Korean J. Chem. Eng., 34(10), 2584-2590 (2017)
DOI: 10.1007/s11814-017-0191-1

pISSN: 0256-1115
eISSN: 1975-7220

A prioritization method for replacement of water mains using rank aggregation

Go Bong Choi*, Jong Woo Kim*, Jung Chul Suh, Kwang Ho Jang**, and Jong Min Lee*,†**

*School of Chemical and Biological Engineering Institute of Chemical Processes, Seoul National University,
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
**Samchully Co., Ltd., 42 Gukjegeumyung-ro 6-gil, 9 Yeongdeungpo-gu, Seoul 07328, Korea
(*Received 13 March 2017 • accepted 4 July 2017*)

Abstract−Pipe breaks in municipal water distribution networks may cause serious damage economically and socially. Existing methods for replacement scheduling of pipes do not provide practical indicators for replacing an individual deteriorated pipe. This work formulates the selection problem as the decision of preference ordering or ranking and proposes a bipartite ranking-based approach. The suggested approach also considers loss from broken pipes in terms of the costs associated with broken water main and its repair. We use rank aggregation method to integrate multiple ranks into replacement order of water mains. The suggested framework prioritizes current pipe sections for replacement based on the aggregated ranks. Multiple ranks given by the reliability of water pipe sections are aggregated and a cost effective policy for pipe replacement is derived.

Keywords: Water Pipe Management, Optimal Management, Bipartite Ranking, RankboostB, Rank Aggregation

INTRODUCTION

Systematic management of the water distribution network is becoming more important. Recently, the Ministry of Environment in South Korea reported that the ratio of pipes buried under the ground that aged more than 20 years exceeded 30% nationwide, and the average age of the pipelines in the capital region is over 30 [1]. These deteriorations may bring about pipe breakage, property damage and social issues such as contaminated drinking water. Indeed, the major cause of pipe breakage incidents is deterioration of water mains; about 30% of the total incidents from 2008 to 2013, and the total damage from the breakage of water mains was estimated to be about 8.5 million USD in South Korea [2].

Economic and sustainable maintenance of pipe networks first requires reliable estimation of the pipeline condition. Nevertheless, it is difficult to identify or estimate the condition of pipeline in a proactive manner because the water main is a very complex and intricate network under the ground without much real-time information available. Most accurate and conclusive techniques to determine the states of water mains are visual examination with the naked eye after digging in the ground and nondestructive inspections using expensive equipment. However, if applying these methods to pipes buried in a wide area, the total cost including direct costs for digging and indirect costs for traffic jam and water outage may be prohibitive.

Alternatively, two types models, physical and statistical, are often used to address the cost issue. Physical or mechanistic models can be used to analyze the operational and environmental stresses imposed on a pipe from various external sources and the pipe capacity to hold the loads. Though the physical model can describe the degradation mechanism of a pipe, its application to real pipelines in predicting water pipe breakage is limited because the deterioration mechanism is very different for pipe material types and structural properties. Furthermore, modelling the physical degradation process is time consuming and requires expensive, extensive data. On the other hand, statistical models can be formulated only by using historical records of pipe break events with failure time and environmental data.

There exist previous studies using simple regression models to predict the time to next break of water pipe based on several factors [3]. However, these techniques have been deemed as insufficient to establish a policy for water main management because they exclude the data of unbroken water pipes. The most representative method that can utilize all types of data is the survival analysis [4], which predicts the probability of pipe breaks. In addition, advanced statistical tools such as artificial neural networks [5], fuzzy logic [6], and Bayesian belief networks [7] as well as a simple Poisson distribution [8] are also used.

While many statistical models have been suggested, there is still much room for further improvement for predicting pipe breakage. Debon et al. conclude the generalized linear model (GLM) is the best based on the receiver operating characteristic curve [9], whereas Yamigala et al. report even GLM still needs to be further improved for better prediction accuracy [10]. To the best of our knowledge, existing methods do not provide a straightforward indicator for selecting an individual pipe that needs to be replaced, but indirect, abstract information such as prediction of the total number of failures or the annual frequency of breakages per unit length. Wang et al. propose a rank algorithm to provide candidate pipes for preventive maintenance based on their failure risks [11]. Rank algorithm can use historical data and draw a preference ordering for further investigation. The algorithm shows satisfactory results

†To whom correspondence should be addressed.
E-mail: jongmin@snu.ac.kr

for proactive inspection given limited budgets. For instance, 50% of pipe breaks of waterworks in 2011 could be prevented by inspecting 6.98% of the pipes in advance using data set available at the end of 2010.

