SIMPLE INUNDATION EVALUATION SYSTEM BASED ON MACHINE LEARNING USING INUNDATION AND RAINFALL RECORDS

Toshihiko TAKAHASHI*, Yasunori HIRAYAMA**, Reiko AMANO*,
Makoto TAKEDA***, Naoki MATSUO***
Kajima Corporation*
Kajima Information Communication Technology Co., Ltd**
Chubu University***

ABSTRACT**:** Recently, inundation analysis models have been vigorously proposed as one of the effective tools to evaluate outflow and drainage under torrential rain dynamically and precisely. As some advanced models sufficiently take characteristics of urban infrastructures, such as roads, buildings and sewer systems, into consideration, their applicability to actual urban areas has been widely verified. However, in order to conduct numerical simulations based on such advanced models, inevitably high costs and much labor are required. When enough budget is not secured, especially in daily safety management at construction sites and in the initial phase of planning buildings, one cannot afford to conduct such expensive simulations and usually has no choice but to employ a simple method to evaluate inundation. Therefore, it is highly expected to develop simple but relatively reliable inundation evaluation methods.

In this paper, in order to meet the above needs, a new simple inundation evaluation method based on the support vector machine, which is one of the most reliable machine learning methods, using inundation and rainfall records in the past as training data, is proposed. Formulation on the application of the support vector machine to the inundation evaluation is fully described. For Tokyo Metropolitan City, the system, in which the inundation evaluation method as well as visualizing tools is implemented, has been developed. It is demonstrated that the system can be readily operated in user-friendly manner and is very useful to grasp the changing trend of the inundation in Tokyo under torrential rain. By comparison of the inundation hazard predicted by this system and the actual inundation that occurred recently, the effectiveness and applicability of this system are discussed.

KEYWORDS: machine learning, support vector machine, inundation evaluation

1. INTRODUCTION

Recently, inundation analysis models have been vigorously proposed as one of the effective tools to evaluate outflow and drainage under torrential rain dynamically and precisely. As some advanced models sufficiently take characteristics of urban infrastructures, such as roads, buildings and sewer systems, into consideration, their applicability to actual urban areas has been widely examined and verified by Kawaike et al., Sekine et al., Toda et al., and so on. However, in order to conduct numerical simulations based on such advanced models, inevitably high costs and much labor are required.

From a practical point of view, when enough budget is not secured, especially in daily safety management at construction sites and in the initial phase of planning buildings, one cannot afford to conduct such expensive simulations and has no choice but to employ a simple method to evaluate inundation. Specifically, inundation hazard maps open to the public by Ministry of Land, Infrastructure, Transport and Tourism and local government offices are widely referred. Considering this current situation, it is highly expected to develop simple but relatively reliable inundation evaluation methods.

In this paper, in order to meet the above needs, a new simple inundation evaluation method is proposed, as a completely different approach to the inundation evaluation. This method is based on the support vector machine, which is one of the most reliable machine learning methods, using inundation and rainfall records in the past as training data. Formulation on the application of the support vector machine to the inundation evaluation is fully described. As applications of the support vector machine to civil engineering fields, several methods to evaluate danger of slopes and synthetic health of infrastructures have been proposed by Oishi et al. and Sugimoto et al., but its application to the inundation evaluation under torrential rain has not been seen so far. As the first step, an application of the support vector machine to the inundation evaluation in Tokyo Metropolitan City was considered, since its inundation and rainfall records in the past are totally open to public. The system, in which the inundation evaluation method based on the support vector machine as well as visualizing tools is implemented, has been developed. It is demonstrated that the system can be readily operated in user-friendly manner and is very useful to grasp the changing trend of the inundation in Tokyo under torrential rain. By comparison of the inundation hazard predicted by this system and the actual inundation that occurred recently, the effectiveness and applicability of this system are discussed.

2. MACHINE LEARNING METHODS

Machine learning is a general term for academic fields, in which architecture of learning system and learning algorithms are studied for various kinds of learning problems such as pattern recognition and game strategy. Neural networks are well known as one of the most representative machine learning methods. In 1980's, they were actively applied to various fields with the implementation of multi-layered networks and back-propagation algorithms, but some problems, such as overtraining and relatively slow convergence, have been often pointed out.

