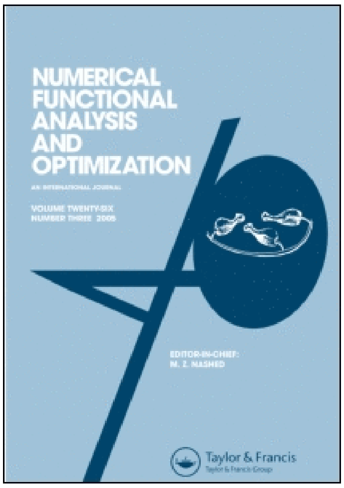

Numerical Functional Analysis and Optimization
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597287

An Automatic User-Recognition Approach Within a Reputation System Using a Nonlinear Hausdorff-Derived Metric

Lenua Alboaie [a]; Tudor Barbu [b]
[a] Department for Computer Sciences of the University "Al. I. Cuza" of Iai, Iai, Romania [b] Institute for Computer Science of the Romanian Academy, Iai Branch, Iai, Romania

Online Publication Date: 01 November 2008

To cite this Article Alboaie, Lenua and Barbu, Tudor(2008)'An Automatic User-Recognition Approach Within a Reputation System Using a Nonlinear Hausdorff-Derived Metric',Numerical Functional Analysis and Optimization,29:11,1240 — 1251

To link to this Article: DOI: 10.1080/01630560802494673

URL: http://dx.doi.org/10.1080/01630560802494673

Numerical Functional Analysis and Optimization, 29(11–12):1240–1251, 2008

ISSN: 0163-0563 print/1532-2467 online
DOI: 10.1080/01630560802494673

AN AUTOMATIC USER-RECOGNITION APPROACH WITHIN A REPUTATION SYSTEM USING A NONLINEAR HAUSDORFF-DERIVED METRIC

Lenuţa Alboaie[1] **and Tudor Barbu**[2]

[1]*Department for Computer Sciences of the University "Al. I. Cuza" of Iaşi, Iaşi, Romania*
[2]*Institute for Computer Science of the Romanian Academy, Iaşi Branch, Iaşi, Romania*

□ *In this paper, we provide an automatic unsupervised recognition technique for Web community user reputations that uses a special nonlinear metric. First we describe the general framework for reputation systems. Then, we propose a feature extraction approach for the reputation system users. The resulting feature vectors (reputations) are clustered with an unsupervised classification algorithm using a nonlinear distance, derived from the Hausdorff metric for sets.*

Keywords Feature vector; Hausdorff-based metric; Pattern recognition; Ratings; Reputation systems; Unsupervised classification.

1. INTRODUCTION

This paper presents an application of the pattern recognition theory and the Hausdorff metrics in the reputation systems domain. The Hausdorff distance has a variety of applications in many important scientific domains. Among the many areas that have found applications of the Hausdorff metric are computer vision, robotics, medicine, astronomy, and graph theory. Transforming the reputation systems into an application area for both the Hausdorff metrics and the pattern recognition represents an important novelty brought by our paper.

At this moment there are many Web systems that gather various information about thousands or millions of people. This information is

Address correspondence to Tudor Barbu, Institute for Computer Science of the Romanian Academy, Iaşi Branch, Iaşi, Romania; E-mail: tudbar@iit.tuiasi.ro

obtained without even interrupting user actions with questions. The user profile is obtained on the basis of this information using different techniques (e.g., machine learning, statistical methods). Actually, the Web systems are developing in a few different directions; one exists due to the information furnished by users (e.g., Wikipedia). The other direction is based on different algorithms that allow obtaining new information that enhances the user experience. An important example in this sense is Google, which uses links to rank Web pages and at the same time collects and manages data obtained from situations when advertisements are clicked [1].

