Computers & Industrial Engineering 128 (2019) 60–69

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning

Xuehong Gao, Gyu M. Lee*

Department of Industrial Engineering, Pusan National University, Busan, Republic of Korea

ARTICLE INFO

Keywords:
Back propagation network
Bicycle-sharing system
Fuzzy C-means
Hybrid genetic algorithm
Machine learning

ABSTRACT

Bicycle-sharing systems are a new type of transportation service that provides bicycles for shared use; they allow users to rent a bicycle at one station, ride it, and return it to another station in the same city. A predictive model is needed to forecast the rental demand to improve user satisfaction and increase profits. To effectively predict the rental demand in such bicycle-sharing systems, we propose a moment-based model and a new hybrid approach that combines a fuzzy C-means (FCM)-based genetic algorithm (GA) with a back propagation network (BPN). This FCM-based GA is a new unsupervised classification method that is used to pre-classify historical rental records into groups. The classification results are then fed into a BPN predictor, which is trained using these categorized records. After training, the BPN predictor can predict the demand at future moments. Finally, we present a case study based on real-life data to demonstrate the effectiveness and efficiency of the proposed approach.

1. Introduction

Bicycle-sharing systems have grown rapidly in the past few decades (Fishman, 2016). They are a new type of transportation service that provides bicycles for shared use. Such systems have a set of bicycle stations, and users are allowed to rent a bicycle from one station, ride it, and then return it to another station. The first such system began operating in Amsterdam, the Netherlands in 1965 (Shaheen, Guzman, & Zhang, 2010). This system was run on a non-profit basis with a focus on the social and environmental issues. Stimulated by the development of information technology (IT) in the 2000s, this system was employed worldwide. Compared to the previous situation, it is now easier and more convenient for people to rent a bicycle because cellphone-mapping apps show the neighboring stations and allow people to find stations that are available with bicycles or open return slots. Bicycle-sharing systems have so far been launched in more than 712 cities in more than 50 countries (Shaheen, Martin, Cohen, Chan, & Pogodzinski, 2014), such as London, Beijing, and Paris. Indeed, they have become an important part of urban transportation systems because they play a significant role in alleviating traffic, environmental, and health issues.

More advanced bicycle-sharing systems, such as OFO sharing bicycles and Mo-bike in Beijing, enable users to track the locations of unused bicycles and rent them. Bicycles can be parked and locked at any public station in the city after use. To increase the number of bicycles available for rent and improve profits, it has been suggested that operators should deploy a fleet of trucks to collect bicycles in the evening and redistribute them to the main stations (Schuijbroek, Hampshire, & Van Hoeve, 2017) as well as continuously (i.e., dynamically) reallocate them during the day. Intelligent management of bicycle-sharing systems involves several issues associated with policy design (Pfrommer, Warrington, Schildbach, & Morari, 2014), intelligent bicycle redistribution (Contardo, Morency, & Rousseau, 2012), and user journey planning (Gast, Massonnet, Reijsbergen, & Tribastone, 2015). In addition to interesting real-world applications of bicycle-sharing systems, the characteristics of the data they generate make them an attractive research topic (Fanaee-T & Gama, 2014).

This study focuses on improving the reallocation efficiency. The main problem is to determine suitable inventory levels at the main bicycle stations, which typically involves data mining and machine leaning analysis of historical data. The ability to cope with fluctuations in bicycle demand is crucial to the success of such systems. To achieve this, each station's inventory must be reviewed regularly (Raviv & Kolka, 2013), and it is important for both the users and system managers to have a model for predicting bicycle rentals. This increases user

* Corresponding author.
E-mail address: glee@pnu.edu (G.M. Lee).

https://doi.org/10.1016/j.cie.2018.12.023

Available online 08 December 2018

satisfaction by allocating more bicycles to areas of high population density and allows system managers to assess the required bicycle inventory level, improving bicycle utilization and increasing profits.

The number of bicycle rentals is strongly affected by the weather (environment), calendar (season), and limited-time events. These have a cumulative effect, making demand difficult to predict. In addition, their impact on the final number of rental bicycles needed changes over time. However, we believe that at least some of these important events can be detected via mining and analyzing historical datasets showing the rental rates over time. In this paper, we present a novel moment-based model for predicting rental demand that can forecast the number of rentals.

The remainder of this paper is organized as follows. Section 2 reviews the previous studies on bicycle-sharing systems and prediction methods. Section 3 describes in detail the proposed fuzzy C-means (FCM)-based genetic algorithm (GA) and back propagation network (BPN) methods for predicting bicycle rental demand. Section 4 tests the proposed method using actual data obtained from Washington D.C., USA and discusses the results. Finally, Section 5 concludes this study, outlining its contributions and possible directions for future work.

2. Literature review

In recent years, bicycle-sharing systems have been a topic of active study. Several studies have investigated their history and interesting features. For example, DeMaio (2009) discussed their history, covering the first- to third-generation programs, and shared insights into future fourth-generation programs. Fishman (2016) reviewed the research into a range of bicycle-sharing topics, including its documented history and growth, usage patterns, user preferences, and demographics, noting that some researchers had analyzed the spatial and temporal relationships in bicycle-sharing data collected via automated devices. Here, we identify and review several main research areas in the bicycle-sharing literature: rebalancing operations, demand prediction, and main prediction methods.

With the rapid growth of bicycle-sharing systems, many studies have proposed methods for optimizing their rebalancing operations. Contardo et al. (2012) presented a dynamic public bicycle-sharing balancing problem based on the daily operations of such systems, while Raviv and Kolka (2013) introduced an inventory model suited to managing rental bicycle stations and a numerical method for solving it. Erdoğan, Battarra, and Calvo (2015) presented an exact algorithm for efficient rebalancing, which aimed to achieve target inventory for the number of bicycles in each station. This featured a new cluster-first, route-second heuristic that allowed a polynomially sized clustering problem to simultaneously consider the feasibility of the pre-determined service level and the approximate routing cost.

In order to handle the rebalancing problem, several studies have investigated whether predicting the rental demand could result in better rebalancing strategies. These studies have presented approaches to optimizing the bicycle station locations, and methods of rebalancing the inventory and routing the trucks to redistribute bicycles in light of spatial and temporal variations in demand. Vogel and Mattfeld (2011) used data recorded by bicycle-sharing systems to forecast the medium-term demand, aiming to support and improve strategic and operational planning via data mining. García-Palomares, Gutiérrez, and Latorre (2012) proposed a geographical information system (GIS)-based method of predicting the potential demand as a first step toward optimizing the station locations for Madrid's bicycle-sharing program, using population and employment data to estimate the spatial demand distribution. Kadri, Kacem, and Labadi (2016) pointed out that the goal of bicycle-sharing systems must be to satisfy user demand while reducing operational costs, emphasizing the importance of predicting future user demand and bicycle availability based on data analysis and statistics. Schuijbroek et al. (2017) combined two aspects of the problem that had previously been handled separately: determining each station's service level requirements and designing optimal vehicle routes to rebalance the bicycle inventory.

