Time-Series Identification on Fish Feeding Behaviour

Machine Learning in Aquaculture pp 37-47 | Cite as

- Mohd Azraai Mohd Razman (1) Email author (azraai@ump.edu.my)
- Anwar P. P. Abdul Majeed (1)
- Rabiu Muazu Musa (2)
- Zahari Taha (1)
- Gian-Antonio Susto (3)
- Yukinori Mukai (4)

1. Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang, , Pekan, Malaysia
2. Centre for Fundamental and Continuing Education, Department of Credited Co-curriculum, Universiti Malaysia Terengganu, , Terengganu, Malaysia
3. Department of Information Engineering, University of Padua, , Padua, Italy
4. Department of Marine Science, International Islamic University Malaysia, , Kuantan, Malaysia

Chapter
First Online: 03 January 2020

Part of the SpringerBriefs in Applied Sciences and Technology book series (BRIEFSAPPLSCIENCES)

Abstract

The identification of relevant parameters that could describe the state of fish hunger is vital for ensuring the appropriate allocation of food to the fish. The establishment of these relevant parameters is non-trivial, particularly when developing an automated demand feeder system. The present inquiry is being undertaken to determine the hunger state of *Lates calcarifer*. For data collection, a video analysis system is used, and the video was taken all day, where the fish was fed by an automatic feeding system. Sixteen characteristics of the raw data set have been extracted through feature engineering for 0.5 min, 1.0 min, 1.5 min and 2.0 min, respectively, in accordance with the mean, peak, minimum and variability of each of the different time window scales. Furthermore, the features extracted have been evaluated through principal component analysis (PCA) both for dimension reduction and PCA with varimax rotation. The details were then categorized using support vector machine (SVM), K-NN and random forest tree (RF) classifiers. The best identification accuracy was shown with eight described features in

Mohd Razman M.A., P. P. Abdul Majeed A., Muazu Musa R., Taha Z., Susto GA., Mukai Y. (2020) Time-Series Identification on Fish Feeding Behaviour. In: Machine Learning in Aquaculture. SpringerBriefs in Applied Sciences and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-15-2237-6_4

- First Online 03 January 2020
- DOI https://doi.org/10.1007/978-981-15-2237-6_4
- Publisher Name Springer, Singapore
- Print ISBN 978-981-15-2236-9
- Online ISBN 978-981-15-2237-6
- eBook Packages Biomedical and Life Sciences Biomedical and Life Sciences (R0)

Chapter 4
Time-Series Identification on Fish Feeding Behaviour

Abstract The identification of relevant parameters that could describe the state of fish hunger is vital for ensuring the appropriate allocation of food to the fish. The establishment of these relevant parameters is non-trivial, particularly when developing an automated demand feeder system. The present inquiry is being undertaken to determine the hunger state of *Lates calcarifer*. For data collection, a video analysis system is used, and the video was taken all day, where the fish was fed by an automatic feeding system. Sixteen characteristics of the raw data set have been extracted through feature engineering for 0.5 min, 1.0 min, 1.5 min and 2.0 min, respectively, in accordance with the mean, peak, minimum and variability of each of the different time window scales. Furthermore, the features extracted have been evaluated through principal component analysis (PCA) both for dimension reduction and PCA with varimax rotation. The details were then categorized using support vector machine (SVM), K-NN and random forest tree (RF) classifiers. The best identification accuracy was shown with eight described features in the varimax-based PCA. The forecast results based on the K-NN model built on selected data characteristics showed a level of 96.5% indicating that the characteristics analysed were crucial to classifying the actions of hunger among fisheries. AQ1

Keywords Image processing · Automated demand feeder · *Lates calcarifer* · Pixel intensity · Specific growth rate

4.1 Overview

Primarily, a group of fish exhibits higher movement when searching for food which often describes their state of hunger while the movement tends to decline or reduces when the fish is satiated [1, 2]. Therefore, the swimming ability of the fish may fluctuate with regard to the hunger level of the fish. A research has shown that fish appear to be more aggressive and cover the field when they are starving and therefore have a greater movement as individuals and in a shoal [3]. Another experiment has shown that hunger can be caused by access to fish to the ultradian pattern of light and darkness replicating pulses every day and night [4]. This scenario implies that the formation of the time interval has to be taken into account. However, it is challenging

M. A. Mohd Razman et al., *Machine Learning in Aquaculture*,
SpringerBriefs in Applied Sciences and Technology,
https://doi.org/10.1007/978-981-15-2237-6_4

to state the accurate period as the moment of time and day could alter the fish hunger behaviour couple with other endogenous as well as exogenous influences. Therefore, a time-series analysis has to be employed to mitigate the aforesaid issue which could be addressed by means of feature engineering extraction technique obtain through image processing.