Other examples consist of Web communities that use recommendation systems [2, 3]. In these cases there is collected information like purchasing history and user characteristics, and the system makes proper recommendations based on them (e.g., Amazon, Netflix). Other examples consist of Web systems that use reputation systems [4, 5]. Reputation systems are extremely valuable in those communities where the users have to interact with resources posted by other users or they have to interact with other users (e.g., eBay, YouTube, Slashdot, Flicker). In these situations, using experience of other users is very useful [6].

Generally people have either different or similar profiles, so they are interested in either different or similar resources. We quantize this interest with values that are provided by the user for other users or resources. This interest will have an indirectly computed component, as described in Section 2. We analyze the situation when a user evaluates favorably one or more users. If these users had evaluated favorably a given resource, even if the user does not evaluate directly that resource, we will consider it an implicit favorable evaluation. Thus, the user has the chance to access resources more relevant for him.

In our system there is no absolute value of good or bad resource characteristic. A resource can be good for a set of users but not useful for other set of users. Also, we establish a set of constructions taken into account by the evaluation mechanism. Whenever new users become community members, they can interact with the users corresponding with their preferences. Also, they will be able to access much faster the proper resource set. This represents the general direction our system is based on.

Our goal is to perform a pattern recognition process within a Web community. It consists in a user classification using the user reputations as feature vectors. In the next section, we describe a formal reputation model. Then, a feature extraction approach, consisting in user reputation computing, is performed in Section 3. A special nonlinear Hausdorff-based metric is proposed in the fourth section [9–11]. The unsupervised automatic classification procedure that uses this new distance is described in the fifth section [12]. The paper ends with a conclusions section.

2. THE FORMAL REPUTATION MODEL

In a previous paper [7], we proposed a personalized evaluation system whose goal is to build a flexible and easy way to manage resources in a personalized manner. Our proposed model ensures for every user that his preferences are important and permits the formation of some homogenous groups on the basis of these preferences. The homogeneity is due to the relations resulting from the explicit and implicit evaluations of users. The purpose of the proposed model is to build a flexible way to filter irrelevant resources for users. In this way, a user that is a member of a community based on the PRES model will dynamically see information that he or she is most interested in.

In this work, we study a basis case approached by PRES. For more details related to PRES architecture, see [7]. In the initial system projecting process, we start with the vocabulary defined in the previous models:

- Users that know other users.
- The list of the users considered to be interesting for a given user.
- Users nominated by a community as evaluators. We use notations $\{E_1, \ldots, E_n\}$ to indicate the community evaluators. These evaluators are in fact some reviewers and are useful for the new users that have not established their own knowledge list yet.
- Known person list of a user. Initially, it contains the community reviewers list only.
- Resources: their definition is made accordingly to the definition given by Berners-Lee [8].
- Worth: this parameter is a metric. This metric represents a rating given by a user to another user. Also, the worth can be obtained (quantized) indirectly. This parameter, *Worth*, takes the following values in the following limits: 0–1 = useless/spam; 1–2 = poor; 2–3 = worth attention; 3–4 = good; 4–5 = exceptional. We note this limit with *MaxWorth*, and its value is 5.

We considered a set of constructions which have the following associated semantics. In fact, these constructions can be mathematically considered as functions or, from the implementation point of view, they are considered associative tables:

- Explicit worth of a user: $WE\_UU(user, user)$ represents the rating for a user, and the rating is given manually by the user to another user.
- Implicit (deducted) worth of a user: $WI\_UU(user, user)$ measures how close are his preferences to the others' preferences (the preference can be considered: the accepting degree of a point of view or the appreciation degree of a piece of art). We consider implicitly that $WI\_UU(user, evaluator) = MaxWorth$.

If a user evaluates an evaluator in an explicit manner, then this evaluation, $WE\_UU(user, evaluator)$, will have priority. We consider the function $WU(user, user)$ for each pair of $(user, user)$ from a Web community. Its value is computed as:

$$WU(U_x, U_y) = \begin{cases} WE\_UU(U_x, U_y), & \text{if } U_x \text{ evaluates explicitly } U_y \\ WI\_UU(U_x, U_y), & \text{otherwise.} \end{cases} \tag{2.1}$$

Thus, a rating given by user U_i to the user U_j can be $WE\_UU(U_i, U_i)$ or $WI\_UU(U_i, U_i)$, and we will specify for each case which equality it is. Let us define the manner of computation of the implicit values introduced above.