Predicting the bicycle demand is an important part of any rebalancing strategy, so many studies have focused on this issue. Barnes and Krizek (2005) developed predictive models for estimating the probable range of total bicycle demand in a given geographic area based on data from commute-to-work surveys. Borgnat, Abry, Flandrin, and Rouquier (2009) proposed a statistical model for predicting the number of bicycles hired per hour, which involved several factors including the number of subscribers, the time during the week, the occurrence of holidays or strikes, and the weather (temperature, amount of rain). Kaltenbrunner, Meza, Grivolla, Codina, and Banchs (2010) analyzed human mobility data based on the numbers of bicycles available in the stations, using this to detect temporal and geographical mobility patterns within the city and predict the number of bicycles available in any given station minutes or hours in advance. dell'Olio, Ibeas, and Moura (2011) proposed a method of estimating the potential demand to determine the station locations for picking up and dropping off bicycles with the assistance of a GIS. Corcoran, Li, Rohde, Charles-Edwards, and Mateo-Babiano (2014) studied the effect of weather and calendar events on the total number of trips for all CityCycle stations in Brisbane using a Poisson regression model. Gast et al. (2015) used probabilistic predictions obtained from a theoretical time-inhomogeneous queueing model to forecast future bicycle availability in bicycle-sharing system stations.

More recently, El-Assi, Mahmoud, and Habib (2017) found that the bicycle-related infrastructure (such as cycle lanes and paths) had a crucial influence on the rental demand, also revealing significant correlations between temperature and land use and bicycle-sharing activity. Feng, Hillston, and Reijsbergen (2017) predicted the future availability of bicycles in the stations by analyzing the moment of a continuous-time Markov-chain population model with time-dependent rates. Kim (2018) investigated the different outcomes caused by different weather conditions and temporal characteristics, based on station-level and system-level analysis characteristics. In their cost-effective station-level analysis, they used a clustering method to identify groups of stations with similar properties, considering the effect of temperature and humidity by introducing a temperature–humidity index and a heatwave indicator variable. They also carried out a system-level analysis, showing that certain factors had significant effects at different times of day. In particular, high temperatures, rain, and whether or not it was a work day had different effects on the rental demand at particular times.

However, many of the factors that contribute to the rental demand cannot be accurately predicted using regression analysis (Borgnat et al., 2009; Corcoran et al., 2014). With the development of artificial neural networks (ANN), many researchers have shown that ANNs can outperform traditional prediction methods (Chen, 2003). Based on ANNs, BPNs were proposed to enlarge the range of problems that can be tackled (Russell & Norvig, 2016), and have since been used for prediction in a wide variety of areas (Chen, 2013; Kim & Lee, 1997; Mi, Yang, Li, Zhang, & Zhu, 2010; Wang, 2007). Machine learning and data mining methods have also been applied to neural network models to improve prediction accuracy (Chen, 2007; Tirkel, 2011).

Although many studies have investigated traffic flow and rental demand prediction, few have been concerned with moment-based demand in bicycle-sharing systems. In addition, the variation in bicycle-sharing system demand is very complex and affected by a range of factors. The reason the relationships are so complex is that these factors are always changing. In order to improve the quality of the input data for prediction models, the FCM method has been used to first classify the data into clusters (Chen, 2008; Rezaee, Jozmaleki, & Valipour, 2018; Sun, Guo, Karimi, Ge, & Xiong, 2015). However, this method is sensitive to noise, isolated data, and the initial clustering centers, and is liable to converge to a local extremum (Ding & Fu, 2016). To overcome these defects, our model introduces a GA. GAs are a global, parallel, stochastic search method, founded on Darwinian evolutionary principles. In recent decades, they have been applied in a variety of areas, with varying degrees of success (Gao, Zhou, Amir, Rosyidah, & Lee, 2017; Yokota, Gen, & Li, 1996; Yu & Gen, 2010).

We investigate the problem of predicting the rental demand in a bicycle-sharing system using a moment-based analysis. For a given system, the demand at a future moment can be predicted by learning from the relevant information in historical data. Here, we propose a hybrid approach that combines the FCM method with a GA to classify historical rental records into several different categories, because the FCM method allows for flexible classification and reduces the sensitiveness of records. Note that the fuzzy similarity of all the categories for the records, as defined as FCM function in following section, as the fitness function in GA. The FCM-based GA classifier's output is then fed into a BPN predictor, to train it on rental records in different categories. The BPN predictor can then predict the rental demand at future moments.

3. Method

In this section, we present our moment-based hybrid GA and BPN model and explain how it predicts the future rental demand. It considers more than ten factors (attributes) to determine the rental demand, including the date, time, weather (e.g., temperature, humidity, and wind speed) and season, all of which make different contributions to the final demand. The collected data is divided into two subsets, one for training the model and the other for testing. Before we explain the method in detail, we first introduce the notation used in the model.

3.1. Notation

The notation used in this model is as follows.

Sets	
B	All historical bicycle rental records, indexed by $i \in B$.
T	Bicycle rental records used for testing, indexed by $j \in T$.
R	Bicycle rental record attributes, indexed by $k \in R$.
Parameters	
v_{ik}	Value of attribute k of bicycle rental record i.
hd_i	Historical demand for bicycle rental record i.
w_k	Weighted value of attribute k.
C	Number of categories (clusters), indexed by $c, l \in C$.
Decision variables	
o_{ck}	Centroid of the c-th cluster for attribute k.
u_{ic}	Membership value of bicycle rental record i in the c-th cluster.
E_c	Expected bicycle rental demand for category c.

3.2. FCM clustering

The FCM clustering algorithm was developed by Dunn (1973), and later improved by Bezdek and Dunn (1975). It is used to assign patterns or data to different clusters for soft partitioning, where each data point is allowed to belong to several clusters with different degrees of membership. These membership degrees represent the extent to which each point belongs to each cluster, and are also used to update the cluster centers. FCM allows for flexible rental record classification and reduces the sensitivity to noise caused by isolated records. With this hybrid approach, the number of clusters is pre-determined and each data point is then assigned to one or more clusters.

The FCM algorithm can be seen as a fuzzy version of the k-means algorithm and is based on minimizing an objective function called the c-means function (Bezdek & Dunn, 1975; Kenesei, Balasko, & Abonyi, 2006). This takes three input parameters, namely the number of categories C, fuzziness exponent $m > 1$, and termination tolerance $\phi > 0$. Given a set of historical bicycle rental records B, including a set of record attributes R, FCM attempts to minimize the objective function J using the following steps.

(1) FCM function: The objective function J measures the fuzzy similarity of all the categories for all the historical records. Note that, at each iteration, the algorithm calculates the center of each cluster.

$$J = \sum_{c \in C} \sum_{i \in B} d_{ic} \cdot u_{ic}{}^{m}. \tag{1}$$

Here, d_{ic} is the distance from bicycle rental record i to the centroid of category c.

