Agricultural Land Use Information Extraction in Miyajimanuma Wetland Area Based on Remote Sensing Imagery

Baofeng SU and Noboru NOGUCHI

Graduate School of Agriculture, Hokkaido University,
Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido 060-8589, Japan

(Received May 3, 2012; Accepted June 15, 2012)

The availability of agricultural land use information allows decision makers and managers to establish short-term and to long-term plans for land conservation and sustainable use. The objective of this study was to develop a method for extraction of agricultural land use information based on remote sensing imagery. By combining particle swarm optimization (PSO), k-means clustering algorithm and minimum distance classifier, a PSO-k-means-based minimum distance classifier for agricultural land use classification was developed. Crop planting information was collected and divided into five classes: water bodies, paddy fields, bean fields, wheat fields and others (windbreak, roads, rare areas, and buildings, etc.). K-means, a widely used algorithm in pattern recognition for unsupervised classification, became a part of supervised classification by using PSO to find the optimal initial position vectors in a training sample pretreatment process. The optimal cluster of each subclass was finally used for minimum distance classification. The results obtained from Miyajimanuma wetland land use information extraction showed that merely using a small feature space composed of the first three principal components of a SPOT 5 image enabled classification accuracy of 93%.

Keywords: agricultural land use classification, PSO-k-means, remote sensing, sample pre-classification

INTRODUCTION

Effective extraction of agricultural land use information could provide a method for sustainable management of land resources and policy development. Remote sensing imagery-based land use classification is an effective means for extraction of agriculture land use information, and it would help to underpin the principles of sustainable development of agriculture.

Automatic land use classification based on remote sensing data has become a powerful tool that can provide repetitive and spatial information on the landscape (Chust et al., 2004). Remote sensing imagery, which provides large-scale and up-to-date information on surface conditions, is considered to be a promising source of information for land management decisions (Song et al., 2009) and had been widely used for accurate and efficient land use classification which characterizes important information on the natural landscape and human activities on the earth’s surface (Gong et al., 2011).

It is common to observe the same object with different spectrums in the remote sensing imagery. Several factors contribute to this problem including differences in atmospheric and light

Corresponding author: Noboru Noguchi, fax: +81-011-706-4800,
e-mail: noguchi@bpe.agr.hokudai.ac.jp
conditions on the days of image acquisition and differences in soil moisture levels. In agricultural planting, plant growth status is affected by various factors including irrigation conditions, plant diseases and insect pests, and natural disasters, and even the same kind of crop would have different spectral values in the same remote sensing images. In previous studies, differences in spectral values within the same sample were always ignored. In most studies on imagery classification, more attention was paid to the feature selection and classifier design. Complex classification problems were often hoped to be resolved by complicate feature and the so-called powerful classifier. So it is easily found that edge, shape, texture, and even band ratio with the original band data as the input were used to train various classifiers, such as SVM or neural network. However, the complicated feature space increased the difficulty for training progress, and the stability of the algorithm and generalization ability became very weak. Therefore, critical major problem in classification of remote sensing data is that once classifiers are trained on a data set related to a specific image, they seldom attain acceptable classification accuracies on different images (even if the land cover classes present in the entire image are the same) (Bruzzzone et al., 1999).

The objective of this work was to develop a classification method and apply it to achieve a good classification mapping with high accuracy as well as to simplify the training progress by analyzing the difference within the training data.

The basic idea of the newly developed method is that a training sample pretreatment process is added between the feature selection and classifier training processes, and then the pre-classification results are used to do supervised classification, thereby obtaining a good classification result. The method is based on the cluster technique method, viz., k-means clustering, optimization technology and minimum distance classifier. In this method, the means of samples used to calculate the distance to each sample were instead of by the cluster centers got in the pre-classification progress. The particle swarm optimization (PSO) was used to search for the optimal initial position for k-means, thereby obtaining global optimal results. The optimal cluster of each kind of subclass obtained in the pre-classification process was used for minimum distance classification.

In this study, the newly developed PSO-k-means-based minimum distance classifier was compared with standard minimum distance and k-means-based minimum distance classifiers by using two feature spaces, original remote sensing image and principal components of the remote sensing imagery, respectively.

