예측적 공간 데이터 마이닝을 이용한 산불위험지역 예측

한 종 규† · 연 영 광†† · 지 광 훈††† · 류 근 호††††

요 약

이 논문에서는 공간적 통계기법에 근거한 예측적 공간 데이터 마이닝 방법을 제안하고, 산불위험지역을 예측하는데 적용하였다. 제안된 방법은 조건부 확률과 우도비를 이용한 방법으로 과거 산불발생지역에 대해 산불과 관련된 공간데이터 집합들 사이의 정량적 관계에 의존적인 예측 모델이다. 두 가지 방법을 이용하여 산불위험지역 예측도를 만들고, 각 모델의 예측력을 평가하기 위해 산불위험율(FHR : Forest Fire Hazard Rate)과 예측률곡선(PRC : Prediction Rate Curve)을 이용하였다. 제안된 두 가지 예측모델의 예측력 비교분석 결과, 우도비 방법이 조건부 확률 방법보다 더 우수한 것으로 나타났다. 이 논문에서 제안된 산불위험지역 예측모델을 이용하여 작성된 산불위험지역 예측도는 산불예방과 산불감시장비 및 인력의 효율적인 배치 등 산불관리의 효율성을 높이는데 많은 도움을 줄 것으로 기대된다.

Prediction of Forest Fire Hazardous Area Using Predictive Spatial Data Mining

Jong Gyu Han† · Yeon Kwang Yeon†† · Kwang Hoon Chi††† · Keun Ho Ryu††††

ABSTRACT

In this paper, we propose two predictive spatial data mining based on spatial statistics and apply for predicting the forest fire hazardous area. These are conditional probability and likelihood ratio methods. In these approaches, the prediction models and estimation procedures are depending on the basic quantitative relationships of spatial data sets relevant forest fire with respect to selected the past forest fire ignition areas. To make forest fire hazardous area prediction map using the two proposed methods and evaluate the performance of prediction power, we applied a FHR (Forest Fire Hazard Rate) and a PRC (Prediction Rate Curve) respectively. In comparison of the prediction power of the two proposed prediction model, the likelihood ratio method is more powerful than conditional probability method. The proposed model for prediction of forest fire hazardous area would be helpful to increase the efficiency of forest fire management such as prevention of forest fire occurrence and effective placement of forest fire monitoring equipment and manpower.

키워드 : 공간 데이터 마이닝(Spatial data mining), 예측모델(Prediction model), 산불위험지역(Forest fire hazardous area)

1. Introduction

Spatial data mining is one of efficient tools to discover interesting, potentially useful and high utility patterns from large spatial data sets. Widespread use of spatial databases [1, 2] is leading to an increasing interest in mining interesting and useful but implicit patterns [3, 4]. Efficient tools for extracting information from geo-spatial data generate and manage large geo-spatial data sets. The focus of this work can be of importance to the organization which own large geo-spatial data sets. Data mining products can be a useful tool in decision-making and planning just as they are currently in the business world. Knowledge extraction from geo-spatial data has also been highlighted as a key area of research in a recently [5]. The organization which make decisions based on large spatial data sets spreads across many domains including ecology and environmental management, public safety, transportation, public health and business. In this study, we focused on the application domain of forest fire prevention where forestry managers are interested in finding spatio-temporal distribution of forest fires and predicting forest fire hazardous areas from large spatial/non-spatial data sets such as forest maps, topography maps and fire history data. Forest fire provides a good example to study spatio-temporal representations for GIS applications because of its spatio-temporal variability. A key element in the forest fire hazardous area prediction modeling is a forest fire hazard model, which estimates the fire hazard potential based upon forest attributes, forest utilization, and topography.

※ 이 연구는 한국 과학재단 RRC(청주대 정보통신센터)의 연구비 지원으로 수행되었음.

† 정 회 원 : 한국지질자원연구원 국가지질자원정보센터 재직
†† 준 회 원 : 한국지질자원연구원 국가지질자원정보센터 재직
††† 정 회 원 : 한국지질자원연구원 국가지질자원정보센터 부장
†††† 종신회원 : 충북대학교 전기전자 및 컴퓨터공학부 교수

논문접수 : 2002년 5월 16일, 심사완료 : 2002년 8월 2일

In this study, a forest fire risk prediction model using predictive spatial data mining is developed to a forest in the Youngdong region of Kangwon province, Republic of Korea. We show that, by analyzing historical data on fire ignition point locations, we can gain the necessary predictive capability, making it possible to quantify ignition probability in space. The analysis is performed using inductive approaches in a raster geographic information system (GIS), and it explores the information contained in the spatial attributes of the phenomenon.