(2) Classification result (initial): We calculate the degree of membership u_{ic} of each bicycle rental record i in category c. Note that, at the start of the algorithm, these membership degrees u_{ic} are initialized randomly. The fuzziness coefficient $1 < m < \infty$ represents the required clustering tolerance.

$$u_{ic} = 1 \Bigg/ \sum_{l \in C} \left(\frac{d_{ic}}{d_{lc}}\right)^{2/(m-1)}. \tag{2}$$

(3) Category centroid: We calculate the centroid for each attribute k in each of category, then update the distance from each bicycle rental record i to the centroid of each category c and the membership values u_{ic}.

$$o_{ck} = \sum_{i \in B} v_{ik} \cdot u_{ic}{}^{m} \Bigg/ \sum_{i \in B} u_{ic}^{m}. \tag{3}$$

$$d_{ic} = \sum_{k \in R} w_k \cdot |v_{ik} - o_{ck}|. \tag{4}$$

(4) Termination condition: The accuracy required for objective function value J determines the number of iterations needed. We represent the accuracy at iteration n as J^n and calculate FCM termination tolerance ϕ as follows:

$$|J^{n+1} - J^n| < \phi \tag{5}$$

3.3. FCM-based GA

The FCM algorithm is sensitive to noise, isolated data, and the initial clustering centers, and is liable to converge to a local extremum (Ding & Fu, 2016), so we use a GA to improve its performance by preventing the cluster centers from falling into local extrema too easily. Many studies have shown that the combination of GA and FCM methods has obvious advantages. Ding and Fu (2016) presented a combination of GA and the kernel-based FCM improved the clustering performance. Wikaisuksakul (2014) tested many datasets and showed that his FCM-NSGA (non-dominated sorting GA) achieved the best partitioning over the other techniques. Ye and Jin (2016) proposed a clustering algorithm based on quantum GA that demonstrated a better accuracy of clustering than the general FCM clustering algorithm. In this study, we develop an FCM-based GA that combines a GA with the FCM clustering algorithm. This uses center-based string encoding, nonlinear ranking selection, and crossover and mutation strategies, which are discussed in detail below.

3.3.1. Chromosome design

Our FCM-based GA's chromosomes represent the cluster centers by encoding them as is center-based string. The R-dimensional centers of the C clusters are considered genes and are concatenated into a string, as shown in Fig. 1. As defined, R is the number of record attributes. An individual or chromosome is consisted of C cluster centers. Hence, the FCM clustering explained in 3.2 needs to be implemented whenever it tries to find C cluster centers, given attributes v_{ik} for $i \in B$ and their membership values u_{ic}.

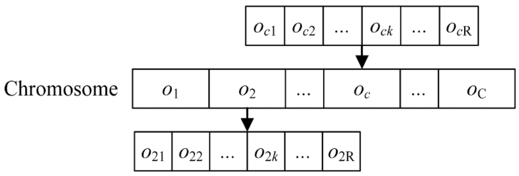

Fig. 1. String encoding for the cluster centers.

3.3.2. Fitness function

The fitness function is the measure used to judge and evaluate individuals (chromosomes) until either a maximum number of generations is reached or the average fitness has converged. In GAs, individuals with higher fitness values are considered better and are more likely to survive. In our case, we used the reciprocal of the objective function as the fitness function to evaluate each individual's fitness, as follows:

$$F = 10^6 \Big/ \left(1 + \sum_{c\in C}\sum_{i\in B} d_{ic}\cdot u_{ic}{}^m\right). \tag{6}$$

3.3.3. Genetic operators

The genetic operators that drive the search process are as follows.

(1) Selection operator

Here, we use constant ratio selection to identify the individuals that will undergo genetic operations. This selection operator is used to screen out some bad individuals to be used in the genetic operations. We create a group of randomly selected individuals at the pre-determined percentage against the population (i.e., constant ratio) and choose the best one among the group. We repeated this process as many times as the size of the population. We use this optimal preservation strategy that avoids destroying the current best individual. With this strategy, fitter individuals have higher survival probabilities, although this does not guarantee that the fittest individual will be selected.

(2) Crossover operator

First, we generate a random number τ in the interval [0,1] and compare it with the crossover probability P_c. If $\tau < P_c$, then the crossover operator is applied to two parents to generate two new children. Here, we use a single-point crossover operator, and the crossover point is selected based on a random integer $c\in[1, C]$. The crossover process is shown in Fig. 2.

(3) Mutation operator

For each individual, we generate a random value ρ in the interval [0,1] and compare it to the mutation probability P_m. If $\rho < P_m$, we mutate that individual. The mutation operator involves first generating a random integer $c\in[1, C]$ and replacing the c-th gene with a random R-dimensional center $O_c^*(o_{c1}^*, o_{c2}^*, \cdots, o_{ck}^*, \cdots, o_{cR}^*)$, as shown in Fig. 3.

3.3.4. Applying the FCM-based GA

The following steps needs to be implemented to apply the FCM-based GA.

Step 1: Set the parameters, namely the number of categories C, weighted values w_k of each attribute k, number of generations G, population size Q, crossover probability P_c, mutation probability P_m, fuzziness m, and termination tolerance ε.

Step 2: Initialize the population. Q chromosomes are generated using the FCM clustering as many times as Q. To produce a

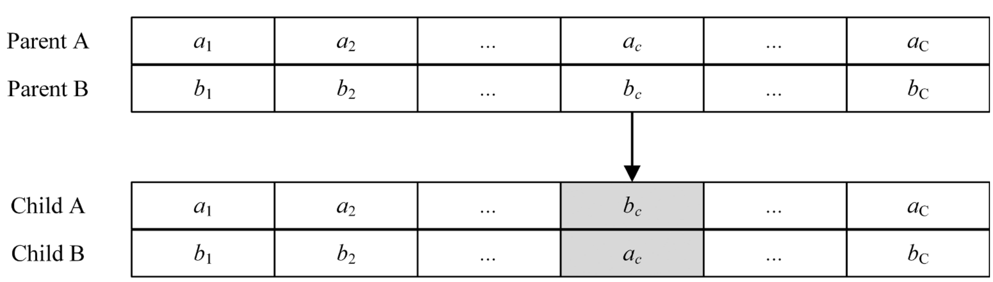

Fig. 2. Generating two children from two parents via crossover.

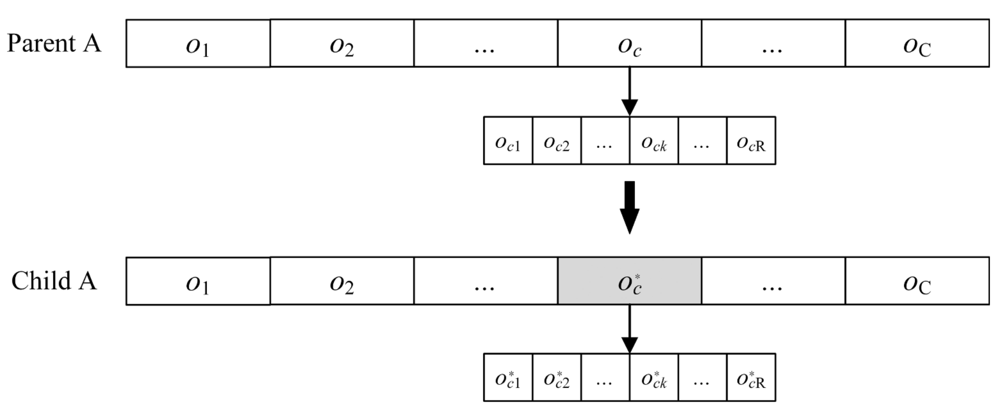

Fig. 3. Mutating one individual.