On the other hand, support vector machines are one of the learning algorithms in order to solve two-valued discrimination problems. They have been considered very reliable, according to the facts that their excellence is theoretically guaranteed and practically verified through their applications to various kinds of problems. The theory of support vector machines can be extended to nonlinear discriminant functions, but only linear support vector machines with linear discriminant functions are considered in this paper. The linear support vector machine deals with learning of two-valued discrimination problems. It is assumed that a training data set is given, in which for each input vector, its corresponding output of 1 or -1 is assigned. Under this assumption, the linear discriminant function is defined as a special type of hyperplane, the so-called optimal separating hyperplane, by which the training data can be optimally separated into two classes, each of which has only output of 1 and -1, respectively. The linear support vector machine is considered the method to construct the optimal separating hyperplane, for a given training data.

3. FORMULATION OF INUNDATION PREDICTION BY LINEAR SUPPORT VECTOR MACHINE

First, each map showing inundation areas that occurred under the events of torrential rain in the past is uniformly divided into many square grids. For each grid, it is decided if the grid is regarded as inundated or not under an appropriate criteria. As a result, the table is generated for each grid, by which it can be easily referred if it was inundated under any event of torrential rain in the past or not. Based on this table, a set of training data is constructed and its corresponding support vector machine is formulated.

It is assumed that the input vector can be reasonably specified based on rainfall data recorded at precipitation stations. The output for each input vector is 1 or -1, corresponding to the indicator that the mesh is inundated or not.

For each mesh, a set of the training data can be represented as follows:

$$S = \{ (x_1, y_1), \ldots\ldots, (x_\ell, y_\ell) \}, \quad x_i \in R^N, \quad y_i \in \{1, -1\}, \tag{1}$$

where S is the set of the training data, x_i is the input vector whose components are defined as representative values for rainfall data recorded at the precipitation stations, y_i is the output of 1 or -1 when the grid is inundated or not, ℓ is the number of the training data and N is the number of the precipitation stations appropriately selected .

Suppose the training data contained in S can be separated by a hyperplane as is shown in Fig. 3.1. Candidates of the hyperplane exist between the two black lines in Fig. 3.1. They are generally represented by the following formula:

$$w^T x + b = 0, \tag{2}$$

where $w = (w_1, w_2, \ldots\ldots, w_N) \in R^N$ is a weight vector and $b \in R$ is a constant.

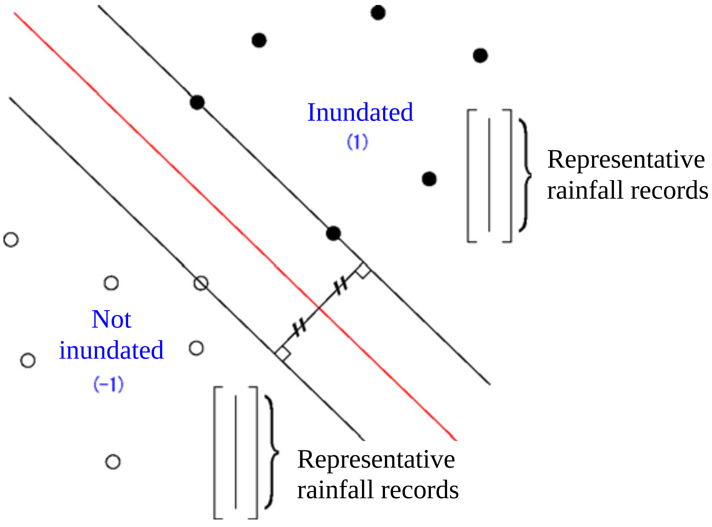

Figure 3.1 Separation of training data by hyperplane

The linear discriminant function is defined as follows:

$$f(x) = \mathrm{sgn}(w^T x + b), \tag{3}$$

where $\mathrm{sgn}(z)$ is 1 if z is non-negative, and -1 if z is negative, respectively.

The margin d_i of (x_i, y_i), which is one of the training data, for the hyperplane is computed by the following formula:

$$d_i = y_i (w^T x_i + b) / \|w\|, \quad i = 1, \ldots\ldots, \ell, \tag{4}$$

where

$$\|w\| = \sqrt{\sum_{j=1}^{N} w_j^2}. \tag{5}$$

If d_i is positive, then (x_i, y_i) is correctly discriminated and d_i is equal to the distance between x_i and the hyperplane.