MATERIALS AND METHODS

Study area and data

Miyajimanuma, located in Bibai City, Hokkaido, Japan (43°20’N, 141°43’E), is a permanent freshwater lake of 30 ha in the middle basin of Ishikari River (Fig. 1. (a)). Human activities, mainly agricultural production activities, in the surrounding area have resulted in water quality deterioration and eutrophication was getting worse. Agricultural land use information within the catchment area has become important for determining the relationships between agricultural activities and water quality change thereby protecting the wetland.

The acquired image data was a sub scene of SPOT5 satellite images taken on May 15, 2010 in a 10-kilometer radius around Miyajimanuma. Resolution of images acquired by SPOT5 are 2.5 m in panchromatic, 10 m for multi-spectral data (green band, red band, and near IR band), and 20 m for shortwave IR band (SWIR). SWIR 20 m resolution data were first fused with multi-spectral data in this study. Atmospheric correction of satellite imagery was performed by the company. Georeferencing was carried out for the image using ground control points. To achieve the most realistic response to spatial structures of agricultural land use, the image was georeferenced to the coordinate system used in the topographic maps was the Universal Transverse Mercator (UTM) zone.
Fig. 1  SPOT 5 remote sensing image of the Miyajimunuma area and survey data.

Fig. 4  Results of classification. (a) Result of standard minimum distance classification, (b) result of k-means-based minimum distance classification and (c) result of PSO-k-means-based minimum distance classification.
54N with spheroid and datum WGS84.

To obtain the required ground-truth data for supervised classification, fieldwork involving collection of field information at 35 locations was carried out in July 16 2010. In the fieldwork, a GPS unit was used to visit each location and the location was placed into one of the five agricultural land use classes. The survey data are shown in Fig. 1. (b).

**Correlation analysis (CA) and principal component analysis (PCA)**

A satellite image contains much useful information, but noise information is also recorded by the satellite sensor due to the influence of the atmosphere and the sensor itself. Such information is not correct and cannot reflect real land information. Therefore, extraction of important information as well as reduction of the influence of interference information from the satellite image is necessary for remote sensing imagery classification. CA can be used to analyze the correlation between each band data, and PCA can make a linear transform from the original data space to another space, thereby obtaining the principal components.

The correlation between two bands is a statistical measure of the relationship between each band data. It is expressed by the correlation coefficient and is represented by a value within the range of $-1.00$ to $+1.00$. PCA is often used as a method for data compression. It allows redundant data to be compacted into fewer bands, thus reducing dimensionality of the data. There are several advantages of using PCA for imagery classification. (1) Since only the dominant principal components are used for analysis, the effect of noise in the data is alleviated. (2) Principal components are the fundamental patterns in the data. In this study, even only four band data in the original data, but the huge number of pixels will lead to a disaster of calculation; therefore, compressing data is essential.

**PSO-k-means-based minimum distance classifier**

Based on the results of analysis of diversity of spectral value within the same kind of crop in remote imagery, a cluster process was added before the progress of minimum distance classification. After sampling from the feature space, the sample data were first clustered, and then the cluster center was used for minimum distance classification.

**Minimum distance classifier**

The minimum distance classifier is used to classify unknown data to classes that minimize the distance between the sample data and the mean of the class in multi-feature space. The distance is defined as an index of similarity, so that the minimum distance is equivalent to the maximum similarity. The following distances are often used in this procedure: Euclidian distance, Normalized Euclidian distance and Mahalanobis distance.

**K-means clustering algorithm**

The k-means algorithm is a fast iterative algorithm that has been used in many center-based clustering applications (Forgy, 1965). It is a point-based clustering method that starts with the cluster centers initially placed at arbitrary positions and proceeds by moving at each step the cluster centers in order to minimize the clustering error, and then each sample was classified into a certain class. It works as follows.

After choosing $k$ cluster centers $C_1, C_2, \cdots, C_k$ randomly from space $\{x_1, x_2, \cdots, x_n\}$ points, the algorithm proceeds by alternating between two steps, an assignment step and an update step.