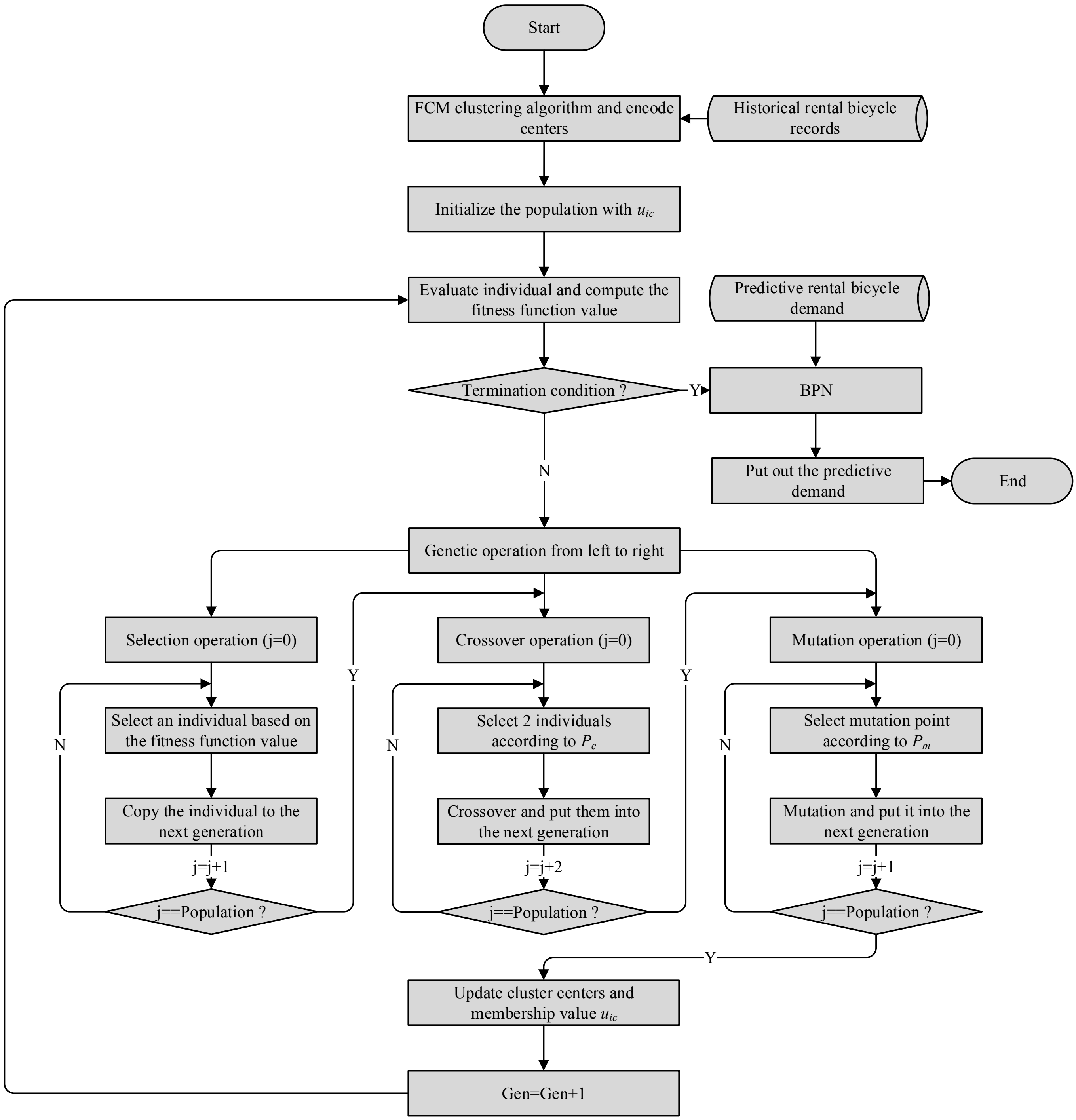

Fig. 4. Flowchart of the FCM-based GA and the BPN method.

chromosome, we calculate membership values u_{ic} for each data record i and each cluster c by randomly generated numbers of ϑ_{ic}, $0 \leq \vartheta_{ic} \leq 1$, as follows.

$$u_{ic} = \vartheta_{ic} \Big/ \sum_{c \in C} \vartheta_{ic}. \tag{7}$$

Step 3: Apply genetic operations: The value of the fitness function F (Eq. (6)) can be calculated for each individual. We then use genetic operations, namely the selection, crossover, and mutation operators, to improve population diversity.

Step 4: Apply optimal preservation. For each generation, the fitness values are re-calculated after the genetic operations have been applied, to evaluate each individual. Individual with higher fitness values are more likely to be chosen for survival.

Step 5: Check termination condition. In this study, the iteration terminates either after a given number of generations G or when a given fitness variance value ε is achieved. If either of these conditions is satisfied, then evolution stops. Otherwise, we return to Step 3. For a population of Q, the variance ε is calculated as follows:

$$\varepsilon = \sum_{s=1}^{Q} \left(F_s - \sum_{s=1}^{Q} F_s \Big/ Q \right)^2. \tag{8}$$

After the evolution process is complete, we can obtain the R-dimensional centroids for each cluster (category) and hence the

