Seismic Damage Prediction of Buried Pipeline by KDD Method

M.B. Javanbarg¹, S. Takada², and Y. Kuwata²

1. Graduate School of Science and Technology, Kobe University, Kobe, Japan
2. Department of Architecture and Civil Engineering Department, Kobe University, Kobe, Japan
email: takada@kobe-u.ac.jp

ABSTRACT: Damage prediction of buried pipeline under earthquake environments is the first stage for the seismic risk analysis. In this paper, we use a Knowledge Discovery in Database (KDD) method for the pipeline damage prediction even though many studies have been performed so far with the aid of empirical, statistical, and/or theoretical methods. By employing the KDD method, much higher accurate damage prediction could be done for better understanding of pipeline damage distribution. Related factors were analyzed by a GIS based model of the Kobe water buried pipelines in the 1995 Kobe Earthquake, and a decision tree of pipeline damage classification was developed based on the Classification and Regression Tree (CART) method. A verification of the method was focused to the modeled area, and accuracy of the proposed prediction method was confirmed in comparison with an actual damage as well as predicted ones by commonly used formula of damage estimation. Results of the developed KDD model showed that the model could predict correctly the number of damage in pipeline network. The proposed method by KDD turned out the distribution of damage better than other damage estimation methods.

Keywords: Seismic damage prediction; Buried pipeline; Knowledge Discovery in Database (KDD)

1. Introduction

Seismic risk analysis of a buried pipeline system is a methodology developed to minimize the probability of system breakdown and reduce the losses from damage due to future earthquakes. Seismic risk analysis requires the evaluation of pipeline damage under earthquake environments. For instance, a comprehensive investigation of water pipeline damage after the 1995 Kobe Earthquake undertaken by the Japan Water Works Association has been described by Shirozu et al [1]. An estimation formula of seismic damage for pipelines was also proposed based on the detailed investigation of the buried pipeline damage in the Kobe Earthquake by Takada et al [2]. Takada also employed this formula to estimate the physical damage and interruption effects in Tehran water pipeline system [3]. Another example of the usage of the estimation formula is due to failure of the pipeline system in the 1976 Tangshan earthquake [4].

A high accurate damage prediction methodology must be capable of better understanding the distribution of pipeline damage due to earthquakes. On the other hand, KDD has recently become a very valuable data analytic process for detecting the association of different related factors in large data sets such as damage prediction in structural mechanics [5]. Javanbarg et al [6] employed KDD technique for damage analyses of two suffered areas in the 2003 Tokachi-Oki and the 2004 Niigata Chuetsu Earthquakes. By analyzing the factors affecting pipeline damage, they found that the preliminary factor of mining pipe damage was not seismic intensity, but geomorphology under a certain level of ground motions. In those areas, however, there were not enough data of predictors affecting the damage distribution. Accordingly, due to the comprehensive data sets of the 1995 Kobe Earthquake, in this study, the focus is addressed on
prediction of pipeline damage distribution in Kobe City based on the KDD technology.

2. KDD Theory [7]

Knowledge discovery from data refers to the process of extracting interesting, non-trivial, implicit, previously unknown, and potentially useful information or patterns from data. Here, we explain a summary of the KDD theory from reference [7].

There are two keys to success in KDD. First is to come up with a precise formulation of the problem that we are trying to solve and the second key is to use the right data.

The Knowledge Discovery in Databases process is comprised of a few steps leading from raw data collections to some forms of new knowledge. The iterative process consists of the following steps:

- Data cleaning: a phase in which noise data and irrelevant data are removed from the collection.
- Data integration: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.
- Data selection: at this step, the data relevant to the analysis is decided and retrieved from the data collection.
- Data transformation: also known as data consolidation, is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- Data mining: is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- Pattern evaluation: in this step, strictly interesting patterns representing knowledge are identified based on given measures.
- Knowledge representation: the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

These kinds of patterns can be discovered depending on the KDD tasks employed. By and large, there are two types of KDD tasks; descriptive KDD tasks that describe the general properties of the existing data, and predictive KDD tasks that attempt to do predictions based on inference on available data.