Therefore, let us consider two users U_i and U_j from the Web community. The value of $WI\_UU(U_i, U_j)$ indicates the deducted worth based on explicit evaluations made by users to each other. Let the users, whom we have explicit ratings from user U_i, be $\{U_i^1, \ldots, U_i^k\}$. So, we know the values $WE\_UU(U_i, U_i^l), l \leq k$. Also, we consider having explicit ratings from U_i^l to U_j, so we get $WE\_UU(U_i^l, U_j)$ (see the example in Fig. 1).

To evaluate the WI_UU value, we must consider which is the value of the weight corresponding with the explicit ratings. We denote this weight with $PE(U_i, U_i^l)$. It represents an explicit rating weight, in our case the weight of the rating provided by U_i, being computed as

$$PE(U_i, U_i^l) = \frac{WE\_UU(U_i, U_i^l)}{k^* MaxWorth} \tag{2.2}$$

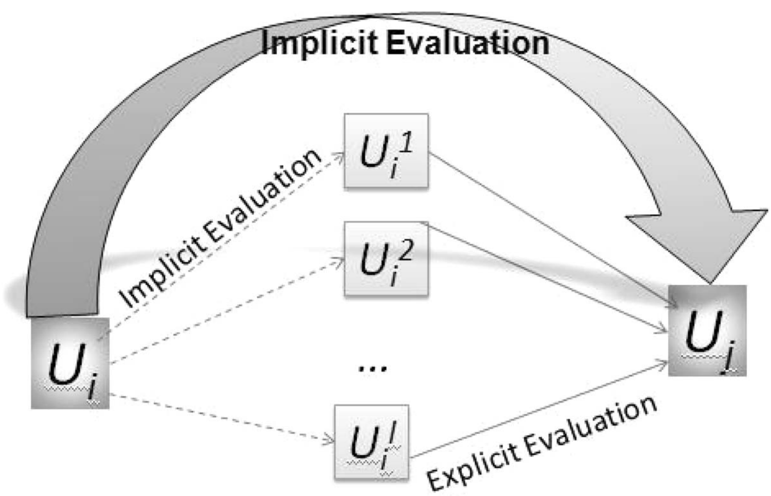

FIGURE 1 WI_UU computation based on explicit evaluations.

where k is the number of the users that were explicitly evaluated by U_i. We compute the implicit rating whom user U_i provided to U_j as

$$WI\_UU(U_i, U_j) = \sum_{l=1}^{k} PE(U_i, U_i^l)^* WI\_UU(U_i^l, U_j) \tag{2.3}$$

where $1 \leq l \leq k$. We have $WI\_UU(U_i^l, U_j) = WE\_UU(U_i^l, U_j)$ if there exists an explicit evaluation from U_i^l to U_j, otherwise we consider an implicit evaluation from U_i^l to U_j. From (2.1) and (2.2), we obtain the final implicit reputation computing formula:

$$WI\_UU(U_i, U_j) = \frac{\sum_{l=1}^{k} WE\_UU(U_i, U_i^l)^* WI\_UU(U_i^l, U_j)}{k^* MaxWorth}. \tag{2.4}$$

3. USER REPUTATION FEATURE EXTRACTION

The first step of any pattern recognition system [12, 13] is the feature extraction procedure. In this section, we present this stage of the Web community user recognition process.