The present chapter is structured in sequential order. The first stage is aimed at identifying the occurrence of events involved in the state of fish hunger behaviour between two classes as examined earlier. The second stage comprised of the dissolution of the previously classified group in order to gauge the appropriate time window for the purpose of addressing the second layer features. The PCA method is then implemented to provide insights on the optimal features that could be used in describing the hunger behaviour of the fish. Subsequently, the SVM model is developed for validating the selected features while hyperparameter tuning is carried out to optimize the classification accuracy of the model. Moreover, correlation matrices, as well as scree plots, were applied to identify the significant features that contribute towards the formation of the effective ML models that best explain the hunger behaviour of the fish in the present investigation. It is worth highlighting that the optimization analysis section is essentially formulated to aid in the identification of the best parameters that could be used to improving both the accuracy as well as the run-time of the model developed in the present investigation.

4.2 Event Identification

It is essential to choose the instances involved in the behavioural changes of fish hunger in order to permit the labelling of the responses exhibited according to the classes that were previously clustered. The often produced performance from the cluster analysis is only capable of producing the best number of clusters as regards data sets; as such the marking of specific and real groups needs to be done on the basis of an expert in the field of fish behaviour. The details in the time series obtained in this analysis are therefore in a real-world situation in which the parameters differ with the period stated below:

$$y(t),\ t = 1, 2, \ldots, n \tag{4.1}$$

where the parameters or features extracted are $y(t)$, and time variable represented by t [5]. The known changes in this study lie when the fish is fed for a given time and as it reaches full satiated phase the behaviour changes to a stagnant motion or as portray by the automated demand feeder is when the activation from the trigger sensor is stopped. One can then conclude that the responses can now be treated as ‘Satiated’ and can be classed as ‘Hungry’ when the sensor is activated.

490571_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:10/12/2019 Pages: 48 Layout: T1-Standard

4.3 Features Selection PCA-Based

As a time sequence issue is understood, it can be represented when features in second layers if behavioural variations are accomplished. The answers can be identified by analysing the activity period or the circadian rhythm moment of famine [6]. *P* feature extraction technique is employed to extract the window size of the features. In this step, the implication of applying *p* features is demonstrated. The input is expressed from $x \in \mathbb{R}^{1\times p}$ is gained by the several features in time series as below:

$$a(t) = \mathbb{R}^{n\times p},\ t = 1, \ldots, U \tag{4.2}$$

where $\mathbb{R}$ is the actual number collection, n is the number of sizes selected and *p* is the number of features of the second level derived from video processing capability. For example, a 0.5 min window has four features (mean, max, min, var) which show the first *p* functionality with four features inside a *p*. Through adding a further 1.0 min window, the *p* features are raised to 2*p*. Hence, 0.5, 1.0, 1.5 and 2.0 min window dimensions are collected in this case, which indicates that sample size should add 4*p* characteristics, as shown in this case $x \in \mathbb{R}^{m\times 4p}$. In addition, it results in 16 features in full. The window size is where the detection of movement is being generalized in which the speed of hunger or any sudden changes or abruption could be detected, the steadiness of the group and the outliers of individual fish from the group could, therefore, be determined.

As far as the PCA concept is concerned, the varimax rotation involves recognition of variables dependent on the PC's own values greater than 1. For each element, the associated function is the varimax factor, which defines the unknown, conjectural and indistinct variables. The variables are positive or negative and find all levels hitting up to 1 and -1 to be a high correlation [7]. AQ2

The labelling of the actual classes of the fish in relation to the hunger behaviour could be determined from the motion of the fish by means of event identification technique. In the time-series analysis, the second layer features analysis could offer information on the circadian rhythm of the fish hunger behaviour. At this juncture, the technique of classifying the hunger behaviour has been exemplified for identifying the fish behaviour through image processing features, translating the features, analysing the significant parameters as well as verifying the period instances of the movement. The processed data is then classified through the application of a variety of ML models, and comparative analysis is conducted to ascertain the best predictive model of the fish hunger behaviour in the present study.

