

Studies in Computational Intelligence

Volume 926

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are submitted to indexing to Web of Science, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink.

More information about this series at <http://www.springer.com/series/7092>

Luefeng Chen · Min Wu · Witold Pedrycz ·
Kaoru Hirota

Emotion Recognition and Understanding for Emotional Human-Robot Interaction Systems

Luefeng Chen
School of Automation
China University of Geosciences
Wuhan, China

Min Wu
School of Automation
China University of Geosciences
Wuhan, China

Witold Pedrycz
Department of Electrical
and Computer Engineering
University of Alberta
Edmonton, AB, Canada

Kaoru Hirota
Tokyo Institute of Technology
Yokohama, Japan

ISSN 1860-949X

ISSN 1860-9503 (electronic)

Studies in Computational Intelligence

ISBN 978-3-030-61576-5

ISBN 978-3-030-61577-2 (eBook)

<https://doi.org/10.1007/978-3-030-61577-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021, corrected publication 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

As robots enter into every aspect of daily life, people put forward higher requirements on robots, hoping that robots come with abilities to perceive human emotions and intentions. Such robots are called emotional robots. Their emergence will change the traditional human-robot interaction mode and realize the emotional interaction between humans and robots. Emotional robot is to use artificial intelligence methods and technologies to endue robots with human-like emotions, so that they have the ability to recognize, understand, and express joy, sorrow, and anger. The robot revolution has entered the era of “Internet + emotion + intelligence”. In the face of the urgent demand for emotional robots in the domestic and foreign markets, it is indispensable to break through the key technologies of human-robot interaction and emotional computing. Therefore, promoting the intelligent robots, make them sense the surrounding environment, to understand human emotion, intention and service demand, adaptively realize human-robot interaction with users, according to the needs of users and the change of environmental information to provide high quality service, has become the trend of the developments of a new generation of intelligent robots. Such development exhibit important research significance and evident application value. Aiming at the development needs of emotional robots and human-robot emotional interaction systems, this book introduces the fundamental concepts, system architecture, and system functions of emotional computing and emotional robot systems. The book focuses on the key technologies and scientific problems involved in the emotional robot system, such as multimode emotion recognition and emotion intention understanding, and presents the design and application examples of human-robot emotional interaction system.

This book is organized into 13 chapters. Chapter 1 introduces the basic knowledge of multimodal emotion recognition, emotional intent understanding, and emotional human-robot interaction system, and explains the complete process of emotional human-robot interaction. In Chap. 2, combined with the characteristics of facial expression, speech, and gesture, the construction method of multimodal emotional feature set is systematically described. In Chap. 3, Softmax regression-based deep sparse autoencoder network is proposed to recognize facial

emotion. Chapter 4 introduces AdaBoost-KNN using adaptive feature selection with direct optimization is proposed for dynamic emotion recognition. In Chap. 5, the weight-adapted convolution neural network is proposed to extract discriminative expression representations for recognizing facial expression. Chapter 6 presents the two-layer fuzzy multiple random forest for speech emotion recognition. In Chap. 7, the two-stage fuzzy fusion based-convolution neural network is presented for dynamic emotion recognition by using both facial expression and speech modalities. Chapter 8 presents the Dempster-Shafer theory based on multi-SVM to deal with multimodal gesture images for intention understanding. In Chap. 9, the three-layer weighted fuzzy support vector regression model is proposed for understanding human intention, which is based on the emotion-identification information in human-robot interaction. Chapter 10 proposes a two-layer fuzzy support vector regression-Takagi-Sugeno model for emotion understanding. In Chap. 11, an intention understanding model based on two-layer fuzzy support vector regression is proposed in human-robot interaction. Chapter 12 introduces the basic construction method of emotional human-robot interaction systems based on multimodal emotion recognition and emotion intention understanding. In Chap. 13, simulation experiments and application results of our emotional human-robot interaction system are shown.

We are grateful for the support of the National Natural Science Foundation of China under Grants 61973286, 61603356, 61210011, and 61733016, the 111 project under Grant B17040, and the Fundamental Research Funds for the Central Universities, China University of Geosciences (No. 201839).