The raster GIS database used in the study contain a layer with the location of ignition events and a set of layers corresponding to potentially explanatory attributes. This spatial data set is analyzed using conditional probability and likelihood ration prediction models.

2. Related Works

Predictive data modeling predicting unknown values of certain attributes of interest based on the values of other attributes in large amount of database is a major task in data mining. Predictive data mining has wide applications, including credit evaluation, sales promotion, financial forecasting, and market trend analysis.

Statistical data analysis and inference have studied many parametric and non-parametric methods towards the prediction problems. The statistical linear regression analysis provides a means to obtain the prediction of the continuous attribute by inserting the new values of the explanatory attributes into the fitted regression equation [6]. This method is a purely parametric approach that assumes that the response attribute has a normal distribution, which sometimes could be violated. Statistical pattern recognition, neural nets, and machine learning techniques deal with the prediction of categorical attributes [7].

Bayesian approach is a probabilistic method, which is designed to yield the minimum overall error rate, the problem can be formulated in precise mathematical terms, and an optimal solution can theoretically be found by the probabilistic theory of Bayesian analysis. Unfortunately, there are serious obstacles to the direct application of this theory [7]. But there are still some successful applications derived from Bayesian analysis. [8] proposes a finite mixture model by adopting the Bayesian approach for predictive data mining. A finite mixture model is learned from instantiated attributes. The conditional predictive distribution of an attribute can be calculated from the model. [8] shows a relatively good performance of this approach by some empirical results.

Neural nets (networks) provide another type of predictive method. Unlike the normal discriminant method described above, most neural network methods are non-parametric : no assumption is made about the underlying population distribution. The back-propagation network (BPN), which is also sometimes referred to as a multilayer perceptron, is currently the most general-purpose, and commonly used neural network paradigm [9].

Decision trees are currently the most highly developed techniques for partitioning samples into a set of covering decision rules. A decision tree is a flowchart-like structure consisting of internal nodes, leaf nodes, and branches. Each internal node represents a decision, or test, on a data attribute, and each outgoing branch corresponds to a possible outcome of the test. Each leaf node represents a class. In order to classify an unlabeled data sample, the classifier tests the attribute values of the sample against the decision tree. A path is traced from the root to a leaf node, which holds the class prediction for that sample. ID3 [10, 11] and CART [12] procedures for induction of decision trees have been well established for highly effective method.

Other procedures, such as SLIQ [13] and SPRINT [14], have been developed for very large training sets. [15] proposed an efficient algorithm of decision tree induction. The algorithm has addressed not only the efficiency and scalability issues, but also the innovative multi-level classification.

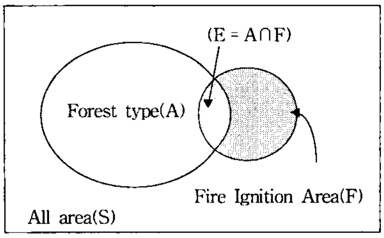
3. Proposed Prediction Model

In this study, we proposed and applied two prediction models for spatial data mining based on spatial statistics [16]. These are conditional probability and likelihood ratio methods.In these approaches, the prediction models and estimation procedures are depending on the basic quantitative relationships of spatial data sets relevant to forest fire with respect to selected the past forest fire ignition areas.

Two models used in this study are described briefly as follows :

1) Conditional probability prediction model.

The left side of (Figure 1) shows the conceptual description of conditional probability prediction model. For example, consider the problem of predicting a forest fire ignition location area in a region that covers an area of 10,000 pixels and suppose that 500 pixels covered by forest fire ignition are known within the region. The average density of known fire ignition area in the region is $N[F]/N[S]$, or 500/100,000 = 0.005, where $N[F]$ is the pixel number of forest fire igni-



N[S] = 100000 (All area)
N[A] = 2500 (Forest type A area)
N[F] = 500 (Fire ignition area)
N[E] = 100 (Area of fire ignitioin on forest type A)
N[*] : pixel number of *

• **Conditional probability**

$$CondP(F \setminus A) = P(F) \cdot \frac{P(A \setminus F)}{P(A)} = \frac{P(A \cap F)}{P(A)} = \frac{N[E]}{N[A]}$$

★ $CondP(F \setminus A)$: The conditional probability of a forest fire ignitioin given the presence of forest type A