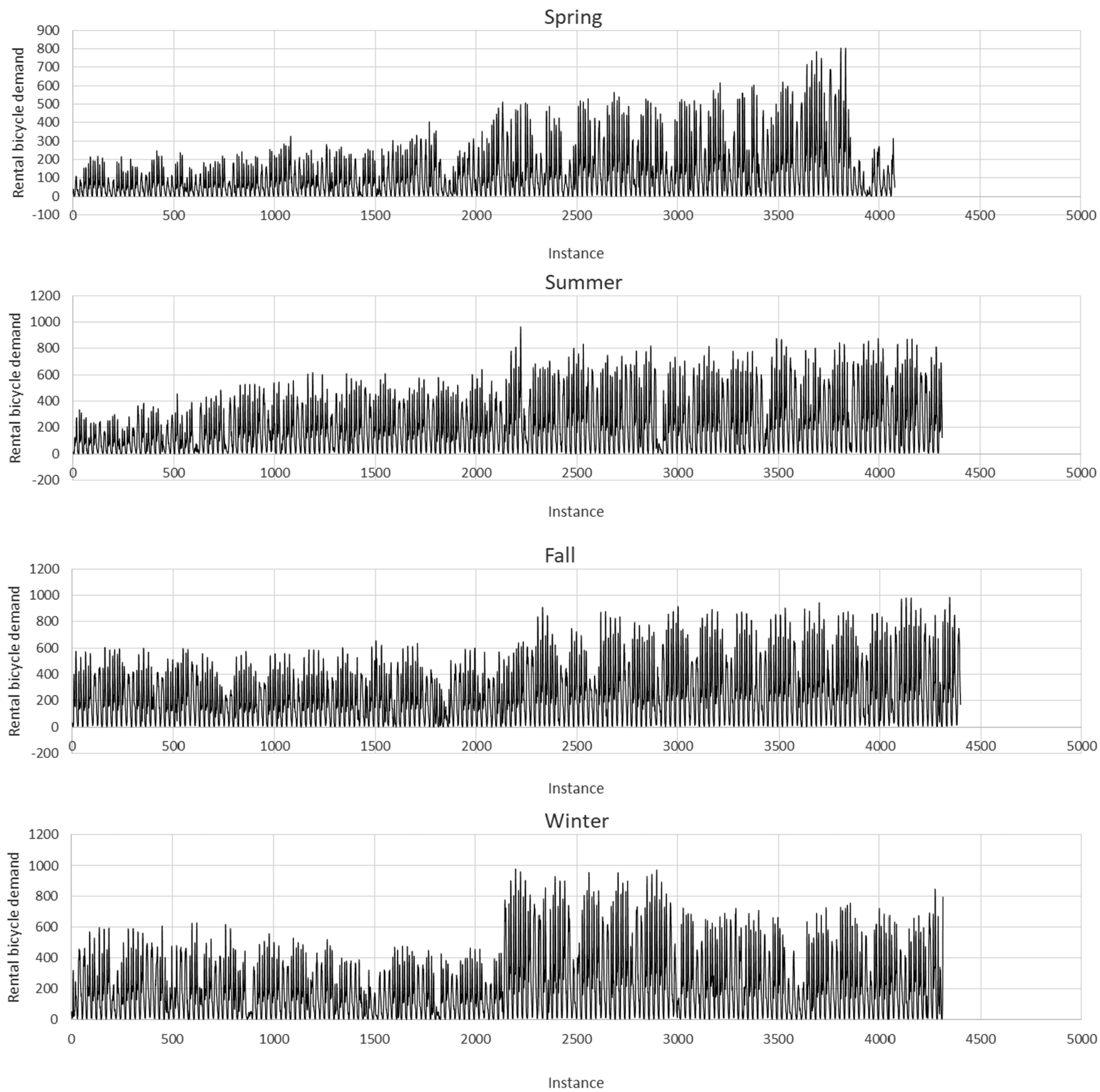

Fig. 5. Historical bicycle rental records divided by season.

membership values. We then associate each bicycle rental record to all categories where its membership u_{ic} is above a given threshold δ. Finally, the expected number of rental bicycles in each category c can be calculated as follows:

$$E_c = \sum_{i\in B} hd_i \cdot u_{ic} \Big/ \sum_{i\in B} u_{ic}. \tag{9}$$

3.4. BPN predictor

After classification is complete, a BPN is trained using historical bicycle rental records belonging to different categories because BPNs are commonly used to fit nonlinear relationships (Chen, 2016). The BPN predictor can then predict the future rental demand. Law (2000) emphasized that the forecast output by BPNs are accurate with a relatively small number of errors because they adjust their weights in the output layer to model the training elements. They compare their output with the actual values, and propagate the error back through the network. This process is repeated until the error falls within the acceptance range, at which point the neural network has been successfully trained. However, determining suitable training and architectural parameters is still difficult. These parameters are usually determined by trial and error or comparing them with each other. The process for configuring and training a BPN is as follows.

(1) Input layer: There are C cluster centers, each with R attributes. The attributes and bicycle rental records for cluster centers are normalized to fall within the interval [0,1].
(2) Single hidden layer: One hidden layer is used as this provides the best convergence.
(3) Output layer: The predicted bicycle rental demand in a normalized form is obtained.
(4) Training method: Gradient descent is the most commonly used training algorithm, and is adopted in this study because it avoids over-fitting.
(5) Activation/transformation function: We employ the following nonlinear sigmoid function (Ripley, 2007) as the activation function:

$$f(x) = \frac{1}{1+e^{-x}}. \tag{10}$$

Table 1
Parameters used in the proposed method.

Parameter	Value
Fuzziness exponent (m)	2
FCM termination tolerance (ϕ)	10^{-4}
Number of clusters (C)	10, 12, 14, 16, 18, and 20
Population size (Q)	40
Selected individuals (ϖ)	15
Number of generations (G)	200
Mutation probability (P_m)	0.2
Crossover probability (P_c)	0.1
Membership threshold (δ)	0.05
Fitness value variance (ε)	10^{-4}
Learning rate (α)	0.9
Number of neurons in the input layer	5
Number of neurons in the hidden layer	13

where x is a random variable.

(6) Learning rate (α): 0.01–1.0.
(7) Convergence criteria: Many measures can be used to determine when to cease BPN training, including the mean average error (MAE), mean squared error (MSE), and root mean squared error (RMSE). We employ the RMSE, a frequently used error measure (Rezaee et al., 2018), calculating it as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2}{N}}. \tag{11}$$

where Y_i is the output of the i-th unit.

Fig. 4 shows a flowchart for the complete FCM-based GA and BPN method.

4. Case study

4.1. Data description

The data used in this study were collected from the bicycle-sharing system in Washington D.C., USA (Fanaee-T, 2018). The core data came from historical logs recorded over a two-year period from 2011 to 2012. The data show that bicycle-sharing rentals are strongly correlated with the weather conditions and time-related characteristics, namely the weather, temperature, apparent temperature, humidity, wind speed, season, date, month, hour, and whether the day is a holiday, weekend, or working day. Fig. 5 shows the data, divided by season.

To demonstrate and evaluate the proposed method's prediction performance and compare its accuracy fairly, we split the data into two parts: the training (95%) and testing (5%) datasets. Because very few days were holidays, we chose to focus on the non-holiday data. For each season, we divided the data into two groups, namely working and weekend days. In order to improve the prediction accuracy, we took each consecutive 2-h period to be one moment and averaged the rental numbers within each period. These moment-based data points were then classified into groups using the FCM-based GA classifier and the output fed into the BPN predictor. The BPN was trained using these grouped rental records, after which it could predict the rental demand at future (2-h) moments.

4.2. Experimental setup

After training, the proposed model's effectiveness was evaluated by applying the BPN predictor to the testing data to predict the future bicycle rental demand. The prediction performance was measured and assessed in terms of the RMSE and MAE, where the MAE was calculated as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|Y_i - \widehat{Y}_i|}{N}. \tag{12}$$

where $|Y_i - \widehat{Y}_i|$ is the absolute difference value for the predicted and real values.

We tested using different numbers of categories for the FCM algorithm and compared the results. In addition, all experiments were carried out using populations with the same number of individuals. The remaining parameters used for the FCM-based GA and BPN are summarized in Table 1.

We implemented our hybrid approach in C++ using Visual Studio 2013. The experiments were run on a computer with an Intel Core i7-4790 CPU@3.6 GHZ and 16 GB of memory under Windows 10 Pro.

4.3. Results

Fig. 6 presents the predicted and observed bicycle rental demand on the weekend for each of the four seasons, showing the average

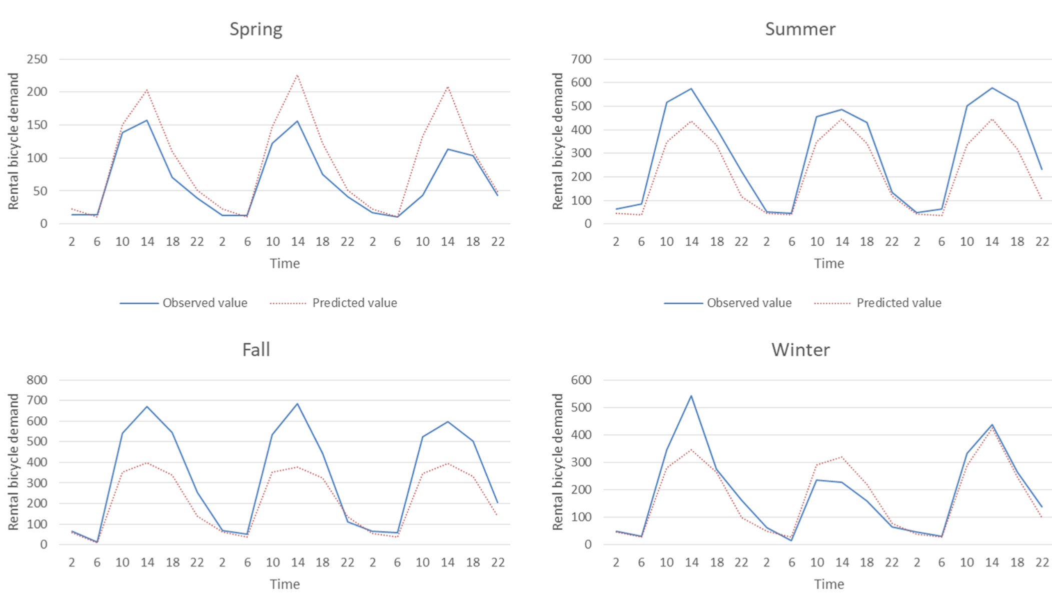

Fig. 6. Predicted 3-day bicycle rental demand on weekends for each of the four seasons.