4.4 Classification Accuracy

In this research SVM, *k*-NN and RF are tested as identification templates. This paper does not expand on the specifics of these designs. The readers are nevertheless

490571_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:10/12/2019 Pages: 48 Layout: T1-Standard

urged to consider the previous literature [8]. SWM builds on the limits of categories commonly referred to as ranges as the largest size. The variable of classification is a feature based on a balance between the optimizing and the penalizing mistake in the classification equation [9].

On the other side, the k-NN method utilizes the range among neighbour's marked input spaces. The method to distance between Euclidean's is often used to measure distances between adjacent individuals and was shown to be successful in classifying fish pictures for freshness detection [10].

The RF system has been developed to improve the accuracy level of classification for traditional decision trees often called classification and regression trees (CART) by sharing the node or by arbitrarily searching for the responses [11]. This design is preferable to the previous version because it reduces the overfitting possibilities. The concept of assembling the trees that satisfy enough for the forest; in other words, the entire model will reduce the overfitting of the classifier. Furthermore, it has the advantages of managing the uncertainties value that is lost during searching, and the model could be described in a definite model.

In this analysis, box constraint, *c* and kernel scale, hyperparameters are chosen as 1 for the linear SVM, while for k-NN the Euclidean distance with the number of neighbours chosen as 1 is used for the Euclidean size.

The criterion of classification accuracy with the percentage of the predicted responses has been correctly classified by computing the confusion matrix that points out the misclassification rate distributed onto the respective classes [12]. The variants of the ML models used in this analysis were tested using reliability of identification, recall or sensitivity and precision. Figure 4.1 demonstrates the confusion matrix as an evaluation of the predictive models.

As illustrated in Fig. 4.1, the confusion matrix evaluates the respective results from the actual against the predicted outcome in terms of accuracy, precision, recall and F_1 score. The understanding of the matrix outcome would be better explained in the analogy of this study to classify hunger or satiated state. For instance, the True Positive (TP) stands when predicting the fish as hungry; however, the False

Fig. 4.1 Confusion matrix table

Confusion Matrix	**Predicted**		
	True	**Hungry**	**Satiated**
Actual	**Hungry**	**True Positive**	**False Negative**
	Satiated	**False Positive**	**True Negative**

Positive (FP) when the hunger state has been labelled into the satiated phase. These conditions conversely followed by the False Negative (FN) in which the expected state of the satiation behaviour is incorrectly labelled as hungry. Nonetheless, the True Negative (TN) accurately marks the satiated responses. The instances generated by the confusion matrix for the accuracy, precision, recall and F_1 score can be expressed through the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.4}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4.5}$$

From the stated equations, the accuracy stands as the overall classification rate on the regularity of the classifier in predicting the classes, precision evaluates the prediction rate of being recurrent yes or correctly predicted. The recall is the fractions of yes being labelled or the positive that is correctly categorized. The F_1 is the harmonic mean that is generated between precision and recall by multiplying the scale by 2. The F_1 score suggests the low volume of classes that are being falsely labelled and hence shows the strength of the prediction from the overall accuracy.

4.5 Results and Discussion

A total number of 59,807 were acquired throughout four days experiment. Both the class identification and the feeding activity dependent on COGy and box size are assessed. The option of these two characteristics among the seven is focused on the recommendation of previous studies [8, 13]. The general intention of the analysis is to measure the hunger between the satiated and the deprived, often the target of a mere human decision that can be skewed and misunderstood. The process begins by removing the fish feed time from the automated feeder until it ceases requesting. Here it is believe that it can be calculated by identifying the feeding cycle that is ‘Hungry’ or ‘Satiated’. For example, since the feed period was at 11:00 am, the feed ends at about five minutes as the rate of COGy is declining, while the pits rise at 11:10 am where the feed time is shown in Fig. 4.2. It indicates that the behaviour changes from hunger to satiation. The feeding cycle of the data set was not included, and only the starving and satiated condition was left. It means that both groups are collected for the corresponding controlled sample evaluation [14].

