International Journal of Information Technology & Decision Making
Vol. 19, No. 1 (2020) 5–47

DOI: 10.1142/S0219622019300052

A Review on Human-Computer Interaction and Intelligent Robots

Fuji Ren*,†,‡ and Yanwei Bao*,§

*School of Computer Science and Information
Engineering of Hefei University of Technology
Key Laboratory of Affective Computing and Advanced Intelligent Machine
Hefei 230009, Anhui Province, P. R. China

†Tokushima University
Graduate School of Advanced Technology & Science
Tokushima 7708502, Japan
‡Ren2fuji@gmail.com
§Baoyanwei007@sina.com

Published 17 February 2020

In the field of artificial intelligence, human–computer interaction (HCI) technology and its related intelligent robot technologies are essential and interesting contents of research. From the perspective of software algorithm and hardware system, these above-mentioned technologies study and try to build a natural HCI environment. The purpose of this research is to provide an overview of HCI and intelligent robots. This research highlights the existing technologies of listening, speaking, reading, writing, and other senses, which are widely used in human interaction. Based on these same technologies, this research introduces some intelligent robot systems and platforms. This paper also forecasts some vital challenges of researching HCI and intelligent robots. The authors hope that this work will help researchers in the field to acquire the necessary information and technologies to further conduct more advanced research.

Keywords: Human–computer interaction; intelligent robots; ective computing.

1. Introduction

Artificial intelligence (AI) technology is a technical science that studies and develops theories, methods, technologies, and application systems for the simulation, extension, and expansion of human intelligence. It has been one of the most popular and widely growing technologies in recent years and has already achieved significant success in many areas such as robots, speech recognition, computer vision, and natural language processing.[1–4] AI is regarded as the most valuable technology, which holds the highest potential to achieve many breakthroughs. It attempts to understand the essence of intelligence and produces intelligent machines that can respond in the

similar form of human intelligence. It signifies that intelligent machine (or robots, agents) with human-like intelligence is the ultimate goal and carrier of AI technology.

Human intelligence is the intellectual prowers of humans, which is marked by performing and solving complex cognitive feats and also including high levels of motivation and self-awareness.[5] Intelligence enables humans to learn, apply logic, reason, recognize patterns, make decisions, solve problems, and think. Through their intelligence, humans possess the cognitive abilities to perceive the world, comprehend truth and good things, and interact with surrounding people and environments through perception, understanding, reasoning, and expressing. Humans can hear the beautiful voices and melodies, read classic literature and masterpiece, gaze at refreshing scenery and artwork, and feel a rich world and affection. Then, they can tell the difference, know what those mean, and express themselves by making dialogs, writing articles, painting, and any other possible ways of expression. And through these processes, humans are able to apply their intelligence and interact in different ways.

The expeditious development in the fields of AI, deep learning technology, intelligent robot, and human–computer interaction (HCI) has achieved a substantial progress in recent years. Now, the intelligent robots possess more and more human-like intelligence and ability, such as the ability of listening, speaking, reading, writing, vision, feeling, and consciousness.[6–9]

The purpose of this paper is to provide a comprehensive overview of technologies involved in the technologies of intelligent robots and HCI, including natural language understanding (NLU), computer vision, deep neural network, and wearable devices. There has been a huge amount of innovative works conducted on intelligent robots and HCI in the AI literature. However, this paper only focuses on the latest research advancements, which are oriented toward interaction between each other. This interaction is vital and closely related to intelligence. Through this research work, the authors hope to provide the intelligent robots and HCI community with useful reference resources.

The rest of the review is organized as follows: in the next section, we will first deliver a general review of intelligent robots and HCI, including its history, definition, and categorization. We will, then, review some current research works on the topic of intelligent robots and HCI from the perspective of abilities that should be mastered by the intelligent robots for aspiring a natural and harmonious HCI environment and experience. In Sec. 4, we will summarize the research processes about affective computing, which is considered to be one of the most important challenges in the field of intelligence. Then, in Sec. 5, we will introduce some successful applications of this topic. Finally, we will discuss some key scientific problems in these fields and conclude the paper with a formulation of future works (including its recommendations) in Sec. 6.