In the assignment step, point $x_i, i=1, 2, \cdots, n$ is assigned to cluster, $C_j, j \in \{1, 2, \cdots, k\}$ using Eq. 1.

$$f(x_i, C_j) = \min \|x_i - C_j\|, \quad i = 1, \cdots, n$$

(1)

Then, new cluster centers $C_1', C_2', \cdots, C_k'$ are computed in the update step using Eq. 2.

$$C_i' = \frac{1}{n} \sum_{x_i \in C_i} x_i$$

(2)
where \( n \) is the number of elements belonging to cluster \( C \).

If the termination criterion is satisfied, the iteration process will stop; otherwise, it continues from the assignment step, and the algorithm is deemed to have converged when the assignments no longer change.

Performance function for measuring goodness of the \( k \) clustering is the total within-cluster variance or the total mean square quantization error (MSE), Eq. 3 (Hamerly and Elkan, 2002).

\[
    f(x, C) = \sum_{i=1}^{n} \min_{j \in \{1, \ldots, k\}} \|x_i - C_j\|^2
\]

Therefore, once getting an initial position, one coincident cluster centers of the final classification will be obtained. However, \( k \)-means clustering algorithm is a local optimization strategy and it is highly sensitive to the choice of initial positions which are often termed “seeds” for the \( k \)-means algorithm (Chan et al., 2006; Kalyani and Swarup, 2011; Peña et al., 1999). Several methods for solving the cluster center initialization problem have been reported (Bagirov et al., 2011; Likas et al., 2003). In this study, \( k \)-means was combined with PSO to obtain the optimal final cluster center for minimum distance classification.

**Particle swarm optimization (PSO)**

PSO (Kennedy and Eberhart, 1995) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search space according to simple mathematical formulae over the particle’s position and velocity. Each particle’s movement is influenced by its local best known position and is the particle is also guided toward the best known positions in the search space, which are updated as better positions found by other particles. This is expected to move the swarm toward the best solutions.

Assuming that the search space is an \( n \)-dimensional vector, the steps of PSO can be summarized as follows (Jiang et al., 2007).

Step 1: Particle swarm is initialized. The initial position of particles is set using an \( n \)-dimensional vector \( X_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \) to represent the particle \( i \) of the swarm; the velocity of this particle can be represented by another \( n \)-dimensional vector \( V_i = (v_{i1}, v_{i2}, \ldots, v_{in}) \).

Step 2: Fitness is calculated. The fitness of each particle can be evaluated according to the objective function of an optimization problem.

Step 3: The position and velocity are updated. The best previously visited position of the particle \( i \) is noted as its individual best position \( P_{iab} = (p_{i1}, p_{i2}, \ldots, p_{in}) \). The position of the best individual of the whole swarm is noted as the global best position \( G_{ab} \). In each step, velocity of the particle and its new position will be assigned according to the following two equations:

\[
    v_i(t+1) = \omega \cdot v_i(t) + c_1 \cdot \text{rand}() \cdot (p_i(t) - x_i(t)) + c_2 \cdot \text{rand}() \cdot (G_{ab} - x_i(t))
\]

\[
    x_i(t+1) = x_i(t) + v_i(t+1)
\]

where \( v_i(t+1), x_i(t+1), v_i(t), x_i(t) \) and denote the new velocity and position and the old velocity and position, respectively; \( \text{rand}() \) is an independently uniformly distributed random variable generated between 0 and 1; the constants \( c_1 \) and \( c_2 \) termed as acceleration/learning factors represent the weighing of the stochastic terms that control the impact of previous velocity of the particle on its current one by pulling each particle toward \( P_{iab} \) and \( G_{ab} \) positions, respectively, the typical value of \( c_1 \) and \( c_2 \) is taken as 2.0 (Eberhart and Shi, 2000; Okamoto and Aiyoshi, 2008). \( \omega \), a linearly decreasing inertia weight, is implemented by starting at \( \omega_{\text{max}} = 0.8 \) and ending at \( \omega_{\text{max}} = 0.2 \) (Lima et al., 2011). The inertia weight \( \omega \) is calculated for each iteration process using Eq. 6 (Okamoto and Aiyoshi, 2008).
\[ \sigma = \sigma_{\text{max}} - \text{iter} \cdot \frac{\sigma_{\text{max}} - \sigma_{\text{min}}}{\text{totaliter}} \]  

where \text{iter} is the current number of iterations, and \text{totaliter} is the total number of iterations. This helps to expand the search space in the beginning so that the particles can explore new areas, implying a global search.

