Purdue University
Purdue e-Pubs

International Refrigeration and Air Conditioning Conference

School of Mechanical Engineering

2016

Fault diagnosis of refrigerant charge based on PCA and decision tree for variable refrigerant flow systems

Jiangyan Liu
Huazhong University of Science and Technology, Wuhan 430074, Hubei, China, liujiangyan@hust.edu.cn

Huanxin Chen
Huazhong University of Science and Technology, Wuhan 430074, Hubei, China, chenhuanxin@tsinghua.org.cn

Jiangyu Wang
Huazhong University of Science and Technology, Wuhan 430074, Hubei, China, 1053523901@qq.com

Guannan Li
Huazhong University of Science and Technology, Wuhan 430074, Hubei, China, guannanli@hust.edu.cn

Haorong Li
University of Nebraska-Lincoln, Lincoln, 68182, NE, USA, hli3@unl.edu

See next page for additional authors

Authors
Jiangyan Liu, Huanxin Chen, Jiangyu Wang, Guannan Li, Haorong Li, and Wenju Hu

Fault diagnosis of refrigerant charge based on Decision Tree for variable refrigerant flow air-conditioning system

Jiangyan LIU[1], Huanxin CHEN[1]*, Jiangyu WANG[1], Guannan LI[1], Haorong LI[2], Wenju HU[3]

[1]Department of Refrigeration and Cryogenics, Huazhong University of Science and Technology, Wuhan, China
Contact Information (+86-027-87558330, chenhuanxin@tsinghua.org.cn)

[2]University of Nebraska-Lincoln, Lincoln, NE, USA

[3]Beijing Key Lab of Heating, Gas Supply, Ventilating and Air Conditioning Engineering, Beijing University of Civil Engineering and Architecture
Beijing, China

* Corresponding Author

ABSTRACT

Variable refrigerant flow (VRF) systems are easily subjected to performance degradation due to refrigerant leakage, mechanical failure or improper maintenance after years of operation. Ideal VRF systems should equip with fault detection and diagnosis (FDD) program to sustain its normal operation. This paper presents the fault diagnosis method for refrigerant charge faults of variable refrigerant flow (VRF) systems. It is developed based on the classification and regression tree (CART) algorithm. Data of the experimental VRF system is used to test the advantages of the CART method. Results show that the decision tree can achieve desirable diagnosis efficiency on undercharge faults, while it is less sensitive to the overcharge faults. Validation study is also conducted using the data of online VRF systems. Results implies that the CART method obtains an outstanding classification efficiency on the VRF system that has the same type as the one provides the training data. But it is unable to identify the data of different type systems

1. INTRODUCTION

Since the building energy consumption occupies the lion's share in total energy consumption, numerous researches and solutions have been implemented for building energy efficiency. For instance, LED light for building illumination, frequency conversion technology for chillers, pumps and fans in the heating, ventilation and air-conditioning (HVAC) systems, application of the building fault detection and diagnosis (FDD) techniques, etc. The HVAC system, which accounts for more than 50% of the building energy (Lin and Liu, 2015), draws the major attention of building energy efficiency enhancement. The variable refrigerant flow (VRF) system is one of the HVAC system, which has been widely installed in commercial and residential buildings. It has outstanding part load performance and energy saving potentials, as well as provide flexible zone comfort control (Aynur, et al., 2009, Lin, et al., 2015). To maintain the healthy operation of the HVAC system, the FDD techniques has been proposed to automatically identify and reject the faults. In comparison to abundant FDD researches on centrifugal chiller systems, vapor-compression air-conditioning systems, the FDD strategy studies for VRF systems are still quite limited.

Kim and Cho (2012) presented a regression algorithm to detect evaporator air blockage faults of multi-heat pump system under heating mode, the fault detection method shows desirable goodness for the multiple indoor units. Shin et al. (2014) employed two model-based fault detection methods to identify the heat exchanger fouling faults and valve stuck faults in a multi-split VRF system. Li et al. (2016) extending the virtual refrigerant charge (VRC) sensor strategy proposed by Li and Braun (2009) with the data-based analysis method. The combination of support vector regression and VRC method are proposed to predict the refrigerant charge amount in VRF systems.