When the set of the training data is linearly separable, there exists a hyperplane such that

$$y_i (w^T x_i + b) > 0, \quad i = 1, \ldots\ldots, \ell, \tag{6}$$

so the hyperplane can be standardized as follows:

$$\min_{1 \le i \le \ell} \ y_i (w^T x_i + b) = 1, \tag{7}$$

with which the margin of the hyperplane is equal to $1/\|w\|$. As it is considered that an optimal hyperplane can be obtained when the margin attains its maximum value, the optimal hyperplane is characterized as a solution of the following minimization problem:, which corresponds to the red line in Fig. 3.1.

$$\min_{w,b} \ \frac{1}{2}\|w\|^2,$$
$$w \in R^N, b \in R \tag{8}$$
$$\text{subject to} \quad y_i(w^T x_i + b) \geq 1,$$
$$i = 1, \ldots, \ell.$$

The minimization problem (8) is one of convex quadratic programming problems and can be solved directly. However, as it has to be convenient to deal with its dual problem, the following Wolfe's dual problem is considered here:

$$\max \ L(w,b,\alpha),$$
$$w \in R^N, \ b \in R, \ \alpha \in R^\ell \tag{9}$$
$$\text{subject to} \quad \nabla_w L(w,b,\alpha) = 0,$$
$$\nabla_b L(w,b,\alpha) = 0, \quad \alpha \geq 0,$$

where α is a dual variable and $L(w,b,\alpha)$ is the Lagrangian function defined as follows:

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{\ell} \alpha_i (y_i(w^T x_i + b) - 1). \tag{10}$$

From the classical Karush-Kuhn-Tucker theorem of necessary and sufficient conditions for the optimality for the convex programming problems, the dual problem (9) is reduced to the following problem:

$$\max_{\alpha} \ -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{\ell} \alpha_i,$$
$$\alpha \in R^\ell \tag{11}$$
$$\text{subject to} \quad \sum_{i=1}^{\ell} \alpha_i y_i = 0, \quad \alpha_i \geq 0,$$
$$i = 1, \ldots, \ell.$$

The dual problem (11) is one of non-concave quadratic programming problems, which tends to be numerically unstable. Therefore, more numerically stable formulation is considered, with introduction of the soft margin in case of the situation in which the training data might be linearly inseparable. By introducing slack variables into the constraints in the minimization problem (8), they can be converted as follows with slight relaxation:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0,$$
$$i = 1, \ldots, \ell. \tag{12}$$

As it has to be desirable that the values of slack variables are kept as small as possible, their penalty terms are added to the objective function in the minimization problem (8). In order to make the formulated minimization problem one of strictly convex quadratic programming problems, the following is considered:

$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{\ell} \xi_i^2,$$
$$w \in R^N, \quad b \in R, \quad \xi \in R^\ell \tag{13}$$
$$\text{subject to} \quad y_i(w^T x_i + b) \geq 1 - \xi_i,$$
$$i = 1, \ldots, \ell.$$

Applying the classical Karush-Kuhn-Tucker theorem above mentioned to (13), the following dual problem can be finally obtained:

$$\max_{\alpha} \ -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \alpha_i \alpha_j y_i y_j (x_i^T x_j + \frac{1}{C}\delta_{ij}) + \sum_{i=1}^{\ell} \alpha_i,$$
$$\alpha \in R^\ell \tag{14}$$
$$\text{subject to} \quad \sum_{i=1}^{\ell} \alpha_i y_i = 0, \quad \alpha_i \geq 0,$$
$$i = 1, \ldots, \ell,$$

where δ_{ij} is the Kronecker's delta, which is equal to 1 if $i = j$ and 0 otherwise, respectively.

The dual problem (14) is one of the strictly convex quadratic programming problems and can be solved numerically stably. Many codes for solving strictly convex quadratic programming problems have been developed, some of which have much accuracy and reliability. By employing one of them properly, an optimal solution α^* is readily obtained, with which the optimal solutions of the original minimization problem, w^*, b^*, and ξ^* are obtained.

Finally, by substituting w^* and b^* into (3), the following linear discriminant function,

$$f(x) = \text{sgn}((w^*)^T x + b^*), \tag{15}$$

is obtained. It can be promptly evaluated if the grid is inundated or not for any x, one of the representative values for supposed rain in the future.