• **Likelihood ratio**

$$LR(A \setminus F) = \frac{P(A \setminus F)}{P(A \setminus \overline{F})} = \frac{N[E] \cdot (N[S] - N[F])}{N[F] \cdot (N[A] - N[E])}$$

★ $P(A \setminus F)$: The conditional probability of a forest type A, given the presence of forest fire ignition area

★ $P(A \setminus \overline{F})$: The conditional probability of a forest type A, given the absence of forest fire ignition area

(Figure 1) Schematic diagram and explanation of the proposed prediction models

tion area and $N[S]$ is the pixel number of total region. This is the probability that 1 pixel, chosen at random (with a random generator for example) contains a known forest fire ignition area. Where no other information is available, this ratio can be used as the prior probability of a forest fire ignition area, $P(F)$. Suppose that a binary indicator map such as a forest type map and that 100 out of the 500 pixels of fire ignition area, where the forest type A is on. Clearly the probability of finding a forest fire ignition area is much greater than 0.005 if the forest type A is known to be present ; conversely the probability is less than 0.005 if the forest type A is known to be absent. The favourability for finding a forest fire ignition area given the presence of the evidence such as the forest type A, can be expressed by the conditional probability :

$$CondP(F \setminus A) = \frac{P(F \cap A)}{P(A)} \tag{1}$$

where $CondP(F \setminus A)$ is the conditional probability of a forest fire ignition area given the presence of a forest type A. But $P(A \cap F)$ is equal to the proportion of total area occupied by F and A together, or $P(A \cap F) = N[A \cap F]/N[S] = N[E]/N[S]$, and similarly $P(A) = N[A]/N[S]$, where $P(A)$ and $N[A]$ are, respectively the probability and area where pattern A is present. So that the conditional probability of a forest fire occurrence given the presence of the forest type A is 100/2500 = 0.04, 8 times larger than 0.005, the prior probability of a forest fire ignition area. Effectively, the prediction for new forest fire occurrence of the same type now becomes more focused, because if the forest type A is used as a critical indicator, the search area is reduced from 100000 pixels to 2500 pixels.

In order to obtain an expression relating the posterior probability of a forest fire occurrence in terms of the prior probability and a multiplication factor, we note that the conditional probability of being on the forest type map A, given the presence of a forest fire ignition area is defined as :

$$P(A \setminus F) = \frac{P(A \cap F)}{P(F)} \tag{2}$$

which for the present case has the value 100/500 = 0.2, Because $P(A \cap F)$ is the same as $P(F \cap A)$, Equations (1) and (2) can be combined to solve for $CondP(F \setminus A)$, satisfying the relationship.

$$CondP(F \setminus A) = P(F) \cdot \frac{P(A \setminus F)}{P(A)} \tag{3}$$

This states the conditional (posterior) probability of a forest fire occurrence, given the presence of the forest type A equals the prior probability of the forest fire ignition area $P(F)$ multiplied by the factor $P(A \setminus F)/P(A)$. The numerator of this factor is 0.2 and the denominator is 2500 / 100000 = 0.025, so the factor is 0.2 / 0.025 = 8. Thus given the presence a forest type A, the posterior probability of a forest fire occurrence is 8 times greater than the prior probability.

2) Likelihood ratio prediction model

Likelihood ratio prediction model has been applied to the problem of predicting in various disciplines. In particular, it has been applied to quantitative medical diagnosis from clinical symptoms to predict disease, e.g. [17-19]. In [20], a Likelihood ratio prediction model was used in ecological GIS application. In geology, the Prospector model originally developed in a non-spatial mode [21, 22], uses a likelihood ratio prediction model to search the potential mineral deposit area in an expert system. [23] applied the same approach to the prediction of base-metal deposits in a greenstone belt.

The likelihood ratio prediction model is described here in the context of forest fire hazardous area mapping, where the goal is to predict the forest fire ignition location. The spatial datasets used in this study, are formatted in set of pixel objects. The set of point objects (past forest fire occurrence area, forest type, digital elevation model (DEM), aspect, slope, road networks, human habitat and so on) are treated as being binary, either present or absent. In fact the model requires that each point (pixel) is treated as a small area object, occurring within a small unit cell.