As mentioned above, most FDD strategies for VRF systems are mainly developed by the model based, semi-model based or statistic based methods. while small amount of experimental data is used to training and test. Although most of the previous studies have emphasized the feasibility of the proposed FDD strategy, they have not sufficiently considered the discrepancy of the experimental system and the actual operating systems. In a real time, the actual systems or online systems operate in complex conditions rather than specific conditions in the laboratory chambers. The data of both systems might be very discrepant and the actual data would be noisy or biased, which would degrade the robustness of the FDD model that developed by the experimental data. Hence, this study aims at filling the knowledge gap. First, the data of both experimental and online systems are collected. Secondly, a typical data mining algorithm, i.e. the classification and regression tree (CART), is presented to classify the refrigerant charge faults for VRF systems. The basic ideal of CART method is described, as well as generalized the data analysis procedure. Thirdly, the data of experimental VRF system is used to training the CART model, and the model is tested by the experimental data. Further, to evaluating the robustness of the FDD strategy, validation studies are implemented with the data of online operating VRF systems.

2. MATERIALS AND METHODS

2.1 Data sources

This analysis used the data of an experimental and there online VRF systems, respectively. All VRF systems investigated in this work are made by the same manufacturer. The experimental VRF system, constructed with one outdoor and five indoor units, is installed in a standard enthalpy difference laboratory. It uses R410A as the refrigerant and the normal charge amount is 9.9 kg. The schematic is illustrated in Figure 1. The online VRF systems, which we labeled as #1, #2 and #3, are all installed in commercial buildings. The online systems operate in the actual conditions rather than specific conditions in the laboratory chambers (for example, different installed environment, unstable indoor and outdoor temperature). Besides, the #1 has the same construction as the experimental system. While the others are inconsistent, since the #2 and #3 have 8 and 12 indoor units, respectively. The data are collected from the built-in controllers in each unit by the original equipment manufacturer (OEM) sensors in both experimental and online VRF systems. In addition, the position of OEM sensors installed in the online systems are consistent to the experimental system, as the recorded parameters of the controllers are the same for both systems.

In this work, the experimental program contains nine refrigerant charge levels (RCLs). There are three cooling modes in each RCL test, i.e. low temperature, medium temperature and high temperature. Each mode includes three indoor unit operating scheme, i.e. one unit operating, three units operating and five units operating. The overall experimental program is shown in Table 1. The experimental data is used to training and test the refrigerant charge fault diagnosis model, while the online data is used to validate the robustness of the model.

For the experimental data, three categories are grouped according to 9 RCLs. The charge level between ±15% of