However, conditions leading to deterioration of water mains are not the only factors that affect the water pipe replacement policy, as other considerations are also necessary, including the cost of damage. The optimal decision that only considers reliability will be short-term cycle replacement of all pipes. Previous studies on optimal scheduling of water mains considering various factors primarily focus on optimization methods. Xu et al. categorize models for pipe replacement optimization into two types [12]. The first calculates optimal replacement time, where the lifespan is estimated by minimizing the total cost in a multi-objective optimization framework [13-15] and considering the threshold break rate [16-18]. The second prioritizes pipes for replacement given budget constraints. It orders the sequence of pipes to replace based on the replacement cost and reliability of pipes [19-22].

The sequence of replacement from rank algorithm can be used as a criterion for prioritizing pipes for replacement, but a total ranking considering reliability, failure and repair cost is necessary. For this, rank aggregation can be applied to integrate multiple ranks for optimal replacement of water mains. Rank aggregation has been mainly used for spam reduction and search engine comparison. It also shows proper performance to draw consensus of multiple ranks from various sources when aggregating ranking functions [23]. Hence, the final rank reflecting essential factors can be used as a criterion to suggest candidate pipes for replacement [24].

Although a framework integrating the steps of diagnosis and corrective action for the replacement scheduling of water pipes is necessary, previous studies only consider one of the steps and require modification of the policy-prescribing step on an ad hoc basis. For example, survival analysis [25], one of the popular methods for estimating the state of infrastructure, yields the probability of breakage over infinitesimal period of time, which is difficult to use for prescribing a replacement policy. Hence, heuristic policies are often employed instead of optimization [24].

This work suggests a prioritization algorithm based on rank aggregation for prescribing a pipe replacement policy. The proposed approach recasts the optimization problem as an ordering problem so that the diagnosis and policy generation steps are seamlessly integrated into a rank aggregation problem without additional modifications. For estimation of the degree of pipe deterioration, a rank algorithm is employed. Based on this reliability rank, replacement cost, and the number of households, the final rank is estimated by rank aggregation. Hence, the proposed approach can generate the list of candidates to be replaced in order of preference.

The rest of the paper is organized as follows: Section 2 presents algorithms for ranking and rank aggregation. Section 3 provides a case study where the proposed approach is applied to real field data obtained in one city of South Korea and shows the efficacy of the proposed scheme. Finally, Section 4 provides concluding remarks.

PROBLEM STATEMENT

This study is concerned with optimally replacing water mains given budget constraints. This type of problem can be formulated as an optimization problem in (1) [19]. With appropriate assumptions, the preference ordering problem can provide the same result with the optimization problem of (1).

$$\mathrm{obj}=\min\left\{\sum_{j=type}\sum_{i=1}^{n}C_{replacement}(i, j)\times x(i)+E[C_{breakage}(i, j)]\times y(j)\right\} \quad (1)$$

where x(i) is 1 if pipe is replaced, y(i) is 1 if pipe is maintained, is the total number of pipes and $E[C_{breakage}(i, j)]$ is the expected cost of damage with type j if pipe i is broken.

$$C_k(i)=\sum_{j=type}C_{k,\, type}(i)\times j_{type}(i) \quad (2)$$

First, replacing one pipe neither increases the failure rate of other pipes nor gives meaningful benefits to other pipes. Second, instead of determining optimal replacement timing, we determine a series of pipes that should be replaced preferentially for a given period. Finally, the breakage cost is greater than the replacement cost, and different types, e.g., ductile irons, PVC, etc., of pipes can also be considered by introducing a binary parameter $u_{type,k}(i)$ that indicates 1 if pipe i belongs to type k. Given these assumptions and replacing y(i) with $1-x(i)$, the objective function in (1) can be reformulated as (5) and (6).