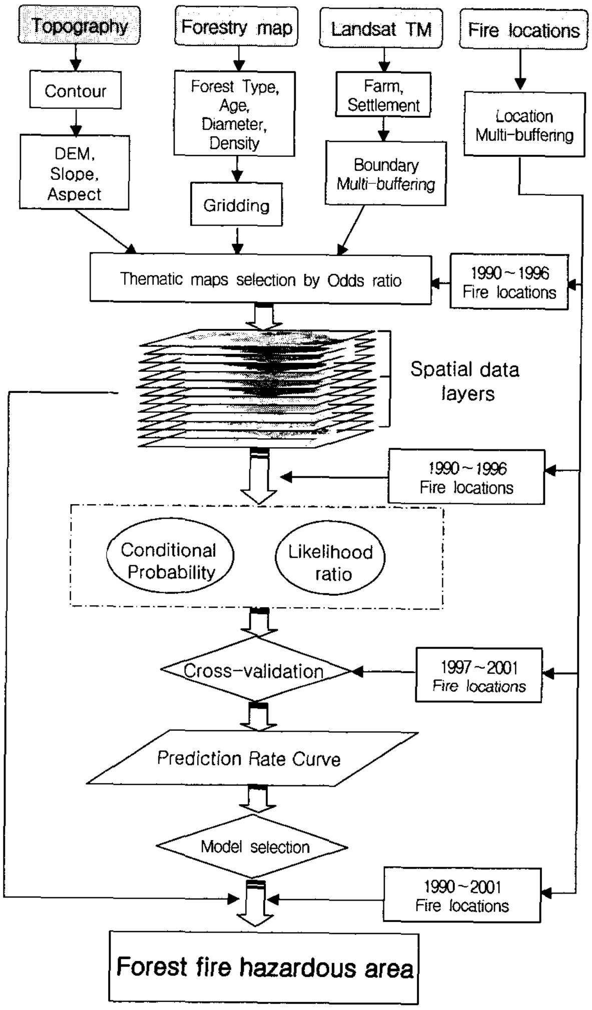
In this study, the likelihood ratio represents the ratio of the two spatial distribution functions, those of the forest fire happened and the non-happened areas. The right side of (Figure 1) shows the conceptual description of likelihood ratio prediction model. In this model, we assume that the spatial distribution functions of the forest fire happened and the non-happened areas should be distinctly different. For example, LR(A\F), the likelihood ratio to happen the forest fire at forest type A in (Figure 1) is $P(A \backslash F)/P(A \backslash \overline{F}) = N[E] \cdot (N[S]-N[F]) / N[F] \cdot (N[A]-N[E])$ and substituting number for the example leads to $P(A \backslash F) = 100 \times (100000 - 500) = 9950000$, $P(A \backslash \overline{F}) = 500 \times (2500 - 100) = 1200000$, and therefor $P(A \backslash F) / P(A \backslash \overline{F}) = 9950000 / 1200000 = 8.2916$. If the value of LR is greater than 1, then the presence of the binary pattern, forest type A, is important positive evidence for forest fire ignition. However, if the pattern is negatively correlated with the forest fire, LR would be in the range (0, 1). If the pattern is uncorrelated with the forest fire, then LR = 1. Thus for each thematic layer, two spatial distribution features for the forest fire happened and the non-happened areas are computed, firstly. Then the likelihood ratio function for the layer is computed. Using the Bayesian combination rule, the likelihood ratio functions for all data layers are combined. And then, the forest fire hazard prediction map is generated from these joint likelihood ratio functions.

This prediction model is not restricted to this case, however, and can be applied to the prediction of the environment change and natural hazard.

4. Forest Fire Hazardous Area Prediction Procedure

The procedure for making the forest fire hazardous area prediction map is as follows (Figure 2) ;

In first step, we established the spatial dataset related to the past forest fire ignition area. And we have constructed thematic maps like forestry maps (forest type, diameter, density, ages), topography maps (elevation, slope, aspect), human activities maps (road networks, farm and building boundaries) and fire history data (ignition location, year, month date, time, climate conditions, cause, burned area). These thematic maps are pre-processed for spatial data mining of forest fire hazardous area prediction. In the pre-processing step, not only gridding but also multi-buffering data processing techniques are used for spatial computation. To extract general forest fire ignition patterns such as spatial distribution features, it may be better to use elevation, slope, road netwoks and farm boundary thematic maps.



(Figure 2) Procedure for making forest fire hazardous area

In second step, we performed relevance analysis to select suitable thematic maps related to forest fire occurrences. Usually, there may be a large number of thematic maps associated with forest fire. It is neither desirable nor feasible to use all the thematic maps to do prediction. In most cases, only a few of them are highly relevant to the response attribute and are valuable for prediction.