2.1. KDD Terminology

In predictive models, the values or classes that we are predicting, are called the response, dependent or target variables. The values used to make the prediction are called the predictor or independent variables. Predictive models are built, or trained, using data for which the value of the response variable is already known. This kind of training is sometimes referred to as supervised learning, because calculated or estimated values are compared with the known results.

2.1.1. Classification

Classification problems aim to identify the characteristics that indicate the group to which each case belongs. This pattern can be used both to understand the existing data and to predict how new instances will behave. KDD creates classification models by examining already classified data (cases) and inductively finding a predictive pattern. These existing cases may come from a database, such as GIS database.

2.1.2. Regression and Decision Tree

Regression uses existing values to forecast what other values will be. In the simplest case, regression uses standard statistical techniques such as linear regression. The same model types can often be used for both regression and classification. Decision trees are a way of representing a series of rules that lead to a class or value. The decision tree analyzes (mines) a set of data values and generates a decision tree that can be used to predict the value of a target variable based on the values of a set of predictor variables. Like a real tree, a decision tree has a root, branches, and leaves. A prediction is made by entering the tree at the root and following the branches left or right based on values of the predictor variables until a leaf is reached. Each leaf shows the most likely value for the target variable given the set of predictor values that led to the leaf. There are two steps to making productive use of decision trees; 1) building a decision tree model, and 2) using the decision tree to draw inferences and make predictions.

2.1.3. Rule Induction

Rule induction is a method for deriving a set of rules to classify cases. Although decision trees can produce a set of rules, rule induction methods generate a set of independent rules which do not necessarily (and are unlikely to) form a tree. Because the rule inducer is not forcing splits at each level,
and can look ahead, it may be able to find different and sometimes better patterns for classification. Unlike trees, the rules generated may not cover all possible situations. Also unlike trees, rules may sometimes conflict in their predictions, in which case it is necessary to choose which rule to follow.

In order to apply the KDD method for pipeline damage prediction, the classification and regression tree (CART) was employed as the KDD model to predict pipeline damage due to the 1995 Kobe Earthquake. In particular, we used the KDD techniques of decision trees. The CART methodology is a relatively new approach to the problem of predicting a response (target) variable on the basis of several predictor variables. A very interesting advantage of CART is the possibility to deal with large numbers of both categorical and numerical variables. Another advantage is that no assumption about the underlying distribution of the predictor variables is required (even categorical variables can be used). Eventually, CART provides a graphical representation, which makes the interpretation of the results easy. Therefore, CART could be a very interesting method to predict the distribution of damage within a pipeline database.

2.2. Classification and Regression Tree [8]

Classification and Regression Tree analysis is a statistical method that explains the variations of target variable using a set of explanatory variables, so-called predictors. In other words, CART analysis is the organization of data in given classes. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is then used to classify new objects. For instance, in the case of pipeline damage analysis under earthquake environment, the model can analyze vulnerability of the pipeline network due to the predictors such as ground condition, geomorphology, liquefaction, seismic intensity, pipe diameter, pipe material and pipe length.

2.2.1. Tree-Growing Process

CART works by splitting the data into mutually exclusive subgroups, called nodes, within which the objects have similar values for the target variable. The process starts from the root or parent node, which contains all objects of data set. CART uses a repeated binary splitting procedure, which means that the parent node is split in two nodes, called child nodes. The process is repeated by treating each child node as a parent node, see Figure (1). Each split is defined by a simple rule, usually based on a predictor. For categorical variables, a split is defined by relating one or more levels of the variable to a specific node. Trees are grown by selecting the splits in such a way that so-called homogeneity and the impurity of the target variable within each node is maximized and minimized, respectively. To achieve this, CART looks at the possible splits for all variables included in the analysis. The resulting splits are compared and eventually, the best splits are chosen by evaluation of the impurity of the formed nodes, based on the statistical criteria. This procedure is repeated consecutive split made in the tree. The splitting procedure is continued until no further split can be performed, i.e. all child nodes are homogenous, or contain one or a user-defined minimal number of observations. The tree thus obtained is called the maximal tree and the terminal nodes, so-called leaves, represent the final groups formed by the tree.