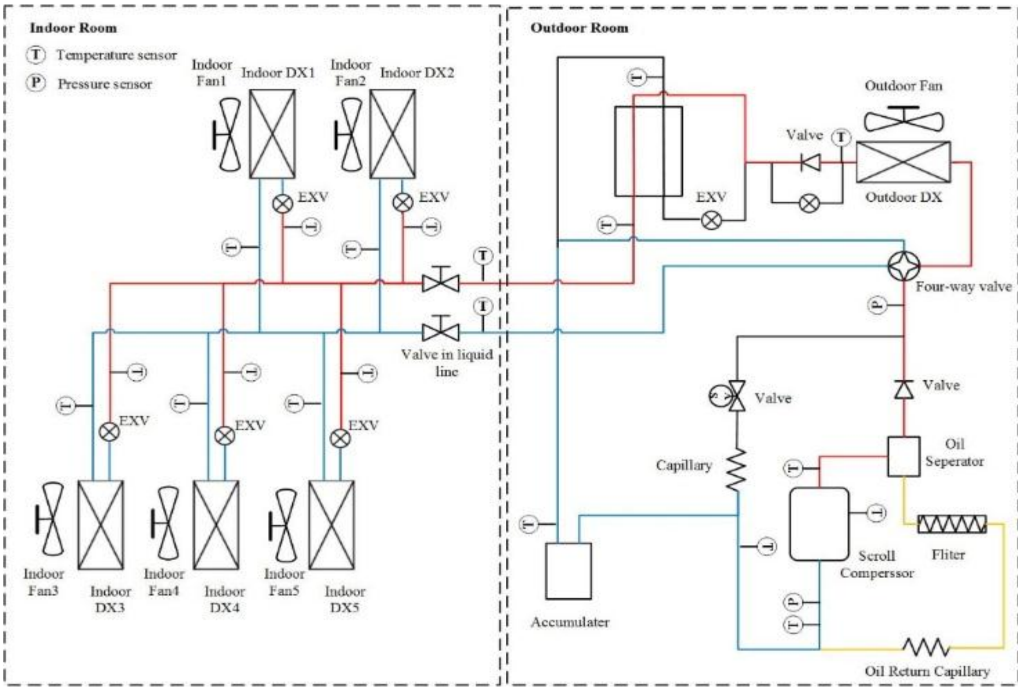


Figure 1: The schematic of experimental VRF system

the 100% refrigerant charge level is defined as "Normal charge", while the charge level less or greater than 15% of the 100% refrigerant charge level are deemed to "Undercharge" or "Overcharge", respectively. For the online data, it is grouped into "Normal charge" since each online system is proved to be normal charged after the on-site test. The data size of each RCL and online VRF system is illustrated in Table 2.

Table 1: Experimental program

Group number	Cooling mode	Indoor unit 'ON' number	Indoor room conditions		Outdoor room conditions	
			Dry-bulb	Wet-bulb	Dry-bulb	Wet-bulb
1	Low	1	23	15	31	23
2		3				
3		5				
4	Medium	1	27	19	35	24
5		3				
6		5				
7	High	1	32	24	40	26
8		3				
9		5				

Table 2: Data size of both experimental and online VRF systems

Data source	Category	Refrigerant charge level	Sample size
Experimental system	Undercharge	63.64%, 75.45%, 80.00%	51500
	Normal charge	84.84%, 95.75%, 103.74%, 111.72%	60627
	Overcharge	120%, 130%	26017
Online VRF systems	Normal charge	#1, #2, #3	6846, 7429, 14893

2.2 Decision tree

The Classification and Regression Trees (CART) is a non-parametric modeling approach that can explain the responses of a dependent from a set of independent continuous variables or categorical variables (Breiman et al., 1984). In comparison to other "black box" likely classification methods, such as support vector machine, artificial neural networks etc. it can generate accurate classification models with understandable and interpretable tree structures. Useful domain information can be extracted from the tree. The CART employs the Gini index to select properties in each node as shown in Eq. (1).

$$Gini(X) = 1 - \sum_{i=0}^{c-1} p_i^2 \tag{1}$$

where X is the data set, c is the predefined number of class, and C_i represents the ith class in X. s_i is the sample number of C_i. $p_i = s_i/S$ is the relative frequency of C_i. The CART adopts a binary recursive partitioning strategy to split the current X into two subsets, i.e. X_1 and X_2. The optimal split point can be obtained by minimizing the weighted sum of the two subsets as shown in Eq. (2).

$$\textbf{Min} \quad Gini_{split} = \sum_{i=0}^{1} \frac{n_i Gini(X_i)}{n} \tag{2}$$

where n_i is the size of the ith subset, and n is the size of the X.
The complexity parameter (cp) is employed to prune the tree and prevent overfitting, which represents the tradeoff between the tree complexity and accuracy (Themeau and Atkinson, 2011). All analyses were implemented by R software (Tippmann, 2015) and the CART model was established with the RPART package (Themeau and Atkinson, 2011).