For each user of the reputation system, we consider its reputation as a feature vector. Thus, our feature extraction process becomes a user reputation computing operation. Therefore, let us consider a Web community consisting of the users to be classified $\{U_1, \ldots, U_n\}$. The matrix R, containing their ratings, is defined as follows:

$$R(i, j) = \text{rating given by } U_i \text{ to } U_j, \quad \forall i, j \in [1, n]. \tag{3.1}$$

The ratings contained by R can be explicit or implicit ratings, so they are obtained from relations (2.1)–(2.4). We propose the following algorithm that computes the ratings' matrix:

1. Insert the explicit ratings in R
2. For each i, $1 \leq i \leq n$
   For each j, $1 \leq j \leq n$
   Find $\{U_i^k, \ldots, U_i^k\}$ satisfying the following conditions:

   - User U_i evaluates each of these users
   - There exists evaluations from users $U_i^k, \ldots, U_i^k$ to U_j

   We get a temporary value for $R(i, j)$ using (3.1).
3. Repeat Step 2 until R does not change anymore.
   We get a final value for R at the computing moment.

Let us assume that the set of users interacting with U_i, as evaluators or evaluated, is $\{U_i^1, \ldots, U_i^t\}$, with $t < n$. Thus, using the previous algorithm, we compute for each user U_i the following feature vector:

$$V_{rep}(U_i) = \begin{bmatrix} R(Id(U_i^1), i) & \ldots & R(Id(U_i^t), i) \\ R(i, Id(U_i^1)) & \ldots & R(i, Id(U_i^t)) \\ Id(U_i^1) & \ldots & Id(U_i^t) \end{bmatrix}, \quad \forall i \in [1, n] \tag{3.2}$$

where $Id(U_i^j)$ is the index of user U_i^j. Obviously, if U_i does not evaluate U_i^j, $1 \le j \le t$, then $R(i, Id(U_i^j)) = 0$. Also, if U_i^j does not evaluate U_i, then $R(Id(U_i^j), i) = 0$.

One can observe that each feature vector within the Web community represents a matrix having three rows but a variable number of columns, depending on the number of users with whom its corresponding user interacts. Thus, the distance between two reputations given as in formula (3.2) cannot be computed using the Euclidean metric, for example. For this reason, we provide a new special nonlinear metric in the next section, to be used in the feature vector classification process.

4. A NONLINEAR HAUSDORFF-BASED METRIC

In this section, we propose a special nonlinear metric that is able to compute the distance between matrices having different sizes. In fact, in our case we must compare two matrices having only one common dimension.

The new distance we provide derives from the Hausdorff metric for sets [9–11]. Let X and Y be two compact subsets of a metric space M. The Hausdorff distance between them, $d_H(X,Y)$, is defined as the minimal number r such that the closed r-neighborhood of any x in X contains at least one point y of Y and vice versa.

Thus, if $dist\,(x, y)$ denotes the distance in M, then the Hausdorff metric can be formally described as:

$$d_H(X, Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} dist(x, y), \sup_{y\in Y}\inf_{x\in X} dist(x, y)\right\}. \tag{4.1}$$

The components of this Hausdorff metric, $\sup_{x\in X}\inf_{y\in Y} dist(x, y)$ and $\sup_{y\in Y}\inf_{x\in X} dist(x, y)$, respectively, are sometimes termed as *forward* and *backward* Hausdorff distances of X to Y [9].

Now, let us consider two matrices A and B, having one common dimension, instead of sets [11]. Let them be $A = (a_{ij})_{n\times m}$ and $B = (b_{ij})_{n\times p}$. We introduce two more helping vectors, $y = (y_i)_{p\times 1}$ and $z = (z_i)_{m\times 1}$. Then, let us compute $\|y\|_p = \max_{0\le i\le p} |y_i|$ and $\|z\|_m = \max_{0\le i\le m} |z_i|$. With these notations, we create a new nonlinear metric d having the following form:

$$d(A,B) = \max\left\{ \sup_{\|y\|_p\le 1} \inf_{\|z\|_m\le 1} \|By - Az\|, \sup_{\|z\|_m\le 1} \inf_{\|y\|_p\le 1} \|By - Az\| \right\}. \tag{4.2}$$