4. SIMPLE INUNDATION EVALUATION SYSTEM

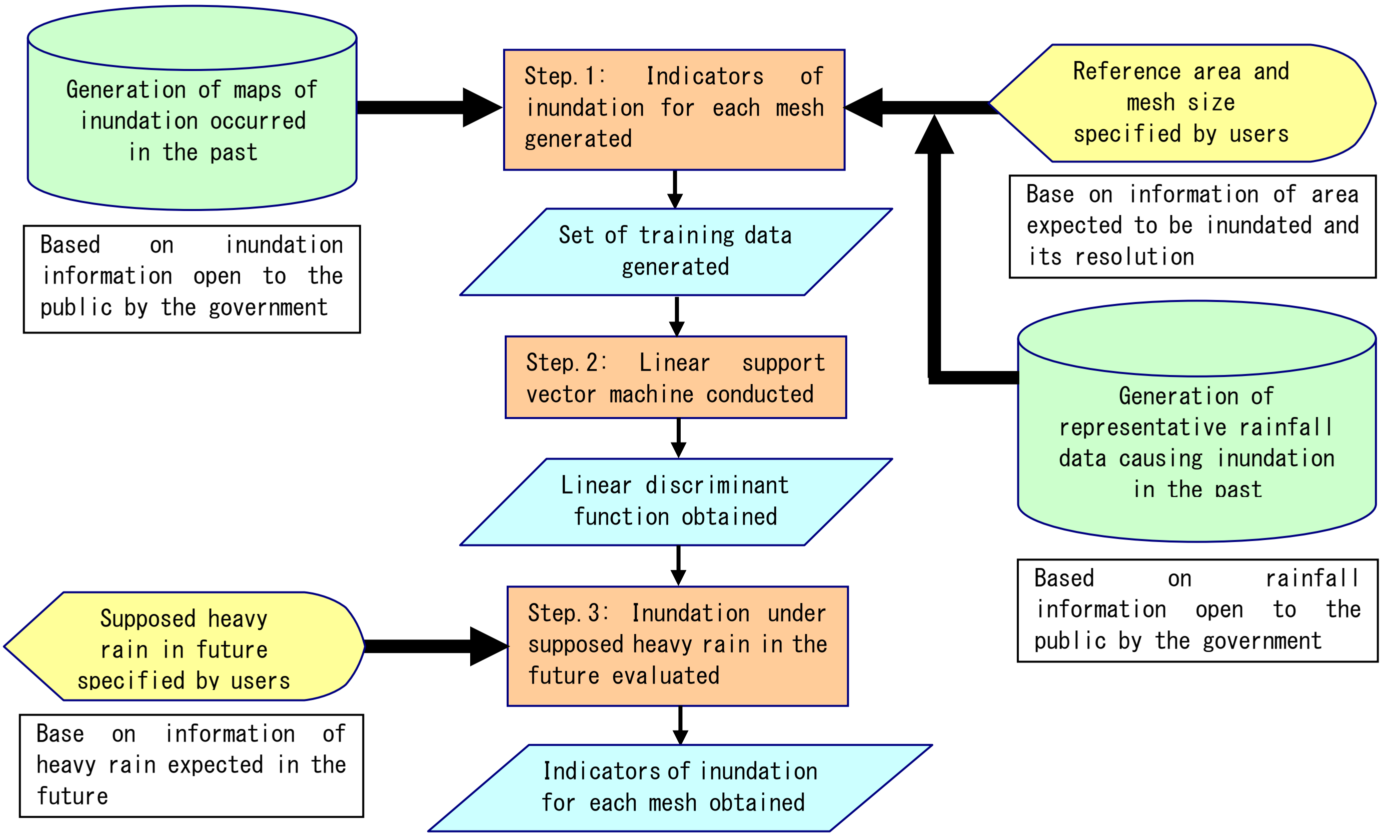

Figure 4.1 Contents of system

Based on the above-mentioned formulation, the simple inundation evaluation system, by which it is easily evaluated if a reference area is inundated or not under any supposed heavy rain, was developed on personal computers. In order to solve the dual problem (14) accurately, the dual method proposed by Goldfarb and Idnani, one of the most reliable and stable methods, was implemented. Fig. 4.1 shows the contents included in the system.

First, inundation maps and their corresponding rainfall records that occurred in the past were collected. Consequently, GIS information for the inundation maps and database for the rainfall records were generated, which corresponds to the green parts in Fig. 4.1. The hourly-maximum precipitation for each precipitation station is employed as a representative value of the input, since it affects the inundation most and its data are readily available at the precipitation stations.