2. Overview

Generally speaking, the technologies of HCI and intelligent robots encompass a huge amount of research fields. Figure 1 summarizes the functions that intelligent robot

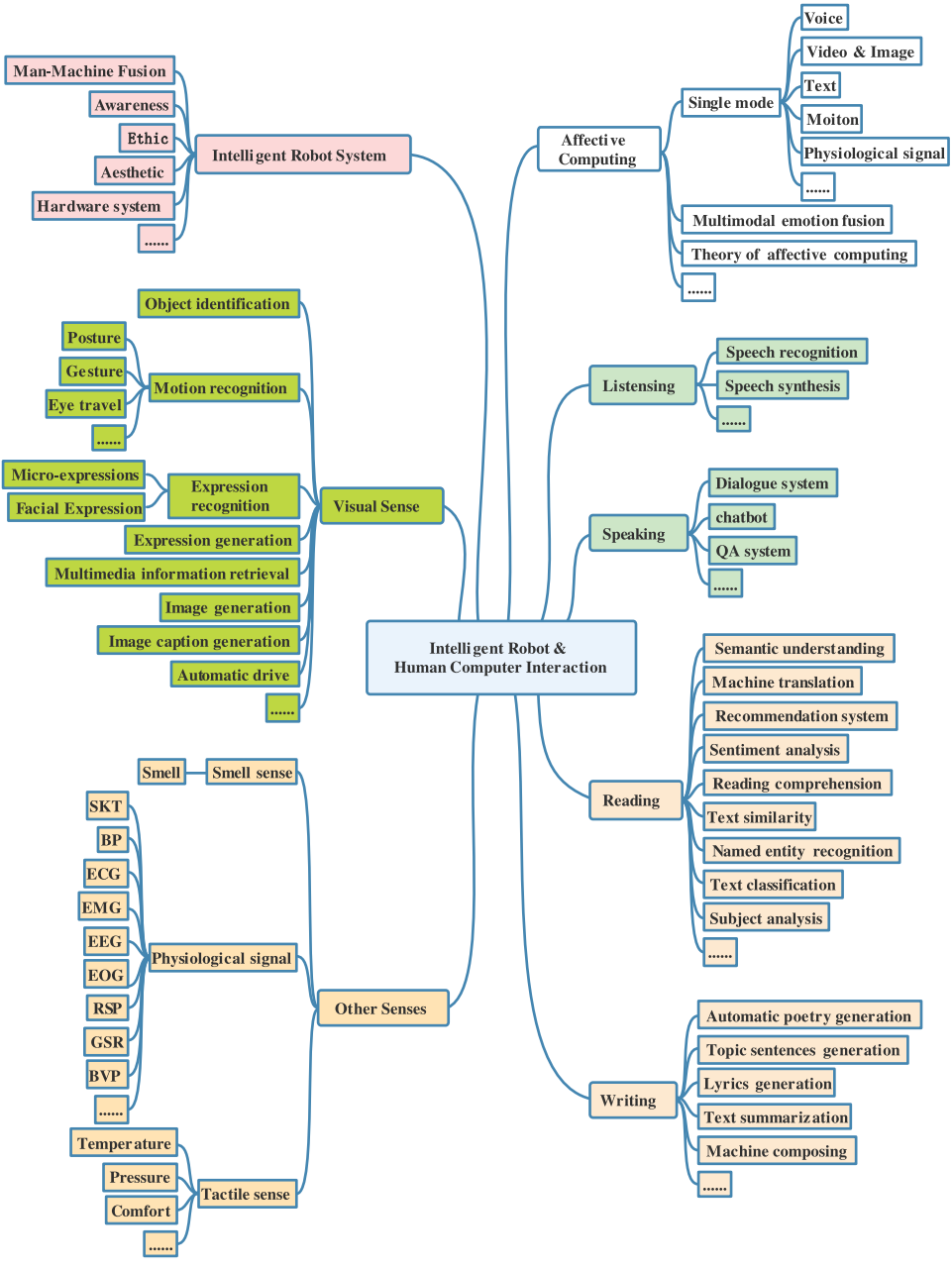

Fig. 1. Subtasks of HCI and intelligent robot system.

should possess for man–machine interactions, and our review will also include literatures in most of the fields in this figure.

2.1. *Definition and categorization of HCI*

HCI or human–machine interaction (HMI) is the syncretic science of computer science, design, behavioral science, AI, and several other subjects, which involves a thorough research of the scientific implications and practices of the interfaces between people and computers or intelligent agents. There are two levels of meaning associated with the related research works (see Fig. 2). On the primary level, it includes the research of ways and design of new technologies to (better) promote the computers as useful tools, whereas on the higher level, it includes the research of intelligent technologies that will adopt the natural ways of interaction between humans and computers, thereby boosting the cause for the computers to become more harmonious as partners to get along with. HCI was first used in 1976,[10] and it was popularized by the book, *The Psychology of Human–Computer Interaction* published in 1983.[11] In 1992, a HCI curriculum was developed by Hewett and other leading HCI educators to serve the needs of the HCI community.[12] In CES 2008, Bill Gates emphasized the role of natural user interface and predicted that the way in which HCI will bring a radical change in the next few years. Thereafter, HCI researchers expounded the definition of a natural HCI by employing different approaches.[13–15]

As far as we know, the development process of HCI has gone through five major stages: manual stage, interactive command language stage, graphical user interface (GUI) stage, network user interface stage, and natural HCI. As their names imply,