Thus, it is necessary to perform an effective relevance analysis to filter out those attributes which are weakly relevant to the response attribute. However, the decision on whether an attribute or combination of attributes is weak or not given its relevance measure is still subjective to individual's or expert's opinions. So the purpose of our relevance analysis is to provide user an index of attribute relevance of each predictor, the choice of the predictors to be used in prediction is user's judgment.

In third step, we employed multiple layer integration method that predicts the probability of forest fire occurrence using the two proposed predictive spatial data mining methods regarding topography, forestry, fire history data and the distance to human built infrastructures. This method integrates the relevance analysis and prediction modeling at single level and enables user to perform prediction at an optimal level abstraction interactively. For flexible and efficient multiple level prediction, we have used to sharing method, which shares some intermediate analysis results with the neighborhood levels of abstraction rather than performing the prediction analysis. FHR (Forest fire Hazard Rate) is used to integrate multiple layers in this study. It is required that each layers used in the proposed prediction models are conditional independence assumption. For example, if two layers (V_1 and V_2) are conditionally independent with respect to a set of fire ignition locations, FHR is computed by multiplying the predicted values at the attribute value at the point thematic map 1 and 2. The general formula is described as follows :

FHR : Forest fire Hazard Rate

$$FHR(p)_{CondP} = CondP(V_1(p)) \times \cdots \times CondP(V_m(p))$$

$$FHR(p)_{LR} = LR(V_1(p)) \times \cdots \times LR(V_m(p))$$

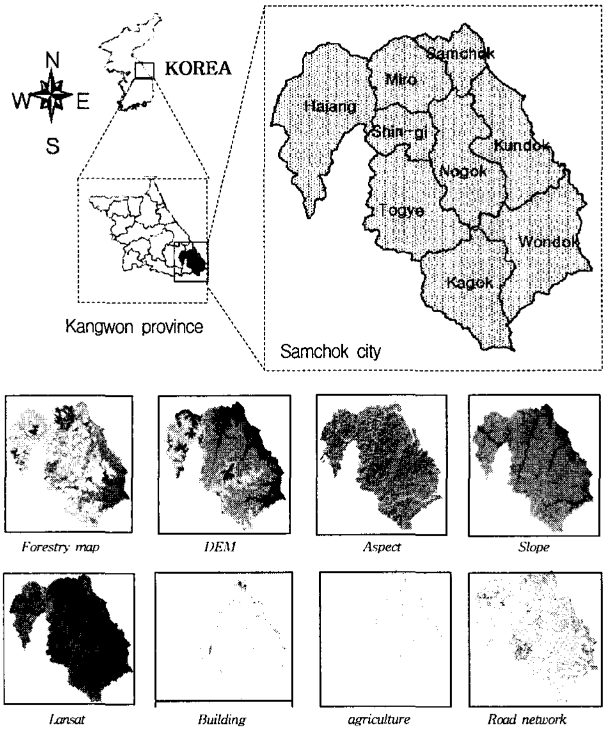
Vi(p) : *Attribute value at the point thematic map(i),* $i = 1, \cdots, m$

LR : *Likelihood radio*

CondP : *Conditional probability*

In last step, we have carried out the efficiency and effectiveness of proposed prediction modeling for forest fire hazardous area. To this step, we have used the strategy of cross validation for the prediction power using different fire history information of two forest fire occurrence periods. And, to validate of prediction accuracy, the prediction rate curve method has been applied.

5. Experiments and Results



(Figure 3) The study area and constructed spatial database for the forest fire hazardous area prediction, Samchok city of Kangwon province, Republic of Korea