We are also grateful for the support of scholars both at home and abroad. We would like to thank Prof. Jinhua She of Tokyo University of Engineering, Profs. Yong He, Weihua Cao, and Xin Chen, Assoc. Prof. Zhentao Liu of China University of Geosciences for their valuable help. Finally, we would like to express our appreciation for the great effort of graduate students Min Li, Kuanlin Wang, Wuanjuan Su, Yu Feng, Wenhao Duan, Pingping Zhang, and Wei Cao of China University of Geosciences.

Wuhan, China
August 2020

Luefeng Chen
Min Wu
Witold Pedrycz
Kaoru Hirota

Contents

1	Introduction	1
1.1	Emotion Feature Extraction and Recognition	3
1.2	Emotion Understanding	8
1.3	Emotional Human-Robot Interaction System	9
1.4	Organization of This Book	10
	References	11
2	Multi-modal Emotion Feature Extraction	15
2.1	Introduction	15
2.2	Facial Expression Feature Extraction	17
2.3	Speech Emotion Feature Extraction	18
2.4	Gesture Feature Extraction	19
2.5	Summary	22
	References	22
3	Deep Sparse Autoencoder Network for Facial Emotion Recognition	25
3.1	Introduction	25
3.2	Softmax Regression Based Deep Sparse Autoencoder Network	27
3.3	ROI Based Face Image Preprocessing	29
3.4	Expand the Encode and Decode Network	30
3.5	Softmax Regression	31
3.6	Overall Weight Training	32
3.7	Experiments	33
3.7.1	Fine-Tune Effect on Performance of Recognition	34
3.7.2	The Number of Hidden Layer Node's Effect on Performance of Recognition	36
3.7.3	Recognition Rate	37

3.8	Summary	37
	References	38
4	AdaBoost-KNN with Direct Optimization for Dynamic Emotion Recognition	41
4.1	Introduction	41
4.2	Dynamic Feature Extraction Using Candide-3 Model	43
4.3	Adaptive Feature Selection Based on Plus-L Minus-R Selection	44
4.4	AdaBoost-KNN Based Emotion Recognition	45
4.5	AdaBoost-KNN with Direct Optimization for Emotion Recognition	47
4.6	Experiments	48
4.6.1	Experimental Environment and Data Selection	48
4.6.2	Simulations and Analysis	49
4.6.3	Preliminary Application Experiments on Emotional Social Robot System	52
4.7	Summary	54
	References	54
5	Weight-Adapted Convolution Neural Network for Facial Expression Recognition	57
5.1	Introduction	57
5.2	Facial Expression Image Preprocessing	59
5.3	Principal Component Analysis for Extracting Expression Feature	60
5.4	Weight-Adapted Convolution Neural Network for Recognizing Expression Feature	60
5.4.1	Feature Learning Based on Deep Convolution Neural Network	61
5.4.2	Softmax Regression for Feature Recognition	62
5.5	Hybrid Genetic Algorithm for Optimizing Weight Adaptively	62
5.6	Experiments	64
5.7	Summary	71
	References	72
6	Two-Layer Fuzzy Multiple Random Forest for Speech Emotion Recognition	77
6.1	Introduction	77
6.2	Feature Extraction	79
6.3	Fuzzy-c-Means Based Features Classification	81
6.4	Two-Layer Fuzzy Multiple Random Forest	82

6.5	Experiments	82
6.5.1	Data Setting	83
6.5.2	Environment Setting	83
6.5.3	Simulations and Analysis	84
6.6	Summary	87
	References	88
7	Two-Stage Fuzzy Fusion Based-Convolution Neural Network for Dynamic Emotion Recognition	91
7.1	Introduction	91
7.2	Dynamic Emotion Feature Extraction	93
7.3	Deep Convolution Neural Network for Extracting High-Level Emotion Semantic Features	95
7.4	Feature Fusion Based on Canonical Correlation Analysis	96
7.5	Decision Fusion Based on Fuzzy Broad Learning System	97
7.6	Two-Stage Fuzzy Fusion Strategy	101
7.7	Experiments	101
7.7.1	Data Setting	101
7.7.2	Experiments for Hyperparameters	102
7.7.3	Experimental Results and Analysis	107
7.8	Summary	112
	References	113
8	Multi-support Vector Machine Based Dempster-Shafer Theory for Gesture Intention Understanding	115
8.1	Introduction	115
8.2	Foreground Segmentation and Feature Extraction	117
8.2.1	Foreground Segmentation Based on Depth and RGB Images	117
8.2.2	Speeded-Up Robust Features Based Gesture Feature Extraction	118
8.3	Encoding Speeded-Up Robust Features: Sparse Coding	119
8.4	Multi-class Linear Support Vector Machines	119
8.5	Dempster-Shafer Evidence Theory for Decision-Level Fusion	120
8.6	Experiments	122
8.6.1	Experimental Setting	122
8.6.2	Experimental Environment and Setup	124
8.6.3	Experimental Results and Analysis	124
8.7	Summary	129
	References	129