For a given reference area, the area is divided into many square meshes with a specified mesh size. For each mesh, pairs of the input vector and the indicator showing if it was inundated or not in the past are constructed in the Step.1 in Fig. 4.1. As they are exactly the elements in the set of the training data, the corresponding linear support vector machine is conducted in the Step.2 in Fig. 4.1. As the linear discriminant function is eventually obtained, it is easily evaluated if the corresponding mesh is inundated or not for any values of hourly-maximum precipitation at the precipitation stations under any supposed heavy rain in the Step.3 in Fig. 4.1.

As a result, whenever inundation maps and their corresponding rainfall records are available, users can easily evaluate if their interest area is inundated or not, with the simple inundation evaluation system, by specifying only any mesh size and supposed heavy rain in the yellow parts in Fig. 4.1.

It is noted that reliability of the linear discriminant function obtained by the linear support vector machine heavily depends on the dimension of the input vector and the number of the training data. It is empirically known that stable learning can be achieved when the number of the training data ranges in orders from ten to hundred times of the dimension of the input vector. In the application of the linear support vector machine to real problems, it is very important to check how well the inundation that occurred in the past can be reproduced by the obtained linear discriminant function.

5. APPLICATION OF THE SYSTEM TO TOKYO METROPOLITAN CITY

As inundation and rainfall records in the past in Tokyo Metropolitan City are totally open to the public by the Government, the simple inundation system using them was constructed. All the detailed data from 1974 to 2009 are open to the public, but locations and combinations of precipitation stations had been updated during this period. As the number of the precipitation stations must be the same for all the events of the inundation in the past in order to generate its training data properly, 31 precipitation stations were chosen from 1995 to 2009. As the total number of the inundation and its corresponding rainfall recorded during this period is 82, the number of the training data is 82.

5.1 VALIDITY OF LINEAR SUPPORT VECTOR MACHINE

Fig. 5.1 shows the map of the inundation under the torrential rain on September 4th in 2005, called the Suginami Heavy Rain, with its corresponding meshes. The blue polygons are exactly the same as the original inundation information open to the public, and the red squares mean the indicators judged as inundated mesh by mesh, respectively. In Fig. 5.1, the mesh size is 50m and each mesh is judged as inundated if its area is included in the blue polygon more than 10 percents.

Using 82 of the training data above-mentioned, the linear support vector machine was conducted and its corresponding linear discriminant function was obtained. Fig. 5.2 shows the inundation map reproduced by substituting the values of the hourly-maximum precipitation in the Suginami Heavy Rain to the linear discriminant function. It has to be amazing that its inundation is precisely reproduced without errors, as the Suginami Heavy Rain is one of the training data. It is also verified that this result can be achieved independent of the mesh size and the judgment criterion for the inundation.