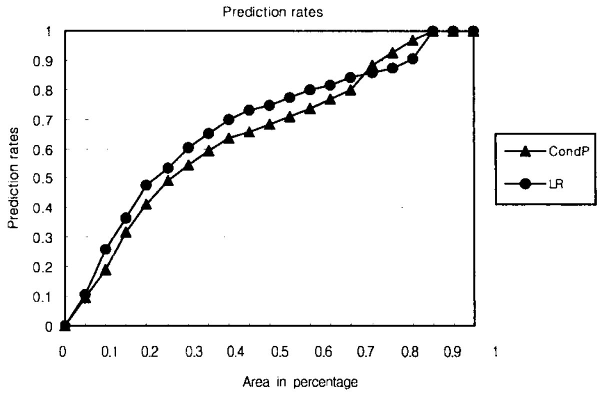
The study area to predict forest fire hazardous area is Samchok city. The largest forest fire in modern history of Korea occurred in April, 2000 in Samchok city, eastern Korea (Figure 3). Samchok city is located on the east of the Taeback Mountains, which divide Kangwon province as Youngdong region and Youngseo region, eastern part and western part of that, respectively. The locations of ignition points are marked on 1 : 25,000 scale topographic maps, and each point is associated with a field form where other information such as time of ignition, area burned, and land use are recorded. These point location data are digitized to a point layer. The geographic database sets for fire ignition prediction analysis were described in section 4. In first, an analysis of the relationship of each individual independent attribute with the response attribute was performed to get an idea of the relative importance of each attribute in explaining fire ignition. The results of the analysis of relevance of each independent attribute for the data sets corresponding to fire ignition is that forest type, elevation, slope, road networks, farm and building boundaries thematic maps have significant attributes in the data set corresponding to fire ignition. However, tree diameter, density, ages and aspect maps have the less significant attributes in the data set.

In the spatial database, it was assumed that the time of the study was the year 1996 and that all the spatial data available in 1996 were compiled, including the distribution of the forest fire ignition locations, which had occurred prior to that year. The occurrences play s pivotal role in constructing prediction models by establishing probabilistic relationships between the pre-1996 forest fires and the remainder of the input data set. The predictions based on those relationships were then evaluated by comparing the estimated hazard classes with the distribution of the forest fire ignition locations that had occurred after 1996, i.e., during the period 1997 to 2001.

To make forest fire hazardous area map using the two proposed methods and evaluate the performance of prediction power, we applied a FHR (Forest Fire Hazard Rate) and a PRC (Prediction Rate Curve) respectively. The FHR for each prediction models was calculated by formula described in section 4. For validation purpose, past fire ignition location data sets were partitioned into two independent data sets (one is the forest fire ignition data set of pre-1996 and the other one is that of during the period 1997 to 2001) for model training and validation, respectively.

(Figure 4) shows the prediction rates of the two proposed models with respect to the 1997 to 2001 forest fire occurrences. In (Figure 4), the prediction rates with respect to the "future" 1997 to 2001 occurrences of "likelihood ratio" are more powerful than "conditional probability".

For the visualization of a prediction pattern, the relative ranks of the models were applied to the pattern. To obtain the relative ranks for prediction pattern, the estimated probabilities of all pixels in the study area were sorted in descending order. Then the ordered pixel values were divided into 11 classes (colored red to blue) as follows. The pixels with the highest 5 percent estimated probability values were classified as the "100"class, shown as "red" in (Figure 5), occupying 5 percent of the study area. The pixels with the next highest 5 percent values were presented in "orange," occupy additional 5 percent of the study area, and were classified as the "95" class. We repeated the classification eight more times, for classes 5 percent apart, and the resulting ten classes are shown in the ten colors. Finally, the "blue" color was assigned to the remaining 50 percent of the area.

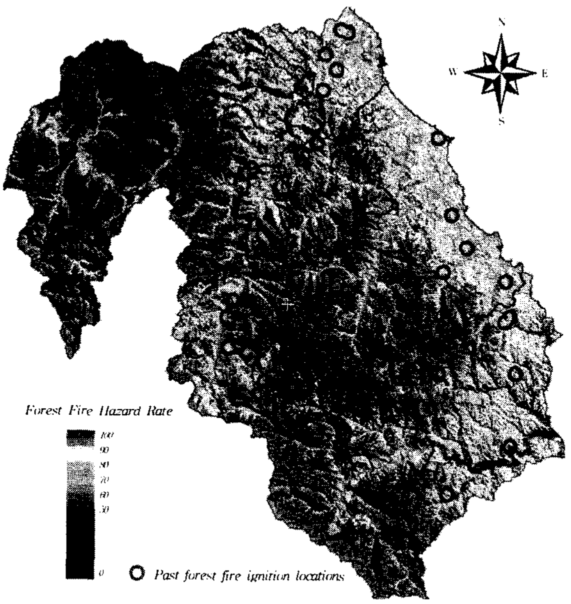


Ignition Probability Class	Prediction rates	
	CondP	LR
0~5%	0.095	0.105
6~10%	0.192	0.261
11~15%	0.317	0.368
16~20%	0.412	0.479
21~25%	0.492	0.536
26~30%	0.546	0.602
31~35%	0.594	0.649
36~40%	0.636	0.698
41~45%	0.656	0.728
46~50%	0.683	0.745
51~55%	0.708	0.773
56~60%	0.736	0.797
61~65%	0.767	0.812
66~70%	0.798	0.839
71~75%	0.884	0.859
76~80%	0.928	0.875
81~85%	0.969	0.905
86~90%	1.000	1.000
91~95%	1.000	1.000
96~100%	1.000	1.000