$$\mathrm{obj}=\min\left\{\sum_{j=type}\sum_{i=1}^{n}C_{replacement}(i, j)\times x(i)+E[C_{breakage}(i, j)]\times y(j)\right\} \quad (3)$$

$$=\sum_{i=1}^{n}E[C_{breakage,\, type}(i)] + \min\left\{\sum_{i=1}^{n}C_{replacement,\, type}(i)\times x(i)-E[C_{breakage,\, type}(i)]\times x(j)\right\} \quad (4)$$

$$=\sum_{i=1}^{n}E[C_{breakage,\, type}(i)] - \max\left\{\sum_{i=1}^{n}E[C_{breakage,\, type}(i)]\times x(i)\right\}-C_{replacement,\, type}(i)\times x(i) \quad (5)$$

$$=\sum_{i=1}^{n}E[C_{breakage,\, type}(i)]-\max\left\{\sum_{i=1}^{n}E[C_{additional\ breakage,\, type}(i)]\times x(i)\right\} \quad (6)$$

where $E[C_{additional\ breakage,type}(i)]=E[C_{breakage,type}(i)]-C_{replacement,type}(i)$. This means minimizing the total cost can be replaced with minimization of the expected cost of damage by changing water mains under limit budget. Therefore, we only need to compare pairwise precedence among pipes. Hence, we can conclude that optimization formulation can be replaced with ranking problem for listing economic policies with reliability guarantee.

If this alternative is feasible, a ranking algorithm, which was shown to be effective for predicting reliability, can be used for deriving a management policy in practice [11]. Because the ranks for pipe replacement need to account for not only reliability of water mains but also other criteria including economic damage, water quality, and the number of households where water can be cutoff, a ranking problem that yields consensus ranking of those factors is necessary. Whereas quantitative information such as costs and the number of households can be easily translated into precedence ordering, it is not straightforward to relate the historical events of pipe breakage and maintenance to the ranks. This work further utilizes

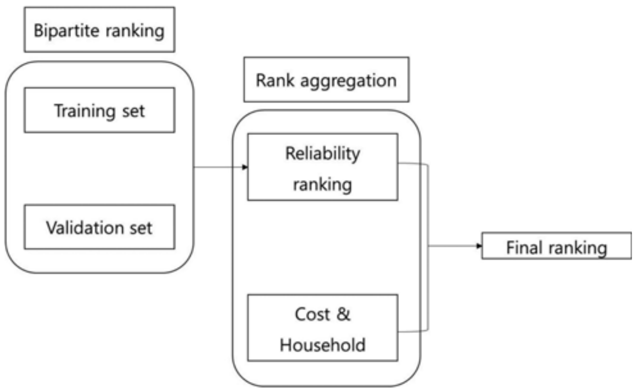

Fig. 1. The overall algorithm of rank aggregation.

such historical data and determines reliability ranks, which are subsequently aggregated to produce the final ranks that prescribe replacement planning of water pipe in numerical orders. Fig. 1 shows the overall procedure of the suggested approach.

RANK ALGORITHM AND RANK AGGREGATION

1. Bipartite Rank Algorithm

Processing of large data sets often requires preference judgments rather than classifications, and the approach to preference judgement is quite different from classification. It only needs ordering to tell what is preferred, not the output with specific meaning or quantitative value allocated to each feature. Rank, a special kind of preference function, has been popularly used in the fields where user preference plays a key role, such as information retrieval, recommender system and autonomous agent. It is also gradually applied in engineering fields such as machine translation [26] and bioinformatics [27]. There exist various ranking methods based on the ranked features and outputs from ranking algorithms such as instance ranking [28], label ranking [29] and objective ranking [30].

Bipartite ranking is one type of ordering problems that occur in real world. It indicates that there are two disjoint sets of instances and all the instances in one set outrank those in the other set. In terms of identifying a pipe that has high possibility to be broken, more preference is given to broken pipes than unbroken pipes, and these two sets of data are disjoint. Hence, the bipartite ranking can be applied to selecting candidate pipes for replacement. For this bipartite ranking problem, the Rankboost B algorithm developed by Freud et al. [31] shows solid performance in practical applications [32]. This technique is one of the boosting methods that can produce highly accurate preference order by iteratively combining various weak rules showing moderate accuracies. At each round, weak ranks can be obtained by a weak learner and the weights of each instance based on whether this ranking gives correct ranks are updated. At last, the final ranking is calculated as a weighted sum of the weak rankings. The weak learner is constructed by comparing the feature score of a given instance with a threshold value, θ. This procedure and more specific information on the threshold and weak learning is available in Freud et al. [31].