2.3 Data analysis procedure

Figure 2 illustrates the whole data analysis process for developing the CART model. First, the data are collected from the both experimental and online VRF systems. Secondly, in the data pre-processing step, feature extraction and outlier remove are implemented on experimental and online data sets, respectively. The feature extraction based on domain knowledge is used to select proper variables (Fan, et al., 2015). The outliers of the data are filtered using the interquartile range rule (Xiao and Fan, 2014). In addition, the recorded data of system starting up process is also removed. Thirdly, in the offline model test process, all the experimental data with three labels ("Undercharge", "Normal charge" and "Overcharge") are randomly split into two subsets, i.e. training data set and test data set. The training data set contains 75% of the total data while the rest is placed into test data set. Then, 10-folds cross-validation (CV) is implemented to obtain the best fitted CART model. The training data set is split randomly into 10 disjoint subsets with equal size. Among 10 iterations, each subset would be tested by the CART model that trained with other 9 subsets. The cross-validation accuracy, which indicates the ratio of correctly classified data to the subset data, would be compared between each fold to select the optimal model. Afterword, the performance of the optimal CART model would be evaluated with the test data set. Fourthly, the robustness of the CART model is validated using the online data.

3. RESULTS AND DISCUSSION

3.1 Data pre-processing

There are 216 and 116 recorded variables in the built-in controller of outdoor unit and each indoor unit, respectively. The total 332 initial variables contain 280 state variables of the VRF systems, which are Boolean value (i.e. TRUE or FALSE) and less beneficial to the fault diagnosis model. Hence, the state variables are removed in the first place. Then, optimal 13 variables of the rest 52 variables (i.e. measurements of temperature, pressure, current, voltage, frequency and openness) were retained after the feature extraction and model validation, i.e. compressor operating frequency (f), condensing saturation temperature (T_{cond}), evaporating saturation temperature (T_{evap}), compressor, compressor discharge temperature ($T_{com,dis}$), shell temperature (T_{shell}), defrosting temperature (T_{def}), liquid refrigerant temperature at the subcooler outlet pipe ($T_{subc.out.L}$), vapor refrigerant temperature at the subcooler outlet pipe ($T_{subc.out.V}$), accumulator outlet pipe temperature ($T_{accu.out}$), accumulator inlet pipe temperature ($T_{accu.in}$), openness of the electronic expansion valve of subcooler (EXV_{subc}), compressor module temperature ($T_{com,mod}$) and compressor current (I).

Afterword, outliers of both experimental and online data are filtered by the interquartile range rule. The sizes of retained experimental data are 50483, 58119, 25318 at undercharge, normal charge and overcharge levels, respectively. Less than 10% of the total data are removed in the process. For the online data, however, in order to validate the robustness of the CART model, the initial data sets without any deletion are used for classification.