(Figure 4) Prediction rate curve of each models ; the curve shows that the likelihood ratio method is more powerful than conditional probability method

(Figure 5) represents the forest fire hazardous area map predicted by likelihood ratio prediction model. From the result of the prediction map (Figure 5), hazardous or non-hazardous areas can be identified. The pixels with the red, orange or yellow color in the predicted map can be interpreted as having higher likelihood of future fire occurrence. On the other hand, less hazardous areas covered by the blue pixels in the predicted map. The circles with the red color are past forest fire ignition locations.

The forest fire hazardous area prediction model described in this paper provided an effective method for estimation of the degree of forest fire hazard. It is based upon the forestry, topography, human activity attributes, which contribute to fire hazard and risk. We have found out the multiple thematic maps emerged as affective to forest fire occurrence are elevation, slope, forest type, road network, farm and habitat zone condition. The effectiveness of the models estimated and tested and showed acceptable degree of goodness. The proposed model developed would be helpful to increase the efficiency of forest fire management such as detection of forest fire occurrence and effective disposition of forest fire prevention resources.



(Figure 5) Forest fire hazardous area map predicted by likelihood ratio prediction model.

6. Conclusion

Spatial data mining is one of efficient tools to discover interesting, potentially useful and high utility patterns from large spatial data sets. In this study, we proposed and applied two prediction models for spatial data mining based on spatial statistics. These are conditional probability and likelihood ratio methods. In these approaches, the prediction models and estimation procedures are depending on the basic quantitative relationships of spatial data sets relevant forest fire with respect to selected the past forest fire ignition areas. To make forest fire hazardous area map using the two proposed methods and evaluate the performance of prediction power, we applied a FHR (Forest Fire Hazard Rate) and a PRC (Prediction Rate Curve) respectively.

Using geographic maps, forestry map and forest fire history, a spatial data mining method such as likelihood ratio and conditional probability has been developed for analyzing the forest fire hazardous area. And then, predictive power of each model has been evaluated, after carrying out cross validation between the models. Next, forest fire hazardous areas have been mapped using the most effective model.

In comparison of the prediction power of the two proposed prediction model, the likelihood ratio method is more powerful than conditional probability method.

The proposed model for prediction of forest fire hazardous area would be helpful to increase the efficiency of forest fire management such as prevention of forest fire occurrence and effective placement of forest fire monitoring equipment and manpower. The ability to quantify ignition risk can be the key to a more informed allocation of fire prevention resources. Additionally, forest fire hazardous area map, when integrated with information on fire propagation risk, can support the optimization of silvicultural practices in specific areas.

References

[1] R. H. Guting, An Introduction to Spatial Database Systems, In Very Large Data Bases Jorunal (Publisher : SpringerVerlag), October, 1994.
[2] S. Shekhar and S. Chawla, Spatial Databases : Issues, Implementation and Trends, 2001.
[3] G. Greenman. Turning a map into a cake layer of information, New York Times, Feb., 2000.
[4] K. Koperski, J. Adhikary, and J. Han, Spatial Data Mining : Progress and Challenges. In In Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD '96), 1996.
[5] D. Mark. Geographical Information Science : Critical Issues in an Emerging Cross-Disciplinary Research Domain, NSF Workshop, Feb.
[6] J. Neter and L. Wasseman, Applied Linear Statistical Models, 4th ed, Irwin.
[7] S. M. Weiss and C. A. Kulikowski, Computer Systems that Learn : Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufman, 1991.
[8] P. Kontkanen, P. Myllym.aki, and H. Tirri, Predictive data mining with finite mixtures. Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, (KDD'96).
[9] D. M. Skapura, Building Neural Networks, ACM Press, 1996.
[10] J. R. Quinlan. Induction of decision trees, Machine Learning.
[11] J. R. Quinlan, C4.5 : Programs for Machine Learning, Morgan Kaufmann, 1993.
[12] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees, Wadsworth International Group, 1984.
[13] M. Mehta, R. Agrawal, and J. Rissanen, SLIQ : A fast scalable classifier for data mining, In Proc. 1996 Int. Conference on Extending Database Technology (EDBT'96), Avignon, France, March, 1996.
[14] J. Shafer, R. Agrawal, and M. Mehta, SPRINT : A scalable parallel classifier for data mining, In Proc. 1996 Int. Conf.