This maximal tree usually contains too many leaves and causes poor predictive abilities for new samples. Therefore, the selection of an optimal tree with a good compromise between model fit and predictive properties is required. Thus, in general, CART analysis consists of three steps; 1) the maximal-tree building, 2) the tree pruning which consists in the cutting-off of nodes to generate a sequence of simpler (i.e. smaller) trees, and 3) the optimal-tree selection.

2.2.2. CART Mathematical Algorithm

CART works by choosing a split at each node in a way that each child node is more pure than its parent node. Here purity refers to the values of the target
variable. In a completely pure node, all of the cases have the same value for the target variable. CART measures the impurity of a split at a node by defining an impurity measure (Gini index). There are four different impurity measures used to find splits for CART models, depending on the type of the target variable. In our approach Gini measure was used and it has been explained in detail in the following section.

In Figure (2), the structure of a node via the CART analysis is presented; in which \( t \) and \( n \) are the node number and the number of cases in target class, respectively, \( i \) and \( j \) are categories of target class, \( N_i(t) \) and \( N_j(t) \) are the number of category \( i \) and \( j \) which take place in node \( t \), respectively. The prior probability (value) affecting the misclassification rates for category \( j \), \( \pi(j) \) can be considered. If \( N_j \) is the number of cases of category \( j \) in the root node, the joint probability of category \( j \) at node \( t \), \( p(j,t) \), can be then defined as the proportion of the number of category \( j \) in node \( t \) to the number of category \( j \) in root node as follows,

\[
p(j,t) = \frac{\pi(j)N_j(t)}{N_j}
\]

The Gini index at node \( t \), \( g(t) \), is defined as,

\[
g(t) = \sum_{j \neq i} p(j/t)p(i/t)\sum_{j} p(j,t)p(t)
\]

\[
p(j/t) = \frac{p(j,t)}{p(t)}
\]

\[
p(t) = \sum_j p(j,t)
\]

When the Gini index is used to find the improvement for a split during the growth, only those cases in node \( t \) and the root node with valid values for the split-predictor are used to compute \( N_j(t) \) and \( N_j \), respectively.

The Gini criterion function \( \Delta g(s,t) \) for split \( s \) at node \( t \) is defined as:

\[
\Delta g(s,t) = g(t) - p_Lg(t_L) - p_Rg(t_R)
\]

Where \( s \) is a particular split, \( g(t_L) \) and \( g(t_R) \) are the impurity of the left and right child nodes, respectively, \( p_L \), the proportion of cases in node \( t \) is sent to the left child node, and \( p_R \), the proportion is sent to the right child node. The proportions \( p_L \) and \( p_R \) are defined as:

\[
p_L = \frac{p(t_L)}{p(t)}
\]

and

\[
p_R = \frac{p(t_R)}{p(t)}
\]

The split \( s \) is chosen to maximize the value of \( \Delta g(s,t) \). This value is reported as the “improvement degree” in the tree. Therefore steps in CART are as follows:

1) Starting from the root node \( t = 1 \), search for a split \( s \) among the set of all possible candidates \( S \) that give the largest decrease in impurity:

\[
\Delta g(s,l) = \max_{s \in S} \Delta g(s,l)
\]

Then split node 1 \( (t = 1) \) into two nodes, \( t = 2 \) and \( t = 3 \), using split \( s \).

2) Repeat the split-searching process in each of \( t = 2 \) and \( t = 3 \), and so on.

3) Continue the tree-growing process until at least one of the stopping rules is met.

2.2.3. Accuracy of the Tree

Once a tree has been generated, it is always important to consider the accuracy of the tree. Accuracy refers to how well the tree predicts outcomes or classifies individuals. Conversely, the inaccuracy of the tree is called the risk. It may then be possible to estimate the risk of the tree. The less risk results the more accuracy. Risk can be calculated in different ways depending on the nature of the target variable. For instance, the risk calculation by resubstitution is presented in the following section.