2. Rank Aggregation Algorithm

Rank aggregation has its origin in attempting to reach a consensus from multiple sources of information. It has been developed in social science for fair election from Borda's election to Condorcet's criterion [33]. It started gaining attention in the fields where there is a need to integrate multiple sources of information, and has finally become a core technique of information retrieval for web search ranking and combating spams. Moreover, in bioinformatics, this technique allows for merging expression data at DNA, RNA or protein level and even across diseases or species [34,35]. As a consequence, rank aggregation approaches have evolved to see their applications in various research areas.

For aggregation, a criterion is required to indicate which aggregated rank sequence is more meaningful. It can be represented as a distance and the most widely used distances are Kendall-tau distance and spear footrule distance. The aggregation that minimizes Kendall distance is called Kemeneny optimal aggregation. However, Kemeneny optimal aggregation requires high computational capacity, and it cannot yield optimal aggregation for real world problems. Condorcet's criterion is often introduced for defeating spamming with simple rank aggregation. It is based on transitivity, which means a higher ranker in terms of Condorcet criterion preferred by more voters defeats any lower rankers.

Dwork et al. [23] classify rank aggregation into three types. The first is the positional method, which finds a final sequence based on the aggregated rank close to the average position of the elements before aggregation. This method has a major advantage of manageable computational time. It also has anonymity, neutrality, and consistency that are meaningful properties in the social science field. However, they cannot satisfy the Condorcet criterion, and it is known that weighted-distance methods cannot satisfy the Condorcet criterion. The most popular technique in positional methods is the Borda method. It finds a sequence based on the positional order that minimizes each distance from the initial ranks. The second method, footrule and scale footrule aggregation are often employed as an approximation scheme for Kemeney optimal aggregation. Diaconis et al. prove that the footrule distance between any two permutations approximates their Kendall-tau distance within a factor of two [36]. This algorithm uses Spearman's footrule distance, defined as a function of permutation σ and ϕ, $F(\sigma, \phi)=\Sigma_j|\sigma(j)-\phi(j)|$. Finally, a hybrid algorithm is suggested to combine positional and comparison based algorithms and it shows better performance than using a single method [23]. MC4 algorithm is one of four MC algorithms proposed by Dwork et al. [23], which is the most widely used algorithm among these. This algorithm is mainly used to combat search engine spamming. Hence, the MC4 algorithm tries to lower aggregate rank of instances that are highly ranked in a minority of lists. The overall process using ranking algorithm and rank aggregation method is summarized in Fig. 1. This study employs two techniques, Borda rule and hybrid algorithm, because the Borda rule is known to provide good approximation of the optimal Kemeney score.

CASE STUDY

1. Data

The proposed approach is applied to the real field data of the water supply system of *E* city in South Korea. This database includes the past history of pipe sections in the network that had been buried, repaired and broken from 1975 to 2005. For enhancing reliability and efficiency of the algorithm, the data sets were preprocessed. First, historical incident data of water pipes are measured in days. The time unit was converted into months considering the decision frequency. We also excluded attributes with missing values, reliability issues and no relevance with pipe deterioration such as damage of gauge.

In the database, there are five types of pipe material: PVC (polyvinyl chloride), PE (polyethylene), CIP (cast iron pipe), HIVP (high impact PVC), and ductile iron. The number of data points for the ductile iron pipe is much larger than the sum of the rest. Hence, this work only considers the ductile iron pipe. This removes the categorical attribute, pipe material, and makes the data set containing only numerical and binary attributes.

Moreover, the database contains only the information of breakage events and pipes themselves, which is not enough to predict pipe deterioration. To improve the accuracy of the proposed approach, we added the GIS data of E city and pH as shown in Fig. 2. We also included atmospheric temperature as a characteristic factor for deterioration. However, because E city is not big enough to have a meaningful distribution in the thermal properties of each pipe, the result using temperature was shown to be less reliable than not. Hence, we did not include the climate data in attaining the ranks.

The number of total instances was 1690 and that of the attributes was five: diameter, land development, pH, burial time and breakage time. It is noted that the breakage time may not be available for all the pipes. Nevertheless, we constructed the list using these factors because we only require the list of pipes to be replaced and repaired preferentially. The ranking algorithm can then be applied to the data sets.

2. Ranking Algorithm for the Reliability

Many previous studies for data-based modeling of water pipe deterioration simply divide the data set of a certain time period into training and validation sets. However, this can result in uneven distributions because the data set tends to have more breakage events in the latter half of the time window. Considering this situation, we partitioned the dataset into training and testing set using random sampling. The proposed approach is composed of two-step algorithms, weak learner and rankboost B, and was implemented in Matlab R2015b.