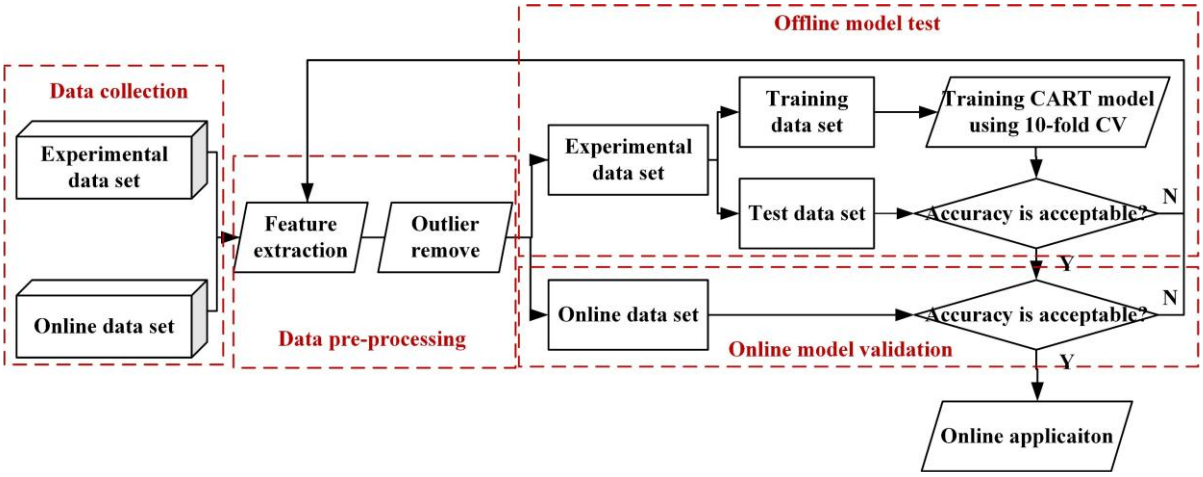


Figure 2: Flowchart of the CART method in refrigerant charge fault diagnosis

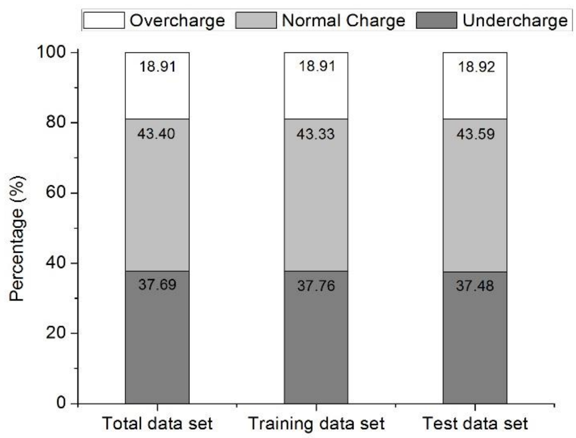


Figure 3: Label distribution of the three data sets

3.2 Test of the decision tree

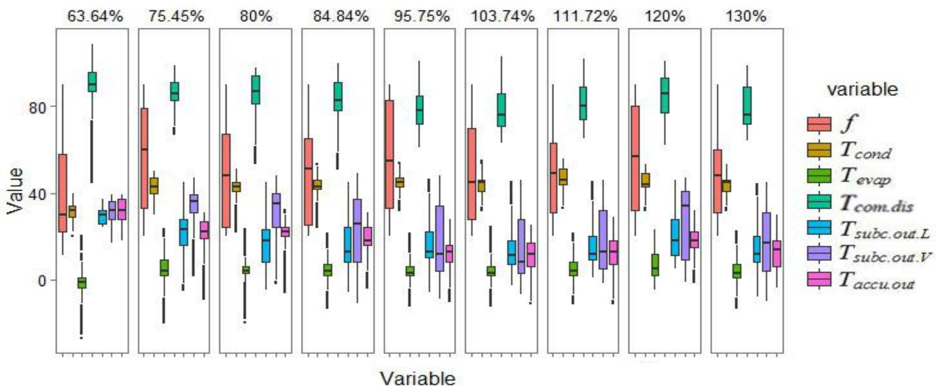
In this step, the experimental data set is grouped into the training data set and test data set. Labels in each subset have the similar distribution as the total data set as shown in Figure 2, which is benefit to the classification efficiency of the model.

The cp of the CART is set as 0.011, the overall accuracy of the training data and test data are 79.90% and 79.73%, respectively. Table 3 shows the fault diagnosis results of the test data. Sensitivity is used to evaluate the sensibility of the model, which indicates the ratio of the correctly classified data to the total data in each class. For the undercharge data, the sensitivity is 87.97%. While the normal charge data has a high sensitivity for 92.04%. For the overcharge data, however, the corrected diagnosis rate is 35.04%, since 59.76% of which is wrongly classified into "Normal charge". This is mainly because the undercharge faults affect the performance of the VRF system greatly while the overcharge faults impact little, the CART model regards the overcharge data as normal one. Figure 3 shows the boxplot of the partial variables at nine refrigerant charge levels, which illustrates the variation range of each variable. As an example, the accumulator outlet pipe temperature ($T_{accu.out}$), as shown by the pink rectangle in in Figure 3(a), increase as the reducing of the refrigerant charge level. When the refrigerant is overcharge, the $T_{accu.out}$ is remaining the similar range as the normal charge condition. Further, the accumulator in the VRF system should account for the low classification efficiency at overcharge cases. The accumulator can store the redundant refrigerant in the overcharge conditions, thus trigger less performance variation on the system. Therefore, the CART model are insufficient to identify the normal charge and overcharge data.