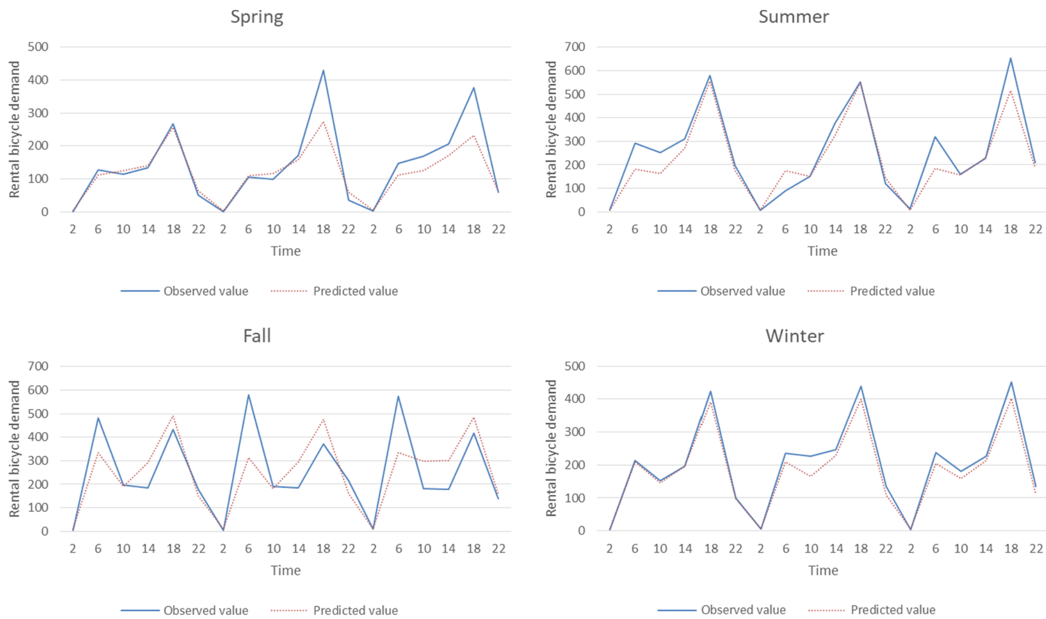

Fig. 7. Predicted 3-day bicycle rental demand on working days for each of the four seasons.

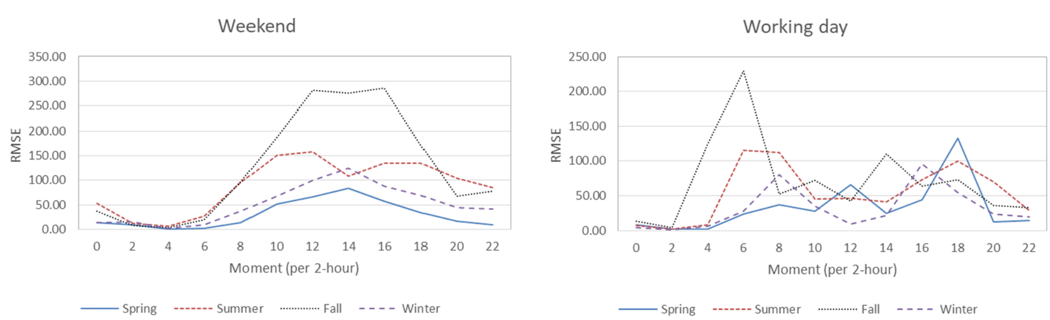

Fig. 8. RMSE over the course of the day for each of the four seasons.

predicted values for the six different numbers of categories (i.e., clusters). Fig. 7 presents a similar comparison for working days. Both figures show the 3-day patterns for testing dataset and that the predicted demand levels follow similar patterns to those of the observed data. In both cases, the trends are a good match when demand is low, but the proposed approach's predictions were not good during peak daytime hours as there were significant variations in the demand, even with similar weather on the same day. Here, we focus on the anticipated demand in typical cases.

Overall, the weather was strongly correlated with the bicycle rental demand. In Figs. 6 and 7, we can observe similar tendencies on both weekend and working days, respectively: more people want to travel by bicycle in the summer and fall, meaning that higher temperatures generally lead to increased bicycle use. In winter and spring, by contrast, more people choose other modes of public transportation. In addition, Fig. 7, shows the demand starting to increase earlier in the morning than in Fig. 6, due to more people choosing to ride a bicycle to work. The demand trends on weekend and working days are different in other ways as well: there are two obvious peaks during common commuting times on working days, emphasizing that people's activities are different on weekend and working days.

In order to investigate the effect of different times of day, the RMSE and MAE are evaluated for different times on weekend and working days during the four seasons, as shown in Figs. 8 and 9, respectively. Again, these are average values over the six different numbers of categories. These figures show similar trends in the RMSE and MAE values for a day, and both indicate obvious demand fluctuations over the day on the same day under the same weather conditions. The level of

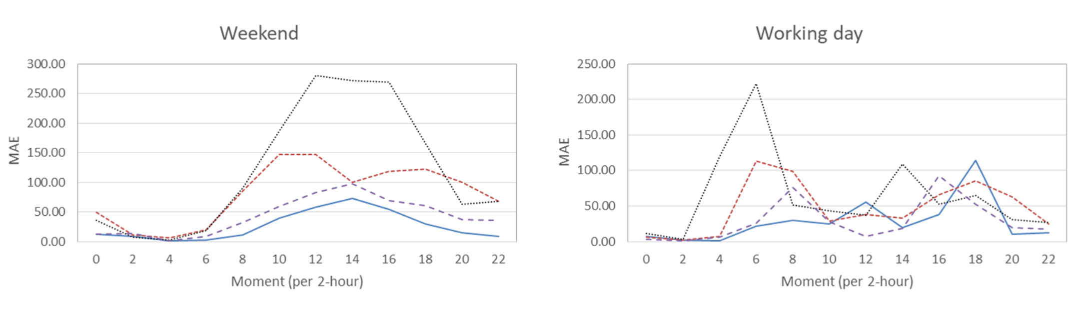

Fig. 9. MAE over the course of the day for each of the four seasons.