To evaluate the performance of the algorithm, a numerical performance index is needed. In bipartite ranking algorithm, the area under the receiver operating characteristic curve (AUC) is most widely used [37]. The AUC of a classifier means the probability that their predicted pairwise ranking is correct. If all the instances in one class are ranked higher than the instances from the other, the AUC would be 1. If the rank classifier cannot discriminate each instance, the AUC would be 0.5. For AUC less than 0.5, reversal of rank can show better performance. Our result shows that the AUC for the validation set was 0.84, which is within the range of reliable results. For performance comparisons, we employed the Cox model, one of the survival analyses. The survival model predicts probability h(t) for failure of a target in infinitesimal time if it survives until time t. There are many previous studies to predict failure rate of water pipes using survival analysis [4,24,38], and this technique is appropriate to compare with our result. AUC in validation set using Cox model is 0.72. It is lower than that of the proposed rank algorithm, which indicates the proposed ranking algorithm shows better performance than the Cox model.

We use reliability rank as just one of the orderings for rank aggregation. Two pipes having the same attributes except for only one are compared. For instance, pipes 19 and 114 have the same con-

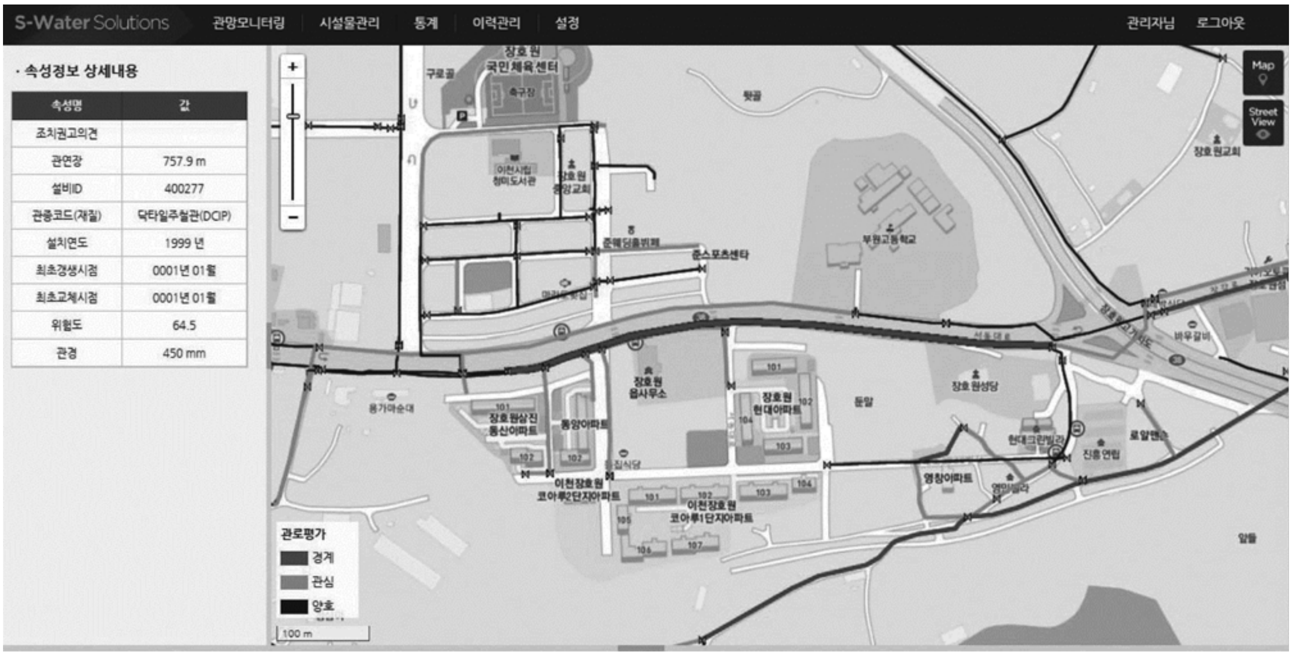

Fig. 2. GIS map of water pipes buried in E city.