Table (1) shows the number of cases corresponding to both the actual data of target variable and prediction by CART model. Definitions of parameters in Table (1) are as follows:

\( N_i \): The number of cases in category \( i \) that are classified as category \( i \)

\( N_j \): The number of cases in category \( j \) that are...
Seismic Damage Prediction of Buried Pipeline by KDD Method

### 4. GIS Database Construction of the CART Model

In order to build the CART model, maps of Kobe buried pipeline including pipeline damage locations as the target class and predictors such as ground condition, liquefied area, seismic intensity and pipe length from JWWA report were digitized and overlaid via a GIS database. In case of predictors such as pipe material and geomorphology classes, Kobe JIBANKUN geo-database was employed [10]. Table (4) shows the classification of the predictors. As it can be seen for pipe material, two classes consist of ductile iron pipe (DIP), and others including cast iron pipe (CIP), vinyl (VP) and steel pipes (SP) were considered. As mentioned in section 2.1, predictors should be independent factors. In order to show the correlation of the predictors, the correlation coefficient for each predictor related to others is calculated and presented in Table (5). The results for ground condition, geomorphology and liquefaction show high correlation. However, in CART algorithm the weight of each predictor is not necessary to be obtained like statistical methods and the order of the predictors could be determined by tree growing process. Therefore, the dependency of each predictor can be considered in CART, automatically.

#### Table 4. Predictor variables for pipeline damage by CART model.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Ground Condition</td>
<td>1</td>
<td>Stiff</td>
</tr>
<tr>
<td>2</td>
<td>Soft</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Reclaimed</td>
<td></td>
</tr>
<tr>
<td>B: Geomorphology</td>
<td>2</td>
<td>Terrace and Fan</td>
</tr>
<tr>
<td>3</td>
<td>Valley Plain and Levee</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Raised River Bed and Reclaimed Area</td>
<td></td>
</tr>
<tr>
<td>C: Liquefied Area</td>
<td>1</td>
<td>0% Liquefied</td>
</tr>
<tr>
<td>2</td>
<td>50% Liquefied</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100% Liquefied</td>
<td></td>
</tr>
<tr>
<td>D: Seismic Intensity</td>
<td>1</td>
<td>5 JMA</td>
</tr>
<tr>
<td>2</td>
<td>5 JMA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6 JMA</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6 JMA</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7 JMA</td>
<td></td>
</tr>
<tr>
<td>E: Pipe Diameter</td>
<td>1</td>
<td>&lt;75mm</td>
</tr>
<tr>
<td>2</td>
<td>100–150mm</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>200–250mm</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>300–450mm</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>&gt;500mm</td>
<td></td>
</tr>
<tr>
<td>F: Pipe Material</td>
<td>1</td>
<td>DIP</td>
</tr>
<tr>
<td>2</td>
<td>Others (CIP, SP, VP)</td>
<td></td>
</tr>
<tr>
<td>G: Pipe Length</td>
<td>1</td>
<td>Length &lt;200m per Mesh</td>
</tr>
<tr>
<td>2</td>
<td>200m ≤ length &lt;400m per Mesh</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Length ≥400m per Mesh</td>
<td></td>
</tr>
</tbody>
</table>
A 100×100 meter mesh (target mesh) overlaid with pipeline path, damage locations, as well as all the predictor classes. Target class was then classified into two classes due to precision of damage within a mesh; 0 related to no damage in mesh and 1 for damaged mesh. Figure (3) shows the overlaying of target mesh with damage locations. It can be seen that in some parts, the density of damage location is higher than the other parts. By employing the decision tree produced by \textit{CART} model, however, it may clear the effect of each predictor causing damage in those areas.


\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Predictor       & Seismic intensity & Ground condition & Geomorphology & Liquefaction \\
\hline
Seismic intensity & 1                 & -0.07            & -0.30         & -0.54         \\
Ground condition  & -0.07             & 1                 & 0.72          & 0.72          \\
Geomorphology    & -0.30             & 0.72             & 1             & 0.81          \\
Liquefaction     & -0.54             & 0.72             & 0.81          & 1             \\
\hline
\end{tabular}
\caption{Correlation coefficients of predictors.}
\end{table}

In order to clarify the class of predictors, the mesh was overlaid with related raster maps. The classification of target mesh due to the ground condition, geomorphology, liquefied area, and seismic intensity are presented via Figures (4) to (7), respectively. In Figures (4) and (5), ground condition and geomorphology have been classified to three and four classes, respectively. The stiff condition is related to the mountainous area in the northern part, soft condition is regarded with plain in the middle part, and the reclaimed area is situated in the southern part including both man-made Rokko Island and Port Island areas. Figure (6) induces a liquefied area that is mostly related to the reclaimed and coastal land. From Figure (7) it can be seen that the seismic intensity shows higher values in the soft ground condition and middle part of Kobe City.