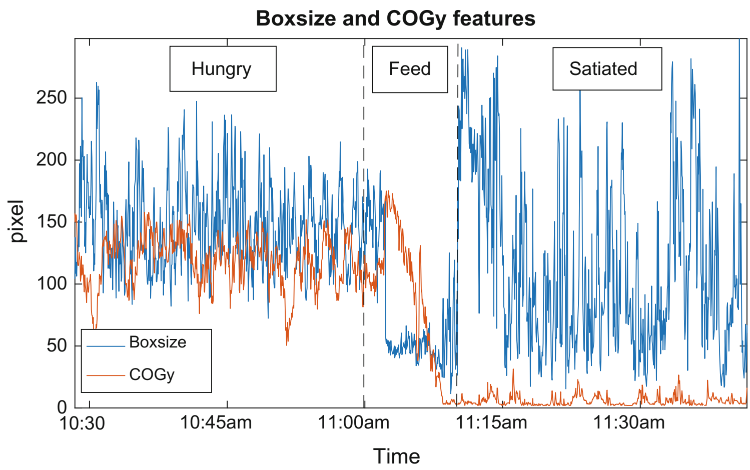

Fig. 4.2 Temporal window extraction for feeding process

The scree interpretation of the PCA study presented in Fig. 4.3 shows that the results that the first two components add up a high percentage of variance in which the quality of the particular principal is higher than one and the corresponding variable axis is suggested [12]. The first variable charging shows the value of 11.2 with the percentage variance of about 73%, and then the second with 3.1 eigenvalue, and thus increasing the percentage of total volatility to approximately 90%. The outcomes

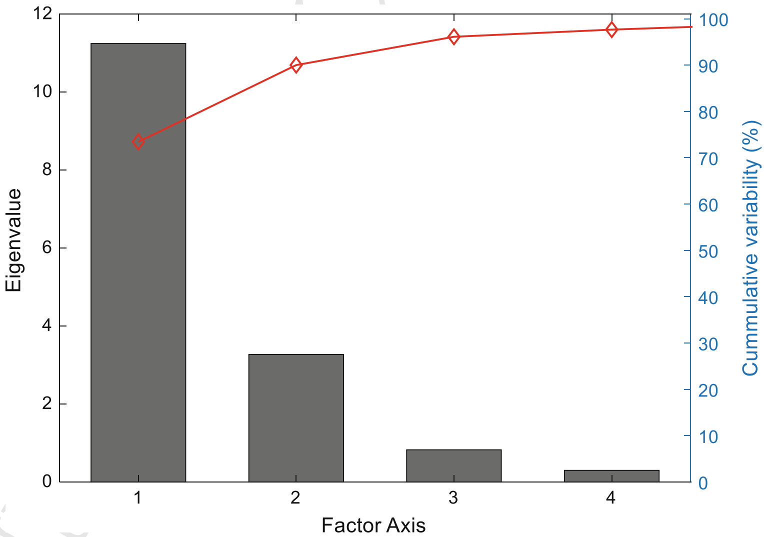

Fig. 4.3 Cumulative variability and eigenvalues scree plot

Table 4.1 Factor loading after varimax rotation

Features		Components	
		C1	C2
0.5 min	Mean	0.97	−0.04
	Max	0.65	−0.64
	Min	0.65	−0.64
	Var	0.07	−0.88
1.0 min	Mean	0.98	−0.04
	Max	0.70	−0.68
	Min	0.70	−0.68
	Var	0.07	−0.96
1.5 min	Mean	0.98	−0.04
	Max	0.71	−0.66
	Min	0.71	−0.66
	Var	0.06	−0.96
2.0 min	Mean	0.98	−0.03
	Max	0.71	−0.62
	Min	0.71	−0.62
	Var	0.05	−0.92

show that both major components are equally important when defining key features in the data set.

Table 4.1 points out the chosen element features from C1 and C2 and illustrates the results of factor loadings on the characteristics tested during varimax rotation. In defining hunger behaviour, bolded typeface values are considered significant. It is clear because, as indicated by the literature, the mean and variance for all windows returned important characteristics [15].

The PCA study concluded that a maximum of eight characteristics are necessary for assessing the fish's hunger behaviour from the initial data developed, while the PCA-based varimax rotation found that two dimensions were sufficient to describe about 90% of the whole data set variability. Hence, the mean square error (MSE) was calculated by SVM classifier to calculate the discrepancy between PCA, PCA with varimax rotation and the 16 features mentioned in Table 4.2 for each function variance.

The MSE percentage and the uncertainty matrix of the three different features that the SVM template assesses are provided in Table 4.2. The eight features from the PCA-based varimax rotation could be observed to show the smallest MSE at 17.40%. It is important to note that the MSE of the 16 features varies little from the two-dimension obtained from the generic PCA study. As a consideration, further analysis of the new data is needed to establish the efficacy for the chosen features of the SVM system.