Table 1. Total cost ($) for replacement of water pipes

Diameter (mm)	Construction	Materials	Destruction	Laying	Incident	Total
80	58.48	19.02	21.45	16.23	234.34	349.53
100	58.48	23.03	23.52	20.58	259.35	384.97
150	58.48	35.50	27.48	34.26	244.70.	414.76
200	58.48	47.53	29.78	36.40	300.05	457.91
300	58.48	79.11	35.80	46.50	337.39	557.29

ditions except pH. Pipe 19 with pH 5.71 is preferred to pipe 114 with pH 8.5 for replacement. For pipes 85 and 86, the rank of pipe 85 with the diameter of 150 mm was higher than that of pipe 86 with diameter of 80 mm. Such results are in accordance with the general tendency of factors that affect water mains.

3. Rank Aggregation

For deriving aggregated ranks, three factors are considered: reliability rank, repair cost, and number of households. The reliability rank is provided from the results in the previous section. In case of breakage, there are direct and indirect costs. The direct cost is based on breakdown cost provided by the Waterworks Bureau as shown in Table 1. The indirect cost involves the additional costs associated with damage and repairment. However, this information cannot be determined specifically. Instead of using direct calculations in an ambiguous manner, weights can be given depending on the land type according to Dandy et al. [13]. To exploit the three factors, we aggregate the total rank using Borda and MC4 method.

The result shows ranks reflecting each factor, but the analysis can be misleading by overestimating the number of households and breakage costs; it can provide an unreasonable result where a normal pipe is on the top of the list because of other high ranks. To address the issue, we consider two alternatives. The first method is to narrow the choice to 100 candidates that are most possible to be broken in the reliability rank and aggregate the ranks in this list only. The second method is to give more weights to the reliability rank. In view of reliability, AUC in the first method is 0.74 based on only 100 candidates. It is slightly lower than that of reliability because the number of failure data in the first method is much smaller than that of the overall test set. However, most of the data with lower reliability were removed so that reliability can be guaranteed. In the second method, reliability weights are given as 2 or 3. Results of cases with the weight more than 3 were almost identical with the case of weight 3. AUC of Borda with the weight of 2 is 0.81 and in case of weight 3, AUC is 0.82. This value is lower than that of the reliability rank but is still higher than that of the Cox and difference with the reliability rank is not significant. Hence, the aggregated ranks can be considered reliable.

More specifically, we compared the top 100 list of each aggregated rank. In case of rank aggregation using Borda method only, the list includes 63 instances that are also contained in the reliability rank. Because the first method is the same as the reliability rank, comparison is meaningless.

If the weight is 2, a total of 91 instances are the same with the two reliability ranks and aggregated rank. If the weight is 3, a total of 99 instances are the same. It means simple rank aggregation can mislead that pipes that are much less likely to be broken should be replaced. Ranks from Borda methods are similar, but the result from MC4 is considerably different. For example, the top 10 list from the Borda method is almost the same. On the other hand, only one pipe in MC4 is included in the list of the Borda method. Nevertheless, it is an interesting result because the results from each rank aggregation shows similar cost savings.

It is not easy to compare the optimal cost for each case. Hence, we compare cost savings for selecting pipes based on the reliability rank only and rank aggregation. We assume we can have data up to one year before the pipe is broken and can inspect 25% of the pipes in the test set. In South Korea, the portion of replaced pipes is lower than 0.2% of the total waterworks annually. Given this,

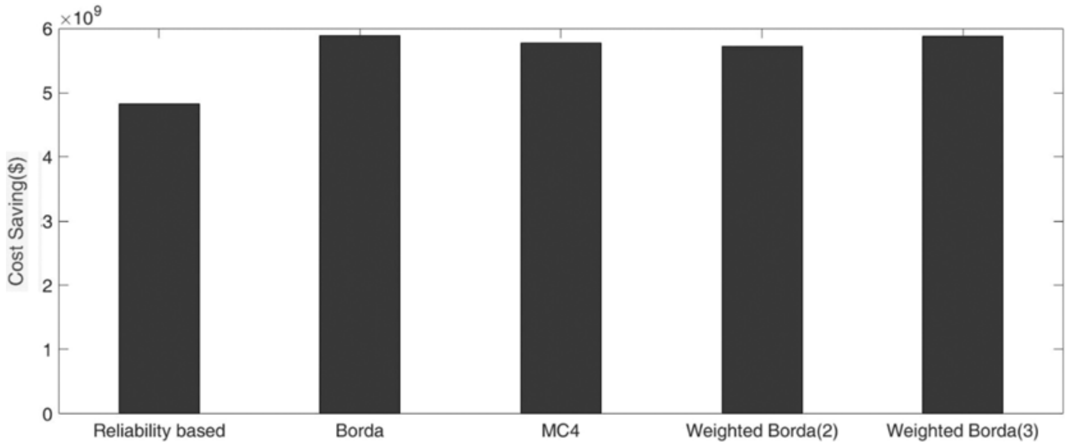

Fig. 3. Cost savings by preventing breakage of water pipes depending on each ranks.