In case of classifying the target mesh due to the pipe diameter, in precision of two or more classes in one mesh, the number of pipes with the same class diameter was derived in each mesh and the class diameter with majority of number of pipes was then selected as the identification of the mesh diameter class.
5. Damage Prediction by Proposed KDD Method

In order to deal with pipe material, four kinds of meshes were addressed, see Figure (8). In case of meshes, which include DIP, the material class was considered DIP (for example mesh K in Figure (8)). In case of meshes with precision of CIP, SP or VP, (mesh L), they were considered as others. In case of meshes which consist of both DIP and other materials, depending on the majority of other material (mesh M) or majority of DIP material (mesh N), were considered as others and DIP respectively. Pipeline length in each mesh was also calculated and classified in tree classes.

Rokko Island are presented. The CART model was employed for this part of the network. In Table (6) risk analysis (misclassification matrix) and improvement degree of tree for the east part of Kobe water pipeline are presented. The meaning of the number in Table (6) have been defined in Table (1). For instance, 118 is the meshes that actually included damage and the model predicted as the damaged mesh. As it can be seen, the risk of the model is about 27%. Therefore, the accuracy of the model is about 73%. As an evaluation criterion, the improvement degree of the model is also presented in Table (6). It is clear that seismic intensity is a prior factor and the most affective parameter to pipeline damage. The structure of the CART model used for this study was an interpretable tree with 5 levels. Among the terminal nodes, 7 nodes were the desired terminal nodes (nodes that hit the category j, damaged category) in which, each node included related meshes. In Table (7), the desired nodes with 5 rules were derived based on the growing tree of the model and conditional probability in each node was
calculated by the equations induced in section 2.2.2. With respect to the fact that each node includes a number of meshes, the probability of each mesh, \( p(j) \) is calculated based on Eq. (10) and presented in Table (7).

\[
p(j) = \frac{p(j/t)}{N_j(t)} \tag{10}
\]

With respect to the tree growing procedure presented in Table (7), it is worthwhile to return to the rules and criteria in each node to state some of the main points:

- In all nodes, due to the priority of the rules, seismic intensity is the prior predictor (factor) to pipeline damage.
- For the JMA seismic intensity more than 6+, pipe diameter is the second factor that influences damage; meanwhile, in case of JMA seismic intensity equal to 6+, pipeline length is the next predictor.
- For the JMA seismic intensity less than 6+, geomorphology is the second factor affecting pipe damage.

The prediction of KDD technique greatly depends on the feature chosen. A more meaningful attribute produces better results. By employing the tree growing criteria of the eastern part, shown Table (7), damage prediction is constructed for the western part of the network and the prediction results of damaged meshes as well as risk summary calculated by Eq. (9) are presented in Table (8). The risk has been estimated 35% which means 65% accuracy of the model. Comparison between results in Tables (6) and (8) shows that the accuracy of Table (6) is higher than Table (8). In Table (6), classification of the damaged meshes in the eastern part based on the actual damage locations was used; meanwhile, in Table (8) we applied the decision tree criteria of the east part to the west part in order to predict the damage distribution.

Figure (10) shows damaged mesh distribution for the western part of Kobe water pipeline network. Locations of the actual and estimated pipeline damage by KDD method show a rough agreement.

The related probability level of damage for each mesh is shown in Figure (11). The probability of damage can be considered as the damage rate, in which it is equal to number of damage per 1km of pipe length. On the other hand, in case of segmented pipelines such as DIP, CIP, SP and VP, the damage rate in mesh can be defined as the number of damage divided by the number of joint/5m segment of pipes in each mesh. Thus, the number of damage in each 100×100 meter mesh can be equal to probability of damage multiplied by the length of 5-meter segmented pipes.
where $N_d$ and $L$ are the number of damage and length of pipeline in each mesh, respectively.