This restriction-based metric represents the Hausdorff distance between the sets $B(y : \|y\|_p \le 1)$ and $A(z : \|z\|_m \le 1)$ in the metric space R^n. Therefore, it can be written as

$$d(A,B) = d_H(B(y : \|y\|_p \le 1), A(z : \|z\|_m \le 1)). \tag{4.3}$$

Let us further process formula (4.2). We have $By - Az = \sum_{k=1}^{p} b_{ik}y_k - \sum_{j=1}^{m} a_{ij}z_j$, so we get $\|By - Az\|_n = \max_{1\le i\le n} |\sum_{k=1}^{p} b_{ik}y_k - \sum_{j=1}^{m} a_{ij}z_j|$. Therefore, the following formula results:

$$\sup_{\|y\|_p\le 1} \inf_{\|z\|_m\le 1} \|By - Az\|_n = \sup_{\|y\|_p\le 1} \inf_{\|z\|_m\le 1} \max_{1\le i\le n} \left| \sum_{k=1}^{p} b_{ik}y_k - \sum_{j=1}^{m} a_{ij}z_j \right|. \tag{4.4}$$

This can be seen as a *max min* optimization problem, and according to the classic J. von Neumann min max theorem we have:

$$\sup_{\|y\|_p\le 1} \inf_{\|z\|_m\le 1} \|By - Az\|_n = \inf_{\|z\|_m\le 1} \sup_{\|y\|_p\le 1} \|By - Az\|_n. \tag{4.5}$$

Moreover, the saddle point (y_0, z_0) of this problem can be computed by solving the system

$$\begin{cases} \nabla_z(\|By - Az\|_n) + \eta_1 = 0 \\ \nabla_y(\|By - Az\|_n) + \eta_2 = 0 \end{cases}, \quad (\eta_1, \eta_2) \in N(y,z) \tag{4.6}$$

where $N(y,z)$ is the normal cone to the set $\{y; \|y\|_p \le 1\} \times \{z; \|z\|_m \le 1\}$ and which can be expressed in terms of the Lagrange multipliers. Finally, $\{\nabla_y, \nabla_z\}$ are gradients taken in generalized sense of convex analysis.

However, because (4.4) is hard to compute for large dimensions of A and B, we shall replace it by a simpler one. Thus, we will replace the set $\{y \mid \|y\|_p \le 1\}$ with $F = \{\underbrace{\{1,0,\ldots,0\},\{0,1,\ldots,0\},\ldots,\{0,0,\ldots,1\}}_{p \text{ components}}\}$ and the set $\{z \mid \|z\|_m \le 1\}$ with $G = \{\underbrace{\{1,0,\ldots,0\},\{0,1,\ldots,0\},\ldots,\{0,0,\ldots,1\}}_{m \text{ components}}\}$. Therefore, we could take:

$$\sup_{y\in F} \inf_{z\in G} \|By - Az\|_n = \sup_{1\le k\le p} \inf_{1\le j\le m} \sup_{1\le i\le n} \left| b_{ik} - a_{ij} \right|. \tag{4.7}$$

Of course, the above formula is not identical with (4.4), but it can be regarded as a good approximation for it. As a matter of fact, we have replaced optimization problem on convex set $\{y; \|y\|_p \leq 1\} \times \{z; \|z\|_m \leq 1\}$ with one on a simpler $F \times G$ on its boundary.