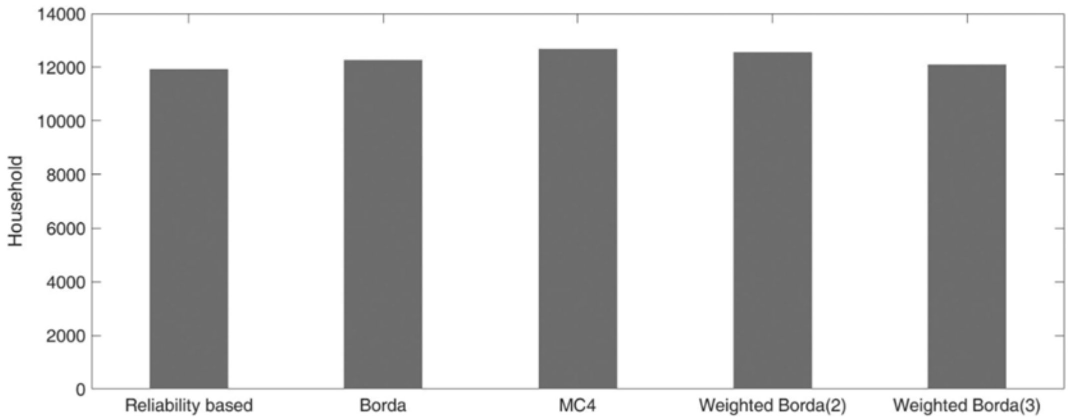

Fig. 4. The number of households that can avoid cut-off by replacing water pipes.

the number of pipes to be replaced is too small and the rank of these pipes is not sufficient enough to serve as indicators. Hence, this study assumes 25% of the total pipes are inspected. Fig. 3 shows the cost saving by the pipe inspections. Replacement policy based on reliability ranking will not replace pipes that will not be broken. On the other hand, rank aggregation-based replacement includes one or two pipes that will not be broken. However, Fig. 4 also shows that rank aggregation-based list is more profitable than the reliability ranking and shows a similar number of households. In terms of the cost, the Borda method using top list and weight of 3 is the most advantageous. In terms of the number of households without water, MC4 is the best, but the cost difference with the rank aggregation method is very small. Although the result is not from optimization and may require more number of data points for generalization, weighted Borda shows the best performance in terms of cost and the number of breakage for rank aggregation. The proposed approach performs better than the simple reliability ranking for total cost while the reliability is also considered. If we keep track of the pipe states after breakage, the proposed rank aggregation method would be even more advantageous than using simple reliability ranks.

CONCLUSION

This study suggests an overall framework to schedule pipe replacements using rank aggregation and validated its efficacy using the data from the water pipes in E city of Korea. The reliability rank showed the same performance with previous studies, indicating the prediction capability is satisfactory. For rank aggregation, general aggregation method may not be appropriate because of the differences of importance of each attribute. Thus, we propose a method that guarantees reliability of pipe deterioration.

The proposed approach can provide a preferential sequence of pipes for repair and replacement under limited budget without complicated optimization. This list also helps to meet practical necessity of the waterworks and beneficial to preventive maintenance on other types of industrial assets. It should be noted that sensitivity analysis of the parameters such as the reliability ranking can be made by removing a single parameter at a time and calculate the resulting total rank. However, the amount of data we tested in this study is not enough to perform such analysis because singling out one parameter yields irrelevant ranks in many cases. Moreover, the suggested approach can provide better results by including more attributes that may be important but are not available in this study. It is likely that additional attributes can further improve the prediction performance and help better understand the deterioration mechanism of pipe.

ACKNOWLEDGEMENTS

This work is supported by the Korea Ministry of Environment as a Project for Developing Eco-Innovation Technologies (GT-11-G-02-001-3) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2015R1A1A1A05001310).