Table 4.2 SVM classifier accuracy on train data set

(MSE) %		Train data set					
		Two-dimension (2D)		Eight features (Varimax)		16 features (All)	
Error		17.42% ± 0.0003%		17.40% ± 0.0005%		17.79% ± 0.0008%	
Confusion matrix (%)	True	H	S	H	S	H	S
	H	0.77	0.23	0.76	0.24	0.77	0.23
	S	0.12	0.88	0.12	0.88	0.13	0.87

Table 4.3 SVM classifier accuracy on test data set

(MSE) %		Test data set					
		Two-dimension (2D)		Eight features (Varimax)		16 features (All)	
Error		17.53%		17.36%		18.34%	
Confusion matrix (%)	True	H	S	H	S	H	S
	H	0.77	0.23	0.77	0.23	0.76	0.24
	S	0.14	0.86	0.13	0.87	0.14	0.86

Table 4.3 demonstrates the comparative reliability between variants relying on chosen characteristics, i.e. eight and seventeen. With eight features, it is clear that the deviation is decreased from 17.4% of training predictions to 17.36% for test error. The evidence presented here indicates that the template of eight characteristics best forecasts on a fresh data set while the 16 characteristics suggest MSE development from 17.79 to 18.34%. Such results combined support a clearer overview of the eight features than the 16 features for SVM models.

Figure 4.4 shows the variations of the data sets among all classifiers involved. Implementing the single PCA to minimize the dimensionality of the data set does not significantly affect the reliability of classification [16]. For RF and *k*-NN classifiers, it shows that the varimax set has lower marginal error with 7.6% and 3.5%, respectively, as opposed to two-dimensional PCA with 17.0% and 20.5%, respectively. Nevertheless, it is clear from this work that the use of PCA with varimax rotation factor loads contributed to a higher identification performance than the traditional PCA methodology. Even though the classification reliability of PCA varimax and the choice of all features are shown to be poor, it is important to note that around half of the information is decreased and that the processing time used for non-trivial real-time classification decreases dramatically [17–19]. The *k*-NN method gives, ultimately, 3.5% error of the system tested by choosing features using PCA with a varimax rotation, the highest classification level.

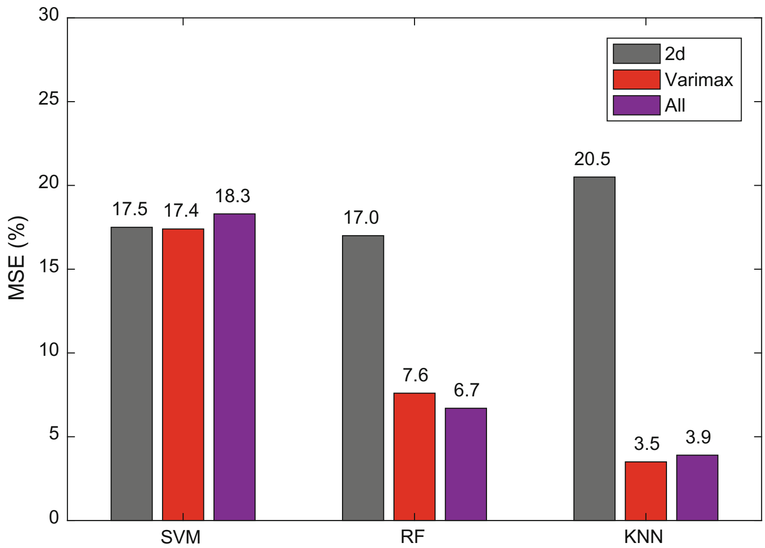

Fig. 4.4 Classification accuracy on the test set

4.6 Summary

Each section has developed a number of applicable features from the different window sizes in which 16 features have been extracted. The PCA evaluation was performed to establish the main characteristics which could characterize fish starvation. In order to extract different characteristics depending on a variable load level, PCA with varimax rotation was also used. In order to evaluate the identification effectiveness of hunger behaviour in relation to above-described functional variance SVM, *k*-NN, RF models are established. Through the presenting study, it was observed that the PCA-based varimax rotation by *k*-NN classifier with eight features could well define hunger in comparison with the different models. The approach suggested is necessary in the knowledge of the circadian rhythm of the *Lates calcarifer* coupled with the feeding schedule, which therefore would not only assist in the laboratory but also in the development and production of highly demanded species [20]. It is important to stress that the system built can be studied beyond and contrasted with other recent models including deep learning, augmented learning and the generative opponent network among many others. For real-time automatic fish feeding systems, it can also be transferred to a microcontroller.