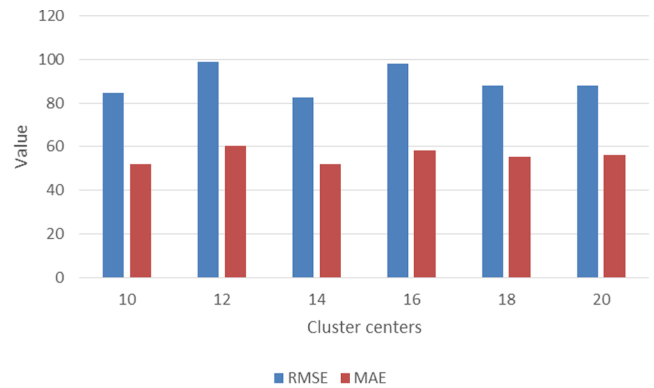

Fig. 10. RMSE and MAE for different number of clusters.

demand also changes with the season. In addition, the demand levels fluctuate more on working days than on weekends.

In order to determine the appropriate number of clusters, we tested six different options, namely 10, 12, 14, 16, 18, and 20 clusters. We used the RMSE and MAE to determine good number of clusters over all the testing data, and the results are shown in Fig. 10. These indicate that the RMSE and MAE do vary, and are smallest for 10 or 14 clusters, meaning that increasing the number of clusters does not provide better demand predictions.

We also tested the effect of using different numbers of clusters for the training data, and partial results are summarized in Tables 2 and 3. Table 2 summarizes the RMSE and MAE results for weekend (N) and working (Y) days in each of the four seasons (1: spring, 2: summer, 3: fall, 4: winter). Table 3 summarizes the RMSE and MAE results for different times of day on a weekend day (N) and a working day (Y) in spring. Table 2 summarizes that we obtain better demand predictions on working days than on weekends, while Table 3 summarizes that the demand predictions are strongly correlated with the time of day.

5. Conclusion and discussion

In this study, we have investigated the relationships between weather and date and bicycle rental demand using historical real-world data. We divided the data into two parts: training and testing data. In

Table 2
RMSE and MAE for weekend (N) and working (Y) days in different seasons.

S	D		RMSE						MAE					
		C	10	12	14	16	18	20	10	12	14	16	18	20
1	N		38.5	52.1	33.7	54.1	43.5	35.6	25.5	32.5	23.8	31.4	28.8	23.1
	Y		44.0	54.1	49.2	44.0	52.8	46.9	26.7	28.1	29.7	25.9	29.9	28.2
2	N		86.6	127.8	93.1	110.4	92.1	125.7	65.3	99.7	76.4	87.3	70.7	98.9
	Y		60.0	72.3	65.8	65.9	75.7	72.3	40.6	50.0	44.0	45.2	52.8	51.2
3	N		176.6	186.3	165.1	189.3	170.6	167.4	135.6	139.2	122.9	143.1	131.3	128.5
	Y		91.6	109.5	90.6	126.3	99.7	84.4	59.4	65.2	63.9	68.4	67.3	62.4
4	N		61.6	68.0	59.6	71.2	72.8	60.2	38.8	43.4	37.3	48.3	46.9	40.9
	Y		45.3	50.1	34.9	50.3	50.2	48.9	18.7	21.4	15.5	21.2	21.2	20.4

S: Season, D: Date.

Table 3
RMSE and MAE at different times on a weekend day (N) and a working day (Y) in spring.

D	M		RMSE						MAE					
		C	10	12	14	16	18	20	10	12	14	16	18	20
N	0–1		13.4	11.6	16.0	13.0	16.1	10.9	11.8	10.1	15.6	10.8	15.1	9.6
	2–3		9.5	10.1	6.1	9.5	8.4	9.7	8.9	9.8	6.1	9.3	7.9	9.0
	4–5		1.1	0.8	1.6	1.1	0.9	1.0	0.9	0.6	1.3	1.1	0.9	0.7
	6–7		1.8	2.5	2.9	2.5	1.9	3.2	1.5	1.9	2.4	2.0	1.7	2.9
	8–9		21.6	10.6	11.4	11.2	14.5	14.2	18.5	8.6	9.0	8.6	11.4	10.9
	10–11		55.7	44.7	56.0	43.8	65.0	43.1	38.7	41.7	47.5	27.0	47.5	34.3
	12–13		77.0	83.0	41.3	118.2	45.1	27.8	64.4	77.3	35.5	106.4	38.7	24.4
	14–15		64.0	89.4	70.2	108.9	93.1	71.7	63.1	73.5	67.4	84.7	85.6	66.6
	16–17		15.1	108.8	37.5	53.2	61.9	70.6	9.5	101.4	36.4	50.9	61.2	65.5
	18–19		29.5	36.6	30.5	39.1	41.4	31.2	25.3	28.8	28.8	33.0	36.5	26.1
	20–21		41.5	6.7	10.6	19.5	12.9	4.8	41.1	6.3	9.8	16.2	10.2	4.0
	22–23		6.8	9.7	11.0	7.1	10.3	8.8	6.6	9.3	10.9	6.6	10.1	8.5
Y	0–1		8.8	9.6	7.3	7.8	8.1	8.4	7.8	8.6	6.4	7.1	7.3	7.6
	2–3		2.0	2.0	1.4	2.4	2.1	1.9	1.8	1.8	1.3	2.3	1.9	1.7
	4–5		1.7	1.8	1.1	2.2	1.4	1.4	1.6	1.8	1.1	1.9	1.3	1.3
	6–7		22.5	19.3	19.9	22.5	23.8	34.7	19.7	17.9	17.9	20.3	21.3	31.6
	8–9		33.5	37.4	32.8	33.4	55.3	30.5	27.8	33.0	26.7	28.4	42.8	22.4
	10–11		24.2	18.9	33.1	26.1	37.4	29.9	20.6	12.5	30.1	23.5	33.4	28.0
	12–13		66.2	62.5	74.5	63.2	64.8	62.0	58.6	49.0	65.2	51.6	52.7	55.2
	14–15		16.3	19.1	31.5	18.7	40.3	21.1	12.4	16.1	26.7	16.9	29.4	17.5
	16–17		56.7	38.0	45.8	50.3	34.5	39.5	48.7	29.7	41.9	41.2	27.5	38.7
	18–19		112.8	163.9	130.6	117.3	144.5	130.6	99.7	146.7	106.7	102.7	118.7	113.6
	20–21		9.4	11.8	25.9	6.7	9.7	8.1	8.1	9.8	23.7	5.6	7.3	7.1
	22–23		17.3	14.2	9.4	11.7	17.7	17.0	13.6	10.6	9.0	9.8	14.9	14.1

D: Date, M: Moment.

order to effectively predict the rental demand in bicycle-sharing systems, we have proposed a moment-based model and a new hybrid approach that combines an FCM-based GA and a BPN. The FCM-based GA represents a new soft unsupervised classification method where the GA is used to improve classification performance. This provides better pre-classification of the training data, which are then used to train a BPN predictor on rental records belonging to different categories. After that, the BPN predictor can be used to predict the rental demand at future moments.

We have also presented a case study based on real-life data to demonstrate the application of the proposed method. This led us to the following conclusions. (1) The classification done by FCM-based GA leads the BPN to more accurate demand predictions. (2) Using the moment-based model, we found that the demand was strongly correlated with the weather conditions. (3) The predictions were more accurate at night than during the day due to significant fluctuations in the daytime demand, even on the same day under the same weather conditions. (4) The demand was also influenced by people's different lifestyles on weekend and working days.