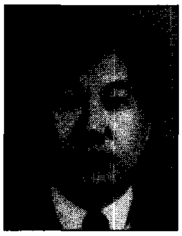
Very Large Data Bases.
[15] M. Kamber, L. Winstone, W. Gong, S. Cheng, and J. Han, Generalization and decision tree induction : Efficient classification in data mining, In Proc. of 1997. Int. Workshop on Research Issues on Data Engineering (RIDE'97).
[16] Jong Gyu Han, Yeon Kwang Yeon, Kwang Hoon Chi and Keun Ho Ryu, Prediction of Forest Fire Hazardous Area Using Predictive Spatial Data Mining, Proc. of Int. Conf. on Information and Knowledge Engineering, pp.348-358, 2002.
[17] Lusted, L. B., Introduction to Medical Decision Making : Charles Thomas, Springfield, p.271.
[18] Aspinall, P. J. and Hill, A. R., Clinical inferences and decisions-I. Diagnosis and Bayes' theorem : Opthalmic and Physiological Optics, Vol.3, pp.295-304, 1983.
[19] Spiegelhalter, D. J. and Knill-Jones, R. P., Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology : Journal of the royal Statistical Society, A, Part 1, pp.35-77, 1984.
[20] Aspinall, R. J., An inductive modeling procedure based on Bayes' theorem for analysis of pattern in spatial data : Internationl Journal of Geographical Information System, Vol.6(2), pp.105-121, 1992.
[21] Reboh, R. andReiter, J., A knowledge-based system for regional mineral resource assessment : Final report, SRI project 4119, p.267, 1983.
[22] McCammon, R. B., Prospector II The redesign of Prospector : AI system in Government, pp.27-31, March, 1989, Washington, D.C., pp.88-92, 1989.
[23] Reddy, R. K., Bonham-Carter, G. F. and Galley, A. G., Developing a geographic expert system for regional mapping of Volcanogenic Massive Sulphide (VMS) deposit potential : Nonrenewable Resources, Vol.1(2), pp.112-124, 1992.

한 종 규

e-mail : jghan@kigam.re.kr
1996년 한남대학교 전자계산공학과 졸업(공학사)
1999년 한남대학교 대학원 컴퓨터공학과 졸업(공학석사)
1995년～현재 한국지질자원연구원 국가지질자원정보센터 재직
1999년～현재 충북대학교 대학원 전자계산학과(박사과정)
관심분야 : 원격탐사, 영상처리, GIS, 공간데이터마이닝, 데이터베이스,

연 영 광

e-mail : ykyeon@kigam.re.kr
1999년 대전산업대학교 산업공학과 졸업(공학사)
2001년 충북대학교 대학원 전자계산학과 졸업(이학석사)
2001년～현재 한국지질자원연구원 국가지질자원정보센터 재직
2002년～현재 충북대학교 대학원 전자계산학과(박사과정)
관심분야 : 데이터베이스, GIS, 시간지원 데이터베이스, 영상처리



지 광 훈

e-mail : khchi@kigam.re.kr
1968년 고려대학교 지질학과 졸업(이학사)
1984년 일본 지바대학교 대학원 사진학과 졸업(공학석사)
1993년 일본 지바대학교 대학원 환경과학과 졸업(공학박사)
1976년～현재 한국지질자원연구원 국가지질자원정보센터 부장
1999년～현재 대한원격탐사학회 부회장
1999년～현재 UN CCOP 지역국가 자원GIS분야의 한국 Focal Point
2000년～현재 우주개발전문위원회 위성활용소위원회 위원
관심분야 : 원격탐사, GIS

류 근 호

e-mail : khryu@dblab.chungbuk.ac.kr
1976년 숭실대 전산과 졸업
1980년 연세대학교 산업대학원 전산전공(공학석사)
1988년 연세대 대학원 전산전공(공학박사)
1976년～1986년 육군군수지원사전산실(ROTC 장교), 한국전자통신연구소(연구원), 한국방송통신대, 전산학과(조교수) 근무
1989년～1991년 Univ. of Arizona 연구원(TempIS Project)
1986년～현재 충북대학교 전기전자 및 컴퓨터공학부 교수
관심분야 : 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS, 객체 및 지식베이스 시스템, 지식기반 정보검색 시스템, 데이터 마이닝, 데이터베이스 보안 및 Bio-Informatics