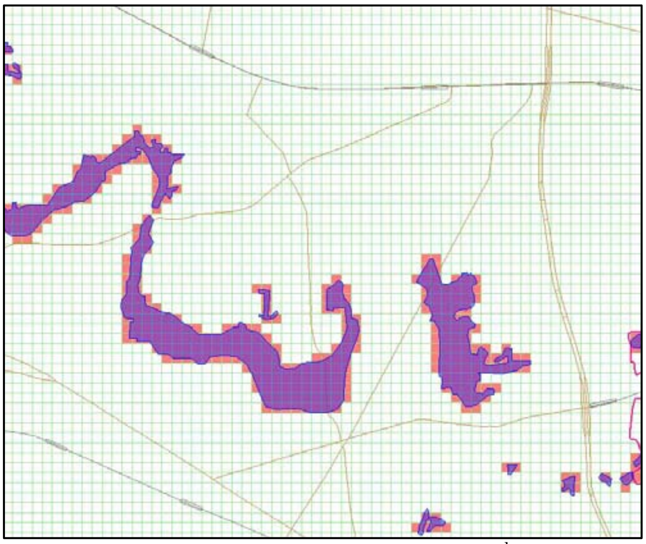

Figure 5.1 Map of inundation on Sep. 4th in 2005

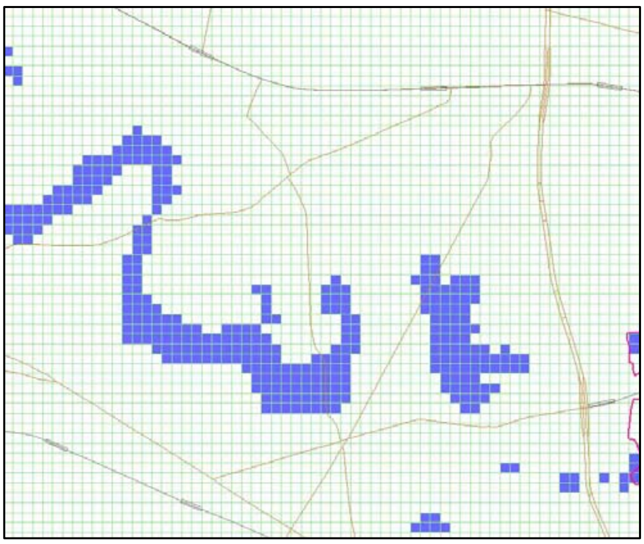

Figure 5.2 Reproduction by linear SVM

5.2 INUNDATION EVALUATION BY LINEAR SUPPORT VECTOR MACHINE

In order to discuss the applicability of the simple inundation system with the linear support vector machine, the comparison of its results with the inundation hazard map open to the public by the local government was conducted.

Fig. 5.3 shows the inundation hazard map of Nogata area in Nakano City. The yellow, green, light blue and dark blue mean that its corresponding inundation water depth ranges from 0.2m to 0.5m, 0.5m to 1.0m, 1.0m to 2.0m and 2.0m to 5.0m, respectively. This inundation hazard map was produced by the numerical simulation under the assumption that the Tokai Heavy Rain occurs on the whole area of Nakano City. The Tokai Heavy Rain occurred in September in 2000, with the whole precipitation of 589mm and the hourly-maximum precipitation of 114mm, and caused disastrous damages in Nagoya Prefecture.

In order to make the assumptions as similar as possible, the hourly-maximum precipitation of 114mm is assigned at all the 31 precipitation stations, based on the record of the Tokai Heavy Rain. Fig. 5.4 shows the results by the simple inundation evaluation system. Similarly, the mesh size is 50m and each mesh is judged as inundated if its area is included in the blue polygon more than 10 percents.

By comparison of Figs. 5.3 and 5.4, it is observed that the tendency of the inundation shows good agreement. It can be detected that some areas of the yellow and green in Fig. 5.3 are not properly evaluated by the simple inundation evaluation system, but its reason is that there are no records that these areas were inundated in the training data between 1995 and 2009.

Locally torrential rain occurred in the western and northern areas in Tokyo Metropolitan City and the western area in Saitama Prefecture on July 5th in 2010, and caused disastrous damages, especially in Kita and Nerima Cities. In Horifune in Kita City, Shakujii River overflowed its banks and some roads were severely flooded. The simple inundation evaluation system was applied to this inundation in order to check how well its actual hazard can be reproduced.

Figure 5.3 Inundation hazard map in Nakano City

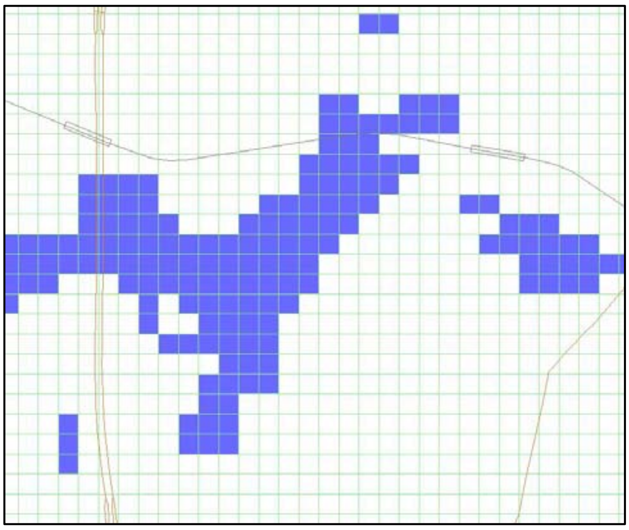
Figure 5.4 Inundation map evaluated by the system

Figs. 5.5 and 5.6 show the maps of the inundation caused by heavy rain during 1974 to 2009 and 1995 to 2009, respectively. The degree of dark colors means how frequently the inundation occurred. By comparison of Figs. 5.5 and 5.6, the inundation has much less occurred in the latter than in the former. It is inferred that countermeasures against inundation constructed between 1974 and 1994, such as reservoirs and pumps, have remarkable effects.