References

1. Kennedy J, Jónsson SÞ, Ólafsson HG, Kasper JM (2016) Observations of vertical movements and depth distribution of migrating female lumpfish (*Cyclopterus lumpus*) in Iceland from data storage tags and trawl surveys. ICES J Mar Sci J du Cons 73:1160–1169. https://doi.org/10.1093/icesjms/fsv244
2. Nakayama S, Johnstone RA, Manica A (2012) Temperament and hunger interact to determine the emergence of leaders in pairs of foraging fish. PLoS ONE 7:e43747. https://doi.org/10.1371/journal.pone.0043747
3. Chapman BB, Morrell LJ, Krause J (2010) Unpredictability in food supply during early life influences boldness in fish. Behav Ecol 21:501–506
4. Sanchez-Vázquez FJ, Madrid JA, Zamora S (1995) Circadian rhythms of feeding activity in sea bass, *Dicentrarchus labrax* L.: dual phasing capacity of diel demand-feeding pattern. J Biol Rhythms 10:256–266. https://doi.org/10.1177/074873049501000308
5. Guralnik V, Srivastava J (1999) Event detection from time series data. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining—KDD'99. ACM Press, New York, New York, USA, pp 33–42
6. Dutta MK, Sengar N, Kamble N, Banerjee K, Minhas N, Sarkar B (2016) Image processing based technique for classification of fish quality after cypermethrine exposure. LWT Food Sci Technol 68:408–417. https://doi.org/10.1016/J.LWT.2015.11.059
7. Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer Ser Stat vol 98, p 487. 10.2307/1270093
8. Razman MAM, Susto GA, Cenedese A, Abdul Majeed APP, Musa RM, Abdul Ghani AS, Adnan FA, Ismail KM, Taha Z, Mukai Y (2019) Hunger classification of *Lates calcarifer* by means of an automated feeder and image processing. Comput Electron Agric 163:104883. https://doi.org/10.1016/J.COMPAG.2019.104883
9. Chang C-C, Lin C-J (2011) LIBSVM. ACM Trans Intell Syst Technol 2:1–27. https://doi.org/10.1145/1961189.1961199
10. Iswari NMS, Wella, Ranny (2017) Fish freshness classification method based on fish image using k-Nearest Neighbor. In: 2017 4th international conference on new media studies (CONMEDIA): 8–10 Nov 2017, Yogyakarta, Indonesia. IEEE, pp 87–91
11. Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
12. Dangeti P (2017) Statistics for machine learning build supervised, unsupervised, and reinforcement learning models using both Python and R. Packt Publishing
13. Mukai Y, Tan NH, Khairulanwar M, Chung R, Liau F (2016) Demand feeding system using an infrared light sensor for brown-marbled grouper juveniles, *Epinephelus fuscoguttatus*. Sains Malaysiana 45:729–733
14. Broell F, Noda T, Wright S, Domenici P, Steffensen JF, Auclair J-P, Taggart CT (2013) Accelerometer tags: detecting and identifying activities in fish and the effect of sampling frequency. J Exp Biol 216:1255–1264. https://doi.org/10.1242/jeb.077396
15. Peré-Trepat E, Olivella L, Ginebreda A, Caixach J, Tauler R (2006) Chemometrics modelling of organic contaminants in fish and sediment river samples. Sci Total Environ 371:223–237. https://doi.org/10.1016/j.scitotenv.2006.04.005
16. Tian XY, Cai Q, Zhang YM (2012) Rapid classification of hairtail fish and pork freshness using an electronic nose based on the PCA method. Sensors 12:260–277. https://doi.org/10.3390/s120100260
17. Vanaja S, Ramesh Kumar K (2014) Analysis of feature selection algorithms on classification: a survey. Int J Comput Appl 975–8887
18. Hira ZM, Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinf. https://doi.org/10.1155/2015/198363
19. Bania RK (2014) Survey on feature selection for data reduction. Int J Comput Appl 94:975–8887

20. Zhao J, Bao WJ, Zhang FD, Ye ZY, Liu Y, Shen MW, Zhu SM (2017) Assessing appetite of the swimming fish based on spontaneous collective behaviors in a recirculating aquaculture system. Aquac Eng 78:196–204. https://doi.org/10.1016/J.AQUAENG.2017.07.008

Author Queries

Chapter 4

Query Refs.	Details Required	Author's response
AQ1	Please check and confirm the edit made in the chapter title and running heads.	
AQ2	Please suggest whether the term 'PC's' can be changed to 'PCA's' in the sentence 'As far as the PCA concept …'.	