Step 4: Steps 2 and 3 are repeated until a user-defined stopping criterion is reached.

Step 5: The global best particle is outputted.

A schematic diagram of the PSO algorithm is shown in Fig. 2.

**PSO-k-means based minimum distance classification**

As an improved minimum distance classifier, PSO-k-means-based minimum distance classifier changed the traditional practice of minimum distance classifier which calculates the distance between the sample point to the mean of the sample set, a pre-classification progress was firstly done by using PSO-k-means, depending on the complexity of different land use type, bean field, paddy field, water body, wheat field and others land use, were firstly divided into 3, 5, 2, 2 and 5 sub-classes using PSO-k-means clustering; the clustering centers of these sub-class were used to do minimum distance classification; finally, based on the original classes, the sub-categories were merged. A flowchart is shown in Fig. 3.

As the core progress, k-means module, in this method, inherits the results of PSO module as the initial clustering centers and continues processing the optimal centers to generate the final results by minimum distance classifier. This method can be summarized as follows.

Step 1: A population of particles with small random positions, \( P_i \), and velocities, \( V_i \), of the \( i_{\text{th}} \) particle in the problem space is initialized. In this method, random position \( P_i \) is a pixel vector in which different dimension data corresponding to different band data.

Step 2: PSO parameters including learning factors \( (c_1, c_2) \) and inertia weight \( (\sigma_{\text{max}}, \sigma_{\text{min}}) \) are initialized.

Step 3: The iterative procedure is started and the iteration count \( t \) is set to 1.

Step 4: K-means clustering for training data in the population is done.

Step 5: Using \( G_{\text{best}} \) as the optimal cluster centers to calculate the Euclidian distance for mini-

![Flowchart of the particle swarm optimization](image)

**Fig. 2** Flowchart of the particle swarm optimization. (i): loop until all particles exhaust; (ii): loop until max iteration.
Fig. 3 Flowchart of the PSO-k-means-based minimum distance classification.

The minimum distance classifier, else increment the iteration count, $t = t + 1$ and loop to Step 4.

**Performance evaluation**

To evaluate the accuracy of the land use classification, reference sampling locations were chosen to encompass a full variety of land use classes across the whole study area. The performance of the PSO-k-means-based minimum distance classifier was rated by evaluating the following measure for test sets (Kalyani and Swarup, 2011).

$$\text{Accuracy} = \frac{\text{Number of samples correctly classified}}{\text{Total number of samples}} \times 100 \quad (7)$$

Due to total accuracy, **Accuracy** (%) does not reveal if error was evenly distributed between classes or if some classes were really bad and some really good. Therefore, as measures of classification accuracy, the producer’s accuracy (PA) and user’s accuracy (UA) were used for each category (Chust et al., 2004). Producer’s accuracy corresponds to error of omission (exclusion). The producer’s accuracy for land use type $j$ (PA$_j$) is the conditional probability that an area classified as category $j$ by the reference data is classified as category $j$ by the map. User’s accuracy corresponds to error of commission (inclusion). The user’s accuracy for land use type $i$ (UA$_i$) is the conditional probability that an area classified as category $i$ by the map is classified as category $i$ by the reference data.

**RESULTS AND DISCUSSION**

Following the flowchart, two kinds of feature space consisting the original data of four bands and the first three principal components of the original data of the four bands, respectively, were used in this study. As shown in Table 1 and Table 2, the correlation coefficient between the red band and near IR band of SPOT 5 is larger than 0.9, and the contribution rate of the first three principal components is up to 99%. That means the first three principal components could replace the original data of the four bands for the feature space. In this study, the first three principal components were therefore used instead of the original data of the four bands for the feature space.
B. SU AND N. NOGUCHI

Table 1  Correlation coefficients between bands.