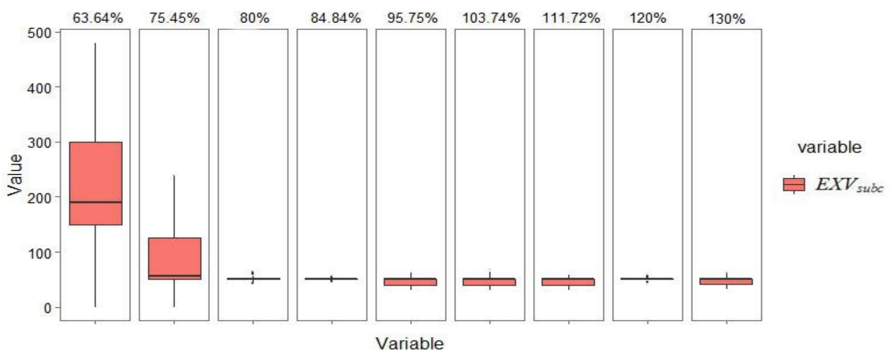
Since the overcharge faults are less recognizable by the CART model, the undercharge and normal charge fault data are used to train the model to obtain more inherent information. Figure 4 shows the decision tree for the classification of two charge classes (i.e. "Undercharge" and "Overcharge"). It can be seen that the tree contains 18 nodes as well as 10 classes (either "Undercharge" or "Overcharge"). The number of correct classifications and the number of observations are also given in each node. Results indicate that 94.44% and 94.43% of the training

Table 3: Confusion matrix of the test data set

Actual \ Prediction	Undercharge	Normal charge	Overcharge	Sensitivity
Undercharge	11042	1430	78	87.97%
Normal charge	416	13433	745	92.04%
Overcharge	329	3787	2220	35.04%



(a)



(b)