Similarly, we could replace $\sup_{\|z\|_m \leq 1} \inf_{\|y\|_p \leq 1} \|By - Az\|$ in formula (4.2) by $\sup_{1 \leq i \leq m} \inf_{1 \leq k \leq p} \sup_{1 \leq i \leq n} |b_{ik} - a_{ij}|$. Therefore, after this elimination of the terms y and z, the transformed formula (4.2) becomes the following Hausdorff-based distance:

$$d(A, B) = \max\left\{ \sup_{1 \leq k \leq p} \inf_{1 \leq i \leq m} \sup_{1 \leq i \leq n} \left| b_{ik} - a_{ij} \right|, \sup_{1 \leq i \leq m} \inf_{1 \leq k \leq p} \sup_{1 \leq i \leq n} \left| b_{ik} - a_{ij} \right| \right\}. \quad (4.8)$$

This resulted nonlinear function d verifies the three essential properties of a distance:

- Positivity: $d(A, B) \geq 0$
- Symmetry: $d(A, B) = d(B, A)$
- Triangle inequality: $d(A, B) + d(B, C) \geq d(A, C)$

Therefore, this Hausdorff-derived function constitutes a metric, although it does not represent the standard Hausdorff metric anymore [11]. It defines a new metric topology on the space of all matrices $\{A\}$ that is not equivalent but comparable with that induced by Hausdorff topology. The newly introduced distance will be successfully used in the reputation classification process.

5. AN AUTOMATIC USER CLASSIFICATION APPROACH

As the next step of the user recognition process, we propose an automatic unsupervised user classification procedure [12, 13]. As we know, the feature vectors, representing user reputations and computed in the third section, are compared using the newly proposed nonlinear metric [9–11].

The classification process must divide the Web community $\{U_1, \ldots, U_n\}$ in several similarity classes. We consider two users to be similar, $U_1 \approx U_2$, and having to be introduced in the same class, if they interact with almost the same users and the received and given ratings are closed enough. Our classification approach is both unsupervised, no training set being needed, and also automatic, no interactivity being present in this process [12]. The number of clusters (classes) is unknown.

The proposed clustering algorithm consists of two parts: the first one operates on the feature vectors, and the second operates on the distances between them. Thus, the first part uses a hierarchical region-growing procedure [12]. It starts with all feature vectors as clusters, unifying the closest clusters at each step and registering the corresponding minimum distances between those. Thus, after performing this clustering process, all regions are finally merged into a single cluster.

The second part analyzes the computed minimum distances between clusters. Another region-growing algorithm is applied to them, these distance values being clustered in two categories: "*large*" distances and "*small*" distances. The small distances are those that will determine the feature vector classes and the corresponding user classes.

There are several techniques to compute the distance between two classes. We may choose the *single linkage clustering* metric, described as

$$d_{Class}(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y), \tag{5.1}$$

the *complete linkage clustering*, defined as

$$d_{Class}(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y), \tag{5.2}$$

or the *average linkage clustering*, described as

$$d_{Class}(C_i, C_j) = \frac{\sum_{x \in C_i} \sum_{y \in C_j} d(x, y)}{card(C_i) \cdot card(C_j)}, \tag{5.3}$$

where d is the Hausdorff-derived metric given by (4.8) and $card(C_i)$ represents the cardinal of the class C_i. We consider single linkage clustering technique the most proper metric for our automatic classifier [12]. This unsupervised automatic classification procedure is described by the following steps:

1. Initialize a distance set: $D = \phi$.
2. Start the classification process with all the feature vectors as the n initial clusters: $C_1 = \{V_{rep}(U_1)\}, \ldots, C_n = \{V_{rep}(U_n)\}$.
3. At each iteration compute the *overall minimum distance* between clusters and merge those being at that distance from each other:

$$\forall i < j, \quad d_{Class}(C_i, C_j) = d_{\min} \Rightarrow C_i = C_i \cup C_j, \quad C_j = \phi, \tag{5.4}$$

where

$$d_{min} = \min_{i \neq j \in [1,n]} d_{class}(C_i, C_j) \tag{5.5}$$

and the distance d_{Class} is computed from (5.1). Minimum distance is then registered:

$$D = D \cup \{d_{\min}\}.$$

4. When a single cluster remains, a new clustering process is performed on the distance set D, using a region-growing algorithm with number of classes $K = 2$. Two classes containing distance values are thus obtained.
5. One element from each class is randomly selected and the two distance values are then compared. The class corresponding with the greater value represents the set of large distances, let it be D_l. The smaller value belongs to the set of small distances, D_s. Obviously, $D = D_l \cup D_s$.
6. Each user receives its initial class label: $\forall i \in [1, n], C(U_i) = i$.
7. For any small distance, it searches for all pairs of vectors corresponding with it and the users related to the feature vectors from each pair will be inserted in the same class:

$$\forall dis \in D_s, \quad \forall i < j \in [1, n], \quad \text{if } d(V_{rep}(U_i), V_{rep}(U_j)) = dis \Rightarrow C(U_j) = i. \tag{5.6}$$

All the Web community users being classified in these clusters, the recognition process is thus finalized.

6. EXPERIMENTS

We have performed a lot of experiments using our recognition technique and obtained satisfactory results. Let us describe one of our numerical tests. We consider a Web community of 9 users and set some explicit ratings. These ratings are given in Table 1.

Then, we compute the implicit ratings on their basis, using relation 2. The feature vectors, or user reputations, are determined next. We get the following values for these vectors:

$$V_{rep}(U_1) = \begin{bmatrix} 4 & 3 & 0 & 0 & 0.8 & 1.2 & 2 \\ 3 & 3 & 1 & 0.6 & 0 & 0 & 0 \\ 2 & 3 & 4 & 5 & 6 & 8 & 9 \end{bmatrix};$$

$$V_{rep}(U_2) = \begin{bmatrix} 3 & 3 & 1.8 & 0 & 0 & 0.48 & 0.72 & 1.2 \\ 4 & 3 & 5 & 0.7 & 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 5 & 6 & 8 & 9 \end{bmatrix};$$

TABLE 1 Table of the explicit ratings.

User Id	1	1	1	2	2	3	3	4	5	5	7	6	9	9
Evaluated user	2	3	9	1	2	2	8	1	2	9	6	8	6	8
Rating	4	3	2	3	3	5	2	1	1	5	1	5	2	4

$$V_{rep}(U_3) = \begin{bmatrix} 3 & 5 & 0 & 0 & 0.48 & 2 & 1.2 \\ 3 & 1 & 0.6 & 0.36 & 0 & 0 & 0 \\ 1 & 2 & 4 & 5 & 6 & 8 & 9 \end{bmatrix};$$

$$V_{rep}(U_4) = \begin{bmatrix} 1 & 0.7 & 0.6 & 0.16 & 0.24 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 6 & 8 & 9 \end{bmatrix};$$

$$V_{rep}(U_5) = \begin{bmatrix} 0.6 & 1 & 0.36 & 2 & 2.05 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 6 & 8 & 9 \end{bmatrix};$$

$$V_{rep}(U_6) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 20.8 & 0.48 & 0.48 & 0.16 & 2 & 1 & 0 & 2 \\ 1 & 2 & 3 & 4 & 5 & 7 & 8 & 9 \end{bmatrix};$$

$$V_{rep}(U_7) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 6 & 8 \end{bmatrix};$$

$$V_{rep}(U_8) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.2 & 0.72 & 2 & 0.24 & 2.05 & 5 & 1 & 4 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 9 \end{bmatrix};$$

$$V_{rep}(U_9) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 2 & 4 \\ 2 & 1.2 & 1.2 & 0.4 & 5 & 0 & 0 \\ 1 & 2 & 3 & 4 & 5 & 6 & 8 \end{bmatrix}.$$

We compute the distances between these feature vectors, using formula (4.8), and obtain the distance matrix shown in Fig. 2.

Using the matrix displayed in the figure above, we apply the clustering algorithm and obtain the minimum distance set $D = \{2, 2.05, 3, 4\}$. Then, applying region-growing to set D, we get $D_s = \{2, 2.05, 3\}$ and $D_l = \{4\}$. Using these distance sets, we get the following classes of Web community users: $\{U_1, U_2, U_3, U_4, U_5, U_6, U_9\}$, $\{U_7\}$, and $\{U_8\}$, respectively.