<table>
<thead>
<tr>
<th>SPOT5</th>
<th>Green</th>
<th>Red</th>
<th>Near IR</th>
<th>SWIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>1</td>
<td>0.426361</td>
<td>0.481666</td>
<td>0.781268</td>
</tr>
<tr>
<td>Red</td>
<td>—</td>
<td>1</td>
<td>0.965033</td>
<td>0.384199</td>
</tr>
<tr>
<td>Near IR</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>0.42179</td>
</tr>
<tr>
<td>SWIR</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2  Cumulative proportion of principal components.

<table>
<thead>
<tr>
<th>PC</th>
<th>Eigen value</th>
<th>Cumulative Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1762.5791</td>
<td>85.35</td>
</tr>
<tr>
<td>2</td>
<td>200.0090</td>
<td>95.03</td>
</tr>
<tr>
<td>3</td>
<td>99.1612</td>
<td>99.84</td>
</tr>
<tr>
<td>4</td>
<td>3.3924</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Samples were extracted from the two feature spaces according to the survey data, of which 80% of each class was used to train the classifier and the others were used to test the classification accuracy.

The innovative and most significant step in this study was the sample pre-classification process using the PSO-k-means clustering algorithm. The clustering centers of sub-classes were used as “means” for the minimum distance classification. In order to evaluate the results of classification by different classifiers, data for two kinds of samples data from different feature spaces were used to train the standard minimum distance classifier, k-means-based minimum distance classifier and PSO-k-means-based minimum distance classifier. For the standard minimum distance classifier, means of five kinds of samples were first computed, and then pixels were divided into certain classes according to Eq. 3. In the case of the k-means-based minimum distance classifier, each sample was first clustered using k-means, and cluster centers of each sample were used as the “means” in the standard minimum distance classifier to calculate the distance with each pixel. The PSO-k-means-based minimum distance classifier used the PSO to search for the optimal initial position for k-means and used the optimal cluster as the “means” to calculate the distance to each pixel. The classification accuracies are shown in Table 3 and Table 4, and mapping pictures are shown in Fig. 4.

As shown in related tables and figures, it is obvious that the classification and mapping results by using different feature spaces were almost the same. The results obtained by the standard minimum distance classifier and PSO-k-means-based minimum distance classifier were exactly the same. This proves that the first three principal components can replace the original data of the four bands in this study, thereby reducing the feature space and simplifying the complicated training process. In the case of feature selection, it acted as an important issue to define the structure of the feature space. This is an open and difficult problem to define an absolutely optimal feature space due to the lack of evaluation criteria (Ng et al., 2007). An increase in the number of input features might introduce additional complexity related to the computational time and the “curse of dimensionality”, overwhelming the expected increase in classes associated with inclusion of additional features (Pacifici et al., 2009). The problem in constructing the feature space in the case of images is even more complicated. By using the proposed method in this study, optimal classification results by merely using the principal components have been obtained.

In addition, compared with using the k-means and the PSO-k-means-based minimum distance classifiers, a large number of paddy fields were incorrectly classified as water bodies when the standard minimum distance classifier was used. In other words, the results obtained by the two pre-classification-based methods were much better than the results obtained by using the standard minimum distance classifier. Through sample pre-classification, better mapping results could be easily
Table 3  Producer’s accuracy (PA) and user’s accuracy (UA) of different classification methods using original four bands data.

<table>
<thead>
<tr>
<th></th>
<th>Standard minimum distance classification</th>
<th>K-means based minimum distance classification</th>
<th>PSO-k-means based minimum distance classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UA</td>
<td>PA</td>
<td>UA</td>
</tr>
<tr>
<td>Bean</td>
<td>0.00</td>
<td>0.00</td>
<td>97.30</td>
</tr>
<tr>
<td>Paddy</td>
<td>90.00</td>
<td>40.91</td>
<td>91.95</td>
</tr>
<tr>
<td>Water Body</td>
<td>71.96</td>
<td>100.00</td>
<td>98.09</td>
</tr>
<tr>
<td>Wheat</td>
<td>36.05</td>
<td>72.09</td>
<td>52.17</td>
</tr>
<tr>
<td>Others</td>
<td>5.26</td>
<td>2.70</td>
<td>100.00</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>56.45</td>
<td>88.86</td>
<td></td>
</tr>
</tbody>
</table>

Table 4  Producer’s accuracy (PA) and user’s accuracy (UA) of different classification methods using first three principal components.