As Fig. 5.6 shows, the inundation near Horifune in Kita City occurred only once under the torrential rain on October 13th in 2003 during 15 years between 1995 and 2009. In the whole area in Kita City, seven times of the inundation were recorded, but none of them have caused severe damages, except the one on October 13th in 2003.

Fig. 5.7 shows the inundation map evaluated by the simple inundation evaluation system with the linear support vector machine by employing the inundation and rainfall records between 1995 and 2009 as the training data. Similarly, the mesh size is 50m and each mesh is judged as inundated if its area is included in the blue polygon more than 10 percents. In order to make the rain conditions as similar as possible, the hourly-maximum precipitation of 100mm and 50m is assigned at three precipitation stations in Kita City and at the others, respectively, based on the record of the torrential rain on July 5th in 2010

In Fig. 5.7, several meshes near Horifune are evaluated as inundated and they contain the areas actually inundated on July 5th in 2010. Considering the record that totally 567 households were inundated in Kita City, but most of them occurred near Horifune, it seems that other areas in Kita City cannot be properly evaluated. However, as the remarkable torrential rain occurred only once between 1995 and 2009 in all the training data, this result is considered quite reasonable.

6. CONCLUDING REMARKS

The new simple inundation evaluation system with the linear support vector machine using inundation and rainfall records as the training data, was proposed and developed, which is a completely different approach from hydraulic ones. By applying it to Tokyo Metropolitan City, its validity and practicability have been examined. As inundation evaluations under various kinds of rain can be conducted handily, this system has great advantages.

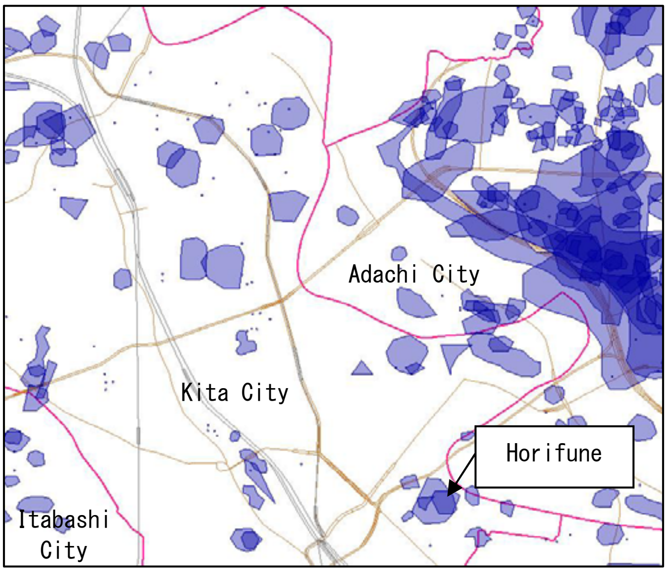

Figure 5.5 Inundation map in Kita City(1974-2009)

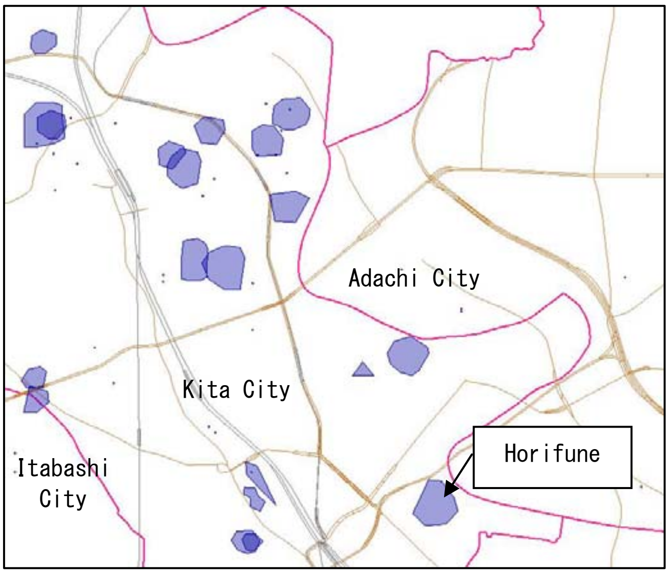

Figure 5.6 Inundation map in Kita City(1995-2009)

