a family A of zero-one sequences of a common length n, one can treat this family as the set of parameters E. Let U be any set of n elements. Then each parameter $\varepsilon \in E$, being a sequence of zeros and ones is a characteristic function of some subset of U, call this subset $F(\varepsilon)$. Then $F : E \to \mathcal{P}(U)$ and the pair $\langle F, E \rangle$ is a soft set. The original set of sequences A can be recovered from this soft set as the family of characteristic sequences of sets $F(\varepsilon)$, for all $\varepsilon \in E$.

References

1. P.Delimatea, M.Moshkow, A.Skowron, Z..Suraj, *Inhibitory Rules in Data Analysis. A Rough Set Approach,* Studies in Computational Inteligence, Vol. 163, Springer-Verlag, Berlin (2008)
2. B. Fryc, Z. Machnicka, M. PaÅĆasiÅĎski, *Remarks on Two Valued Information Systems*, in: T. Pardela, B. Wilamowski (eds.) RzeszÃşw (3rd International Conference on Human System Interaction), pp 775-777 (2010)
3. D. Molodtsov, *Soft set theory - First results*, Comput. Math. Appl., 37, pp 19-31 (1999)
4. K.Pancerz,*Zastosowanie zbiorów przybliżonych do identyfikacji modeli systemów współ bieżnych,* Rozprawa doktorska, IPI PAN, Warszawa (2006)