Figure 4: Boxplot of segmental variables at 9 refrigerant charge levels

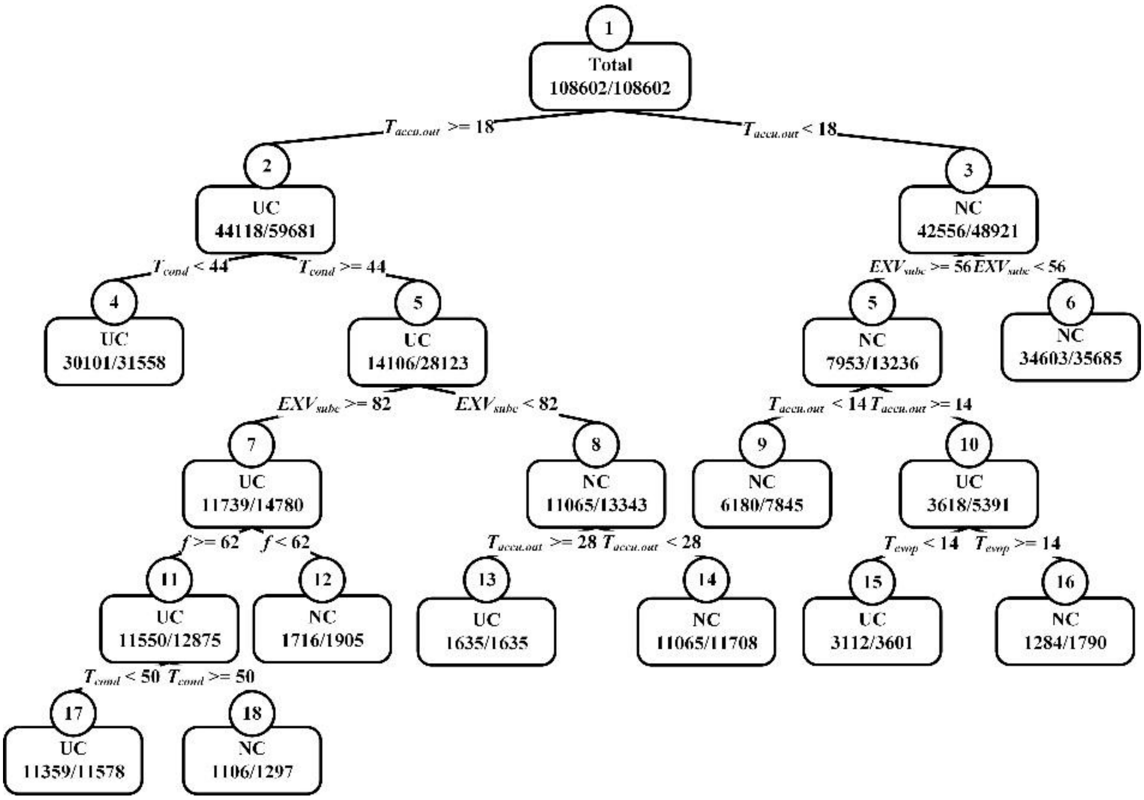


Figure 5: Decision tree for the prediction of normal charge and undercharge

Table 4: Classification result of online data

System number	Indoor unit number	Classification accuracy
#1	5	99.75%
#2	8	26.92%
#3	12	31.83%

records and test records are correctly classified, respectively. In the growing process of the tree, the $T_{accu.out}$, T_{cond} and EXV_{subc} are used repeatedly to split the nodes, which implies that the three variables are significant to identify the undercharge faults. It is reasonable as three variables ($T_{accu.out}$, T_{cond} and EXV_{subc}) vary obviously in the undercharge cases as shown in Figure 4.

According to the decision tree, some interesting information can be gained, which is in accord with the domain knowledge. For instance, as shown in node 4, if the condensing saturation temperature is lower than 44, it more likely the refrigerant is insufficient. It is consistent to the conclusion that the undercharge faults would lead to a lower condensing temperature of the system. Besides, as shown in node 6, the openness of EXV_{subc} is less than 56 indicating that the system is normal charged. It is reasonable in the actual situation, for the openness of EXV_{subc} is low in the normal charge cases as illustrated in Figure 4(b). While it would be high to meet the cooling demand in the undercharge cases.

3.3 Validation of the decision tree

In this section, the CART model based on the experimental data is validated using the data of online VRF systems, i.e. #1, #2 and #3. The CART, modeling with three refrigerant charge data, obtains 79.73% classification accuracy in the test step. Note that the online VRF systems all have the normal refrigerant charge amounts. As illustrated in Table 4, the accuracy of #1, which has the same type with the experimental system, is considerably high for 99.75%. It suggests that 99.75% of the data is identified as “Normal charge”. This is mainly because: 1) the variation ranges of monitoring variables of both experimental and #1 system are uniform, 2) the underlying relationship between variables of both VRF systems are parallel. However, for the system of #2 and #3, the correctly classified ratios are 26.92% and 31.83%. It suggests that CART model failed to identify the refrigerant charge level of both systems. Although the variation range of monitoring parameters of #2 and #3 VRF systems are similar to the experimental VRF system, the underlying relationship between variables might be different. Moreover, the result indicates that the CART method is insufficient to classify the data of VRF systems that have different number of indoor units. Therefore, the CART method is more suitable for diagnosing the refrigerant charge faults of VRF systems which have the same type as the system that provide the training data.

4. CONCLUSION

This work presents a decision tree method to diagnosis the refrigerant charge faults for VRF systems. Two CART models were established to diagnosis one (undercharge) and two (undercharge and overcharge) faults, respectively. Validation study was conducted using the data of online VRF systems. Conclusion are as follows:

1) The overcharge fault data is insufficient to be classified by the CART model, mainly because the accumulator store the redundant refrigerant to maintain the normal operation. Besides, the CART shows desirable diagnosis accuracy for undercharge faults, i.e. 93.78% for training data and 93.83% for test data.
2) The CART model obtains an outstanding classification efficiency on the VRF system that has the same type as the one provides the training data. But it is unable to identify the data of different type systems.