<table>
<thead>
<tr>
<th></th>
<th>Standard minimum distance classification</th>
<th>K-means based minimum distance classification</th>
<th>PSO-k-means based minimum distance classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UA</td>
<td>PA</td>
<td>UA</td>
</tr>
<tr>
<td>Bean</td>
<td>0.00</td>
<td>0.00</td>
<td>94.74</td>
</tr>
<tr>
<td>Paddy</td>
<td>90.00</td>
<td>40.91</td>
<td>91.95</td>
</tr>
<tr>
<td>Water Body</td>
<td>71.96</td>
<td>100.00</td>
<td>98.09</td>
</tr>
<tr>
<td>Wheat</td>
<td>36.05</td>
<td>72.09</td>
<td>54.55</td>
</tr>
<tr>
<td>Others</td>
<td>5.26</td>
<td>2.70</td>
<td>100.00</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>56.45</td>
<td>89.36</td>
<td></td>
</tr>
</tbody>
</table>

obtained regardless of the kind of pre-classification method (k-means or PSO-k-means) than the mapping results obtained by using the standard minimum distance classifier. In the case of standard minimum distance classifier, the mean of different classes was used to stand this class, and calculate the distance between dataset and it, thereby classifying the dataset into different classes. The diversity of the training data was not considered at all. However, the diversity of crop growth status that even the same kind of crop would have different spectral features have decided that samples of remote sensing images prepared for the minimum distance classifier would include different kinds of information, and it would be unreliable to use the mean of one class sample as the cluster center. Thus, it is necessary to perform sample pre-classification. The developed method, by a sample pre-classification process, takes full account of the diversity of the training data, enables to supplement the lack of information on training data in the standard minimum distance classifier.

In this study, k-means, an unsupervised classification algorithm, was first used in remote sensing image classification as part of a supervised classification process by the PSO. The k-means algorithm is traditionally viewed as an unsupervised clustering algorithm and has been widely used in commercial software of remote sensing digital image processing for unsupervised classification. Its basic form is not suitable for use in supervised classification. This study demonstrated that the k-means algorithm could be profitably modified to a supervised classifier algorithm by the PSO.

High classification accuracy is always pursued by researchers for land use mapping by using remotely sensed data. Efficient extraction of image information is the basic requirement for and purpose of remote sensing imagery classification, but in many pattern-recognition applications, the accuracy achieved by the classification system is far from that requested by the end user (Giacinto et al., 2000). The accuracy has always been around 85%, despite almost 30 years of experience and the quality of these maps is often judged to be too low for operational applications. In this study, PSO-k-means-based minimum distance classification was performed for agricultural land use mapping of Miyajimamanuma area with five classes, and good classification overall accuracy of up to 93% was obtained, and the producer’s accuracy beside wheat filed class are over 90%. And the results
were stable as well as highly accurate, and they were comparable to or even better than results obtained in previous studies using automated methods for land use classification.

CONCLUSIONS

A PSO-k-means-based sample pre-classification method was developed for the minimum distance classification with the purpose of studying spatial distribution of different farmlands around Miyajimunuma wetland area. This method can provide better mapping image only using first three principal components, largely reduce the computational complexity. In this method, the k-means clustering algorithm, modified by PSO, was first used as a supervised classification part for pre-treatment to achieve high accuracy for agricultural land use mapping. Classification and mapping results in this study indicated that pixels can be classified into different classes with a high level of accuracy. Therefore, the proposed method could be of interest to researchers involved in environment protection and precision agriculture, since the method provides a good way to obtain an explicit view in a small region and reliable information on agricultural land use.

REFERENCES

Hamerly, G., Elkan, C. 2002. Alternatives to the k-means algorithm that find better clusterings. Paper presented at the proceedings of the eleventh international conference on information and knowledge management, McLean, Virginia, USA.