0	2.0000	2.0000	3.0000	3.0000	4.0000	4.0000	4.0000	4.0000
2.0000	0	2.0000	5.0000	5.0000	3.8000	5.0000	4.0000	3.0000
2.0000	2.0000	0	4.0000	4.0000	5.0000	5.0000	5.0000	4.0000
3.0000	5.0000	4.0000	0	4.6000	4.6000	5.0000	5.0000	5.0000
3.0000	5.0000	4.0000	4.6000	0	2.0500	5.0000	5.0000	5.0000
4.0000	3.8000	5.0000	4.6000	2.0500	0	5.0000	5.0000	3.0000
4.0000	5.0000	5.0000	5.0000	5.0000	5.0000	0	5.0000	5.0000
4.0000	4.0000	5.0000	5.0000	5.0000	5.0000	5.0000	0	4.0000
4.0000	3.0000	4.0000	5.0000	5.0000	3.0000	5.0000	4.0000	0

FIGURE 2 Distance matrix.

7. CONCLUSIONS

This work combined three mathematical and applied mathematics domains: pattern recognition, reputation systems, and Hausdorff metrics. Important personal contributions have been presented in this paper.

The algorithm that computes the implicit ratings from the explicit ones represents the first main contribution. Next, a novel feature extraction approach has been proposed by us. The nonlinear Hausdorff-derived metric is also a very important contribution of our work. The unsupervised automatic reputation classification technique represents a novelty in the pattern recognition domain. Also, some numerical experiments have been presented in this paper.

REFERENCES

1. A. N. Langville and C. D. Meyer (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ.
2. N. Tinarev (2007). Explanations of recommendations. In: *Proceedings of the 2007 ACM Conference on Recommender Systems*. Minneapolis, pp. 203–206.
3. G. Adomavicius and A. Tuzhilin (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowledge Data Eng.* 17(6):734–749.
4. A. Jøsang (2008). Prospectives for online trust management. *IEEE Trans. Knowledge Data Eng.*
5. J. Golbeck and J. Hendler (2004). Accuracy of metrics for inferring trust and reputation in semantic Web-based social networks. *Proceedings of 14th International Conference on Knowledge Engineering and Knowledge Management*. Northamptonshire, UK.
6. J. Golbeck and J. Hendler (2006). Inferring trust relationships in Web-based social networks. *ACM Trans. Internet Technol.* 6(4).
7. L. Alboaie (2008). PReS – personalized evaluation system in a Web community. ICE-B, 2008, Portugal, ISI Proceedings, IEEE.
8. T. Berners-Lee, R. Fielding, and L. Masinter (eds.) (1998). Uniform Resource Identifiers (URI): General Syntax. *RFC 2396, IETF.*
9. O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz (2001). Robust face detection using the Hausdorff distance. In: *Proc. of Third International Conference on Audio and Video Based Biometric Person Authentication*. Lecture Notes in Computer Science, LNCS 2091. Springer, Berlin, pp. 90–95.
10. N. Gregorie and M. Bouillot (1998). Hausdorff distance between convex polygons. *Web Project for the Course CS 507 Computational Geometry*. McGill University.
11. T. Barbu (2004). Discrete speech recognition using a Hausdorff based metric. In: *Proceedings of the 1st International Conference of E-Business and Telecommunication Networks, ICETE 2004*. Vol. 3, Setubal, Portugal, pp. 363–368.
12. T. Barbu (2006). An automatic unsupervised pattern recognition approach. In: *Proceedings of the Romanian Academy*. Series A, Vol. 7, No. 1. Publishing House of the Romanian Academy, Bucharest, Romania.
13. R. Duda, P. Hart, and D. G. Stork (2000). *Pattern Classification*. John Wiley & Sons, Hoboken, NJ.