REFERENCES

1. Korea waterworks management institute, K-water (2012).
2. Deloitte, the Ministry of the environment of Korea (2012).
3. Y. Kleiner and B. Rajani, *Urban Water*, **3**, 131 (2001).
4. S. A. Andreou, D. H. Marks and R. M. Clark, *AdWR*, **10**, 2 (1987).
5. R. Jafar, I. Shahrour and I. Juran, *MComM*, **51**, 1170 (2010).
6. H. Fares and T. Zayed, *J. Pipeline Syst. Eng. Practice*, **1**, 53 (2010).
7. R. A. Francis, S. D. Guikema and L. Henneman, *Reliab. Eng. Syst. Saf.*, **130**, 1 (2014).
8. Y. Kleiner and B. Rajani, In Proceedings, World Environmental and Water Resources Congress, 1 (2008).
9. A. Debón, A. Carrión, E. Cabrera and H. Solano, *Reliab. Eng. Syst. Saf.*, **95**, 43 (2010).
10. S. Yamijala, S. D. Guikema and K. Brumbelow, *Reliab. Eng. Syst. Saf.*, **94**, 282 (2009).
11. R. Wang, W. Dong, Y. Wang, K. Tang and X. Yao, IEEE 29th International Conference on Data Engineering, 1208 (2013).
12. Q. Xu, Q. Chen, J. Ma and K. Blanckaert, *J. Hydro-Environ. Res.*, **7**, 134 (2013).
13. G. Dandy and M. Engelhardt, *J. Water Resources Planning Manage.*, **132**, 79 (2006).
14. L. Dridi, A. Mailhot, M. Parizeau and J.-P. Villeneuve, *J. Water Resources Planning Manage.*, **135**, 344 (2009).

15. J. Nicklow, P. Reed, D. Savic, T. Dessalegne, L. Harrell, A. Chan-Hilton, M. Karamouz, B. Minsker, A. Ostfeld and A. Singh, *J. Water Resources Planning Manage.*, **136**, 412 (2009).
16. G. V. Loganathan, S. Park and H. Sherali, *J. Water Resources Planning Manage.*, **128**, 271 (2002).
17. S. Park, H. Jun, B. Kim and G. Im, *Water Resour. Manage.*, **22**, 1311 (2008).
18. P. Le Gauffre, H. Haidar, D. Poinard, K. Laffréchine, R. Baur and M. Schiatti, *Computer-Aided Civil and Infrastructure Engineering*, **22**, 478 (2007).
19. E. Roshani and Y. Filion, *J. Water Resources Planning Manage.*, **140**, 04014004 (2013).
20. N. S. Grigg, *J. Pipeline Syst. Eng. Practice*, **4**, 04013001 (2013).
21. H. T. Luong and N. N. Nagarur, *J. Water Resources Planning Manage.*, **131**, 299 (2005).
22. J. W. Kim, G. B. Choi, J. C. Suh and J. M. Lee, *Korean J. Chem. Eng.*, **33**, 25 (2016).
23. C. Dwork, R. Kumar, M. Naor and D. Sivakumar, Proceedings of the International Conference on World Wide Web, 613 (2001).
24. S. W. Park and G. Loganathan, *KSCE J. Civil Eng.*, **6**, 539 (2002).
25. R. C. Elandt-Johnson and N. L. Johnson, *Survival models and data analysis*, Wiley, New York (1999).
26. X. Tian, Wright State University (2015).
27. N. J. Risch, *Nature*, **405**, 847 (2000).
28. Y. Hu, M. Li and N. Yu, C, IEEE Conference on Computer Vision and Pattern Recognition (2008).
29. S. Vembu and T. Gärtner, in *Preference learning*, Springer, 45 (2010).
30. A. P. Wierzbicki, *Trends in Multiple Criteria Decision Analysis*, **2**, 37 (2010).
31. Y. Freund, R. Iyer, R. E. Schapire and Y. Singer, *J. Machine Learning Res.*, **4**, 933 (2003).
32. A. M. Frank, *J. Proteome Res.*, **8**, 2241 (2009).
33. M. Truchon, Cahier, 9813 (1998).
34. K. Li, N. Du and A. Zhang, IEEE International Conference on Bioinformatics and Biomedicine, 1 (2012).
35. D. Sengupta, U. Maulik and S. Bandyopadhyay, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **9**, 924 (2012).
36. P. Diaconis and R. L. Graham, Journal of the Royal Statistical Society. Series B (Methodological), 262 (1977).
37. J. Huang and C. X. Ling, *IEEE Trans. Knowledge Data Eng.*, **17**, 299 (2005).
38. Y. Le Gat and P. Eisenbeis, *Urban Water*, **2**, 173 (2000).