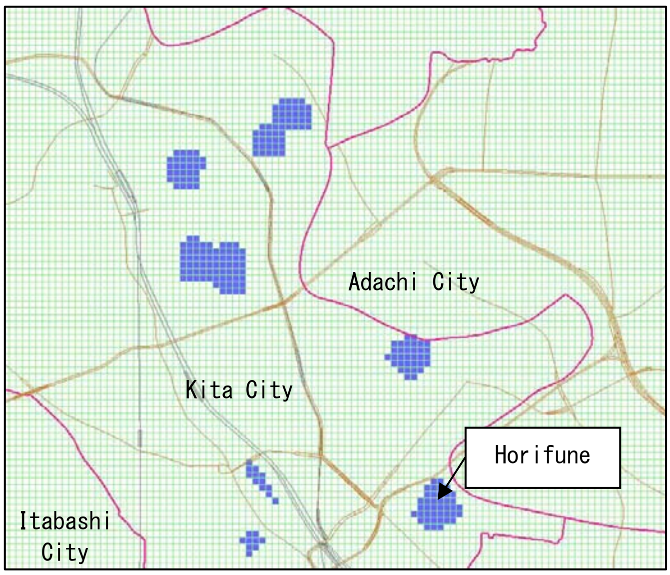

Figure 5.7 Inundation map evaluated by the system

The main findings in this paper can be summarized as follows:

1) It is verified that it can be easily evaluated if a reference area is inundated or not under a supposed heavy rain in the future. However, if the number of the training data, in which many areas are inundated, is limited, the results tend to be affected by these training data. Particularly for some meshes, in which no inundation occurred in the past, they cannot be evaluated as inundated even under unrealistic heavy rain. Although an idea that some artificial data, in which all the meshes are inundated, are added as one of the training data might be considered, it has to be an important subject to be solved.
2) The tendency of the inundation must be changing as countermeasures against inundation are implemented. Therefore, it is required that the training data have to be updated regularly. In addition, it has to be careful which elements of the set of the training data are employed, when the support vector machine is conducted.
3) Generally, it is common to employ inundation and rainfall records open to the public by the government, but it is possible that many results of numerical simulations for inundation, if available, are substituted for them.

REFERENCES

Goldfarb, D and Idnani, A., 1983. A numerically stable dual method for solving strictly convex quadratic programs, *Mathematical Programming*, Vol.27:1-33.

N. Kawaike, K. Inoue, K. Toda and M. Noguchi, 2003. Refinement of inundation flow model applied to Neya River basin, *Annual Journal of Hydraulic Engineering*, *JSCE*, 47^{th}:919-924 (in Japanese with English abstract).

H. Oishi, H. Kobayashi, Y. Yun, H. Tanaka, H. Nakayama and K. Furukawa, 2007. Evaluation of the danger of slopes in view of the effect value of measure constructions using support vector machine, *Doboku Gakkai Ronbunshuu F*, Vol.63, No.1:107-118 (in Japanese with English abstract).

N. Sekine and N. Kawakami, 2003. Numerical simulation of inundation process in urban area with underground space, *Annual Journal of Hydraulic Engineering, JSCE*, 47^{th}:889-894 (in Japanese with English abstract).

H. Sugimoto, K. Ichima, J. Abe and K. Furukawa, 2009. On synthetic health evaluation of infrastructures by SVM and its application to ranking of structures, *Doboku Gakkai Ronbunshuu A*, Vol.65, No.3:658-669 (in Japanese with English abstract).

M. Takeda, N. Matsuo, T. Yamanaka and Y. Morita, 2005. Study on the synthetic system of overland flood flow analysis, *Annual Journal of Hydraulic Engineering, JSCE*, 49^{th}:613-618 (in Japanese with English abstract).

K. Toda, K. Inoue, S. Murase, Y. Ichikawa and H. Yokoo, 2003. Numerical simulation of inundation in Kyoto City due to heavy rain, *Annual Journal of Hydraulic Engineering, JSCE*, 44^{th}:479-484 (in Japanese with English abstract).

Vapnik, V. N., 1998. *Statistical Learning Theory,* John Wiley & Sons, New York.

http://www.kensetsu.metro.tokyo.jp/suigai_kiroku/kako.htm

http://www.city.tokyo-nakano.lg.jp/dept/158000/d002981.html