9	Three-Layer Weighted Fuzzy Support Vector Regressions for Emotional Intention Understanding	133
9.1	Introduction	133
9.2	Support Vector Regression	136
9.3	Three-Layer Fuzzy Support Vector Regression	137
9.4	Characteristics Analysis of Emotional Intention Understanding	140
9.5	Three-Layer Fuzzy Support Vector Regression-Based Intention Understanding Model	141
9.6	Experiments	142
9.6.1	Experiment Setting	143
9.6.2	Experiments on Three-Layer Fuzzy Support Vector Regression Based Intention Understanding Model	143
9.7	Summary	156
	References	157
10	Dynamic Emotion Understanding Based on Two-Layer Fuzzy Support Vector Regression-Takagi-Sugeno Model	161
10.1	Introduction	161
10.2	Dynamic Emotion Recognition Using Candide3-Based Feature Point Matching	164
10.3	Two-Layer Fuzzy Support Vector Regression for Emotional Intention Understanding	166
10.4	Two-Layer Fuzzy Support Vector Regression Takagi-Sugeno Model for Emotional Intention Understanding	169
10.5	Experiments	171
10.5.1	Experimental Environment	171
10.5.2	Self-Built Data	171
10.5.3	Experiments on Dynamic Emotion Recognition and Understanding	173
10.6	Summary	180
	References	180
11	Emotion-Age-Gender-Nationality Based Intention Understanding Using Two-Layer Fuzzy Support Vector Regression	183
11.1	Introduction	183
11.2	Two-Layer Fuzzy Support Vector Regression	185
11.2.1	Support Vector Regression	185
11.2.2	Two-Layer Fuzzy Support Vector Regression	187
11.3	Intention Understanding	189
11.3.1	Emotion Based Intention Understanding	189
11.3.2	Characteristics Analysis	190

11.4	Intention Understanding Model	197
11.4.1	Emotion Recognition	198
11.4.2	Questionnaire	199
11.4.3	ID Mapping	199
11.5	Two-Layer Fuzzy Support Vector Regression Based Intention Understanding	200
11.5.1	Intention Generation by Fuzzy Inference	201
11.6	Memory Retrieval for Intention Understanding	201
11.7	Experiments	202
11.7.1	Experiment Setting	202
11.7.2	Experiments on Two-Layer Fuzzy Support Vector Regression Based Intention Understanding Model	204
11.8	Summary	210
	References	212
12	Emotional Human-Robot Interaction Systems	215
12.1	Introduction	215
12.2	Basic Emotional Human-Robot Interaction Systems	218
12.3	Design of Emotional Human-Robot Interaction System	220
12.4	Summary	221
	References	221
13	Experiments and Applications of Emotional Human-Robot Interaction Systems	223
13.1	Introduction	223
13.2	Emotional Interaction Scenario Setting	225
13.3	Multi-modal Emotion Recognition Experiments Based on Facial Expression, Speech and Gesture	227
13.4	Emotional Intention Understanding Experiments Based on Emotion, Age, Gender, and Nationality	232
13.5	Application of Multi-modal Emotional Intention Understanding	238
13.5.1	Self-built Data	238
13.5.2	Experiments on Dynamic Emotion Recognition and Understanding	239
13.6	Summary	243
	References	243
	Correction to: Dynamic Emotion Understanding Based on Two-Layer Fuzzy Support Vector Regression-Takagi-Sugeno Model . . .	C1
	Index	245