NOMENCLATURE

CART	classification and regression tree	(–)
c	predefined number of class in CART model	(–)
C_i	the ith class in X	(–)
CV	cross validation	(–)
EXV_{subc}	openness of electronic expansion valve of subcooler	(–)
f	compressor operating frequency	(Hz)
FDD	fault detection and diagnosis	(–)
$Gini$	Gini index	(–)
I	compressor current	(A)
n_i	size of the ith subset	(–)
n	size of X	(–)
s_i	sample number of C_i	(–)
S	sample number of X	(–)
$T_{accu.in}$	accumulator inlet pipe temperature	(℃)
$T_{accu.out}$	accumulator outlet pipe temperature	(℃)
T_{cond}	condensing saturation temperature	(℃)
$T_{com.mod}$	compressor module temperature	(℃)
T_{evap}	evaporating saturation temperature	(℃)
$T_{com.dis}$	compressor discharge temperature	(℃)
T_{def}	defrosting temperature	(℃)
T_{shell}	compressor shell temperature	(℃)
$T_{subc.out.L}$	liquid refrigerant temperature at the subcooler outlet pipe	(℃)
$T_{subc.out.V}$	vapor refrigerant temperature at the subcooler outlet pipe	(℃)
VRF	variable refrigerant flow	(–)
X	total data set	(–)

Subscript

accu	accumulator
cond	condenser
dis	discharge
def	defrosting
evap	evaporator
in	inlet
L	liquid
mod	module
out	outlet
shell	compressor shell
subc	subcooler
V	vapor

REFERENCES

B. Lin, H. Liu. (2015). China's building energy efficiency and urbanization. *Energy and Buildings*, 86, 356-365.

T.N. Aynur, Y. Hwang, R. Radermacher. (2009). Simulation comparison of VAV and VRF air conditioning systems in an existing building for the cooling season. *ENERGY AND BUILDINGS*, 41(11), 1143-1150.

X. Lin, H. Lee, Y. Hwang, R. Radermacher. (2015). A review of recent development in variable refrigerant flow systems. *SCIENCE AND TECHNOLOGY FOR THE BUILT ENVIRONMENT*, 21(7), 917-933.

H.S. Kim, M.K. Cho, M.S. Kim. (2012). Experimental Study on Fault Detection Algorithm Using Regression Method for Plural Indoor Units Faults of Multi-Heat Pump System under Heating Mode. *International Refrigeration and Air-conditioning Conference*, West Lafayette, IN

Y. Shin, S.W. Karng, S.Y. Kim. (2014). Indoor unit fault detector for a multi-split VRF system in heating mode. *International Journal of Refrigeration*, 40, 152-160.

G. Li, Y. Hu, H. Chen, L. Shen, H. Li, J. Li, W. Hu. (2016). Extending the virtual refrigerant charge sensor (VRC) for variable refrigerant flow (VRF) air conditioning system using data-based analysis methods. *Applied Thermal Engineering*, 93(908-919.

H. Li, J.E. Braun. (2009). Development, Evaluation, and Demonstration of a Virtual Refrigerant Charge Sensor. *HVAC&R RESEARCH*, 15(1), 117-136.

L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, (1984). *Classification and regression trees*, Belmont (CA): Chapman and Hall/CRC.

T.M. Themeau, E.J. Atkinson, (2011). An introduction to recursive partitioning using the RPART routine. Technical report 61. *Mayo Clinic Section of Biostatistics technical*, Mayo Foundation.

S. Tippmann. (2015). PROGRAMMING TOOLS: ADVENTURES WITH R. *NATURE*, 517(7532), 109-110.

C. Fan, F. Xiao, C. Yan. (2015). A framework for knowledge discovery in massive building automation data and its application in building diagnostics. *Automation in Construction*, 50, 81-90.

F. Xiao, C. Fan. (2014). Data mining in building automation system for improving building operational performance. *Energy and Buildings*, 75, 109-118.

ACKNOWLEDGEMENT

This research work presented in this paper was financially support by National Natural Science Foundation of China (Project No.51576074 & No.51328602), and Beijing Key Lab of Heating, Gas Supply. Ventilating and Air Conditioning Engineering (Project NR2016K02).