Therefore, the damage in the western part has been calculated in total 745 locations, see Figure (12), considering the presented calculation in Eq. (11), while the actual damage for the western part is 723 locations. As it can be seen, the proposed KDD method could predict the number of damage correctly. From Figures (11) and (12), it can be seen that the distribution of damage concluded by KDD method has a rough compatibility with actual damage distribution.

### 6. Pipeline Damage Estimation Formula

In order to compare the results of the DM model for pipeline damage estimation with other methods, and based on the detailed investigation of the buried pipeline damage in the 1995 Kobe Earthquake, the estimation formula of seismic damage for pipelines was proposed as following [2]:

$$ N_d = S_d \cdot C_m \cdot C_d \cdot C_l \cdot L $$

where $N_d$ is number of damage locations, $S_d$ is an averaged damage ratio and can be calculated as,

$$ S_d = 4.11 \times 10^{-9} \text{PGA}^{2.92} \text{ for } \text{PGA} \leq 800 \text{gal} $$

and $C_m$ is a coefficient of pipe material, $C_d$ is a coefficient of diameter, $C_l$ is a coefficient of liquefaction, see Table (9), and $L$ is a total length of pipeline (km).

In Figure (13), damage distribution in the west part of the Kobe water pipeline estimated by formula as well as actual damage locations are compared. Similar to the results for CART model, the risk summary calculated by Eq. (9) is presented in Table (10).

### 7. Comparison of the Results

The total number of damage for all pipes derived by KDD method and commonly used estimation formula of pipeline damage are compared with actual ones in Figure (14). The results show that the proposed KDD method could predict the damage number of pipelines from the earthquake

| Table 9. Coefficients in pipeline damage estimation formula [2]. |
|------------------|------------------|------------------|------------------|
|                  | $C_m$ = Coefficient of Material |                  |                  |
| CP                | VP               | DIP (A, K, T)    | DIP (S, SI)      |
| 3.3               | 1                | 0.3              | 0.1             |

| Table 10. Estimation results and risk summary for estimation by formula. |
|------------------|------------------|------------------|
| Prediction by Formula | Actual Number of Damage | Risk Estimate |
| Category 0        | 0                | 1                |
| 0                 | 1290             | 82               |
| 1                 | 2063             | 361              |
| Total             | 3353             | 443              |

Risk Estimate=0.4862
better than the estimation formula. Comparison of the risk estimation results presented in Tables (8) and (10) shows that risk of the damage prediction by KDD method is less than estimation formula and the KDD method can predict the distribution of damage better than the formula. In other words, according to Table (10) for the estimation formula the amount of predicted damaged meshes with actual damage is 361 which is more than 156 number of damages presented in Table (8). On the other hand, the amount of misclassification of damaged meshes in Table (10) is 2063 that is much more higher than 772 in Table (8), and it causes higher risk for estimation formula. In addition, comparison between predicted distributions of damage with actual damage in Figures (12) and (13) shows that the formula has lower agreement than KDD method.

8. Conclusions

This paper presented KDD model to predict the damage number and distribution of damage in the pipeline network. By using the predictors such as ground condition, geomorphology, liquefaction, seismic intensity, pipe diameter, pipe length and material, a KDD model considered the pipeline database features and classified target class as vulnerable or not vulnerable categories. The results of this research are as follows:

- The development of the KDD model showed that the model could correctly predict the number of damage in the pipeline network due to the earthquake.
- By employing the KDD method, much higher accurate damage prediction could be done for better understanding of pipeline damage distribution.
- The accuracy of the proposed prediction method was confirmed in comparison with an actual damage as well as predicted ones by commonly used formula of damage estimation. Results of developed KDD model showed that the model could predict the number of damage better than the formula.
- The prediction of exact location and severity of the damage in the pipeline network database can be a considerable challenge. However, comparison between the distributions of damage by the proposed KDD method and damage estimation formula showed that KDD model has better agreement to actual damage distribution. Further work needs to be done to extract features that will result in a more accurate KDD model.
- The model showed that in case of Kobe water pipeline damage due to 1995 Kobe Earthquake, seismic intensity was the prior factor to pipeline damage.

References