In future studies, we plan to explore the following interesting problems. In the proposed approach, fuzzy concepts could be incorporated into the BPN to further enhance prediction accuracy. In this study, we have only used 3-layered BPN but the recent advancement in deep learning techniques draws our attention for the further research. Another avenue for future work is to consider logistical decision-making strategies for bicycle redistribution in uncertain environments. Both these questions will be considered in future research.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2015R1A2A1A10054253 and No. 2018R1A2B3008890).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cie.2018.12.023.

References

Barnes, G., & Krizek, K. (2005). Estimating bicycling demand. *Transportation Research Record: Journal of the Transportation Research Board,* (1939), 45–51.

Bezdek, J. C., & Dunn, J. C. (1975). Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. *IEEE Transactions on Computers, 100*(8), 835–838.

Borgnat, P., Abry, P., Flandrin, P., & Rouquier, J.-B. (2009). *Studying Lyon's Vélo'v: A statistical cyclic model. ECCS'09*. Complex System Society.

Chen, T. (2003). A fuzzy back propagation network for output time prediction in a wafer fab. *Applied Soft Computing, 2*(3), 211–222.

Chen, T. (2007). An intelligent hybrid system for wafer lot output time prediction. *Advanced Engineering Informatics, 21*(1), 55–65.

Chen, T. (2008). A hybrid fuzzy-neural approach to job completion time prediction in a semiconductor fabrication factory. *Neurocomputing, 71*(16), 3193–3201.

Chen, T. (2013). An effective fuzzy collaborative forecasting approach for predicting the job cycle time in wafer fabrication. *Computers & Industrial Engineering, 66*(4), 834–848.

Chen, T. (2016). Estimating job cycle time in a wafer fabrication factory: A novel and effective approach based on post-classification. *Applied Soft Computing, 40*, 558–568.

Contardo, C., Morency, C., & Rousseau, L.-M. (2012). Balancing a dynamic public bike-sharing system. *Cirrelt Montreal*.

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: The effect of weather and calendar events. *Journal of Transport Geography, 41*, 292–305.

dell'Olio, L., Ibeas, A., & Moura, J. L. (2011). Implementing bike-sharing systems. *Proceedings of the Institution of Civil Engineers, 164*(2), 89.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation, 12*(4), 3.

Ding, Y., & Fu, X. (2016). Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. *Neurocomputing, 188*, 233–238.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

El-Assi, W., Mahmoud, M. S., & Habib, K. N. (2017). Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. *Transportation, 44*(3), 589–613.

Erdoğan, G., Battarra, M., & Calvo, R. W. (2015). An exact algorithm for the static rebalancing problem arising in bicycle sharing systems. *European Journal of Operational Research, 245*(3), 667–679.

Fanaee-T, H., & Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence, 2*(2–3), 113–127.

H. Fanaee-T. Bike sharing dataset. Available: <http://capitalbikeshare.com/system-data>.

Feng, C., Hillston, J., & Reijsbergen, D. (2017). Moment-based availability prediction for bike-sharing systems. *Performance Evaluation, 117*, 58–74.

Fishman, E. (2016). Bikeshare: A review of recent literature. *Transport Reviews, 36*(1), 92–113.

Gao, X., Zhou, Y., Amir, M. I. H., Rosyidah, F. A., & Lee, G. M. (2017). A hybrid genetic algorithm for multi-emergency medical service center location-allocation problem in disaster response. *International Journal of Industrial Engineering, 24*(6).

García-Palomares, J. C., Gutiérrez, J., & Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography, 35*(1–2), 235–246.

Gast, N., Massonnet, G., Reijsbergen, D., & Tribastone, M. (2015). Probabilistic forecasts of bike-sharing systems for journey planning. *Proceedings of the 24th ACM international on conference on information and knowledge management* (pp. 703–712). ACM.

Kadri, A. A., Kacem, I., & Labadi, K. (2016). A branch-and-bound algorithm for solving the static rebalancing problem in bicycle-sharing systems. *Computers & Industrial Engineering, 95*, 41–52.

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing, 6*(4), 455–466.

Kenesei, T., Balasko, B., & Abonyi, J. (2006). A MATLAB toolbox and its web based variant for fuzzy cluster analysis. *Proceedings of the 7th international symposium on Hungarian researchers on computational intelligence* (pp. 24–25). .

Kim, K. (2018). Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Journal of Transport Geography, 66*, 309–320.

Kim, S. H., & Lee, C. M. (1997). Nonlinear prediction of manufacturing systems through explicit and implicit data mining. *Computers & Industrial Engineering, 33*(3–4), 461–464.

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management, 21*(4), 331–340.

Mi, C., Yang, J., Li, S., Zhang, X., & Zhu, D. (2010). Prediction of accumulated temperature in vegetation period using artificial neural network. *Mathematical and Computer Modelling, 51*(11–12), 1453–1460.

Pfrommer, J., Warrington, J., Schildbach, G., & Morari, M. (2014). Dynamic vehicle redistribution and online price incentives in shared mobility systems. *IEEE Transactions on Intelligent Transportation Systems, 15*(4), 1567–1578.

Raviv, T., & Kolka, O. (2013). Optimal inventory management of a bike-sharing station. *IIE Transactions, 45*(10), 1077–1093.

Rezaee, M. J., Jozmaleki, M., & Valipour, M. (2018). Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange. *Physica A: Statistical Mechanics and its Applications, 489*, 78–93.

Ripley, B. D. (2007). *Pattern recognition and neural networks.* Cambridge, Mass: Cambridge University Press.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach.* Malaysia: Pearson Education Limited.

Schuijbroek, J., Hampshire, R. C., & Van Hoeve, W.-J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research, 257*(3), 992–1004.

Shaheen, S. A., Martin, E. W., Cohen, A. P., Chan, N. D., & Pogodzinski, M. (2014). Public bikesharing in North America during a period of rapid expansion: Understanding business models, industry trends & user impacts, MTI report 12-29.

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board, 2143*, 159–167.

Sun, B., Guo, H., Karimi, H. R., Ge, Y., & Xiong, S. (2015). Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing, 151*, 1528–1536.

Tirkel, I. (2011). Cycle time prediction in wafer fabrication line by applying data mining methods. *Advanced semiconductor manufacturing conference (ASMC), 2011 22nd annual IEEE/SEMI* (pp. 1–5). IEEE.

Vogel, P., & Mattfeld, D. C. (2011). Strategic and operational planning of bike-sharing systems by data mining – A case study. *International conference on computational logistics* (pp. 127–141). Springer.

Wang, H. (2007). Application of BPN with feature-based models on cost estimation of plastic injection products. *Computers & Industrial Engineering, 53*(1), 79–94.

Wikaisuksakul, S. (2014). A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering. *Applied Soft Computing, 24*, 679–691.

Ye, A.-X., & Jin, Y.-X. (2016). A fuzzy c-means clustering algorithm based on improved quantum genetic algorithm. *simulation, 9*(1).

Yokota, T., Gen, M., & Li, Y.-X. (1996). Genetic algorithm for non-linear mixed integer programming problems and its applications. *Computers & Industrial Engineering, 30*(4), 905–917.

Yu, X., & Gen, M. (2010). *Introduction to evolutionary algorithms.* London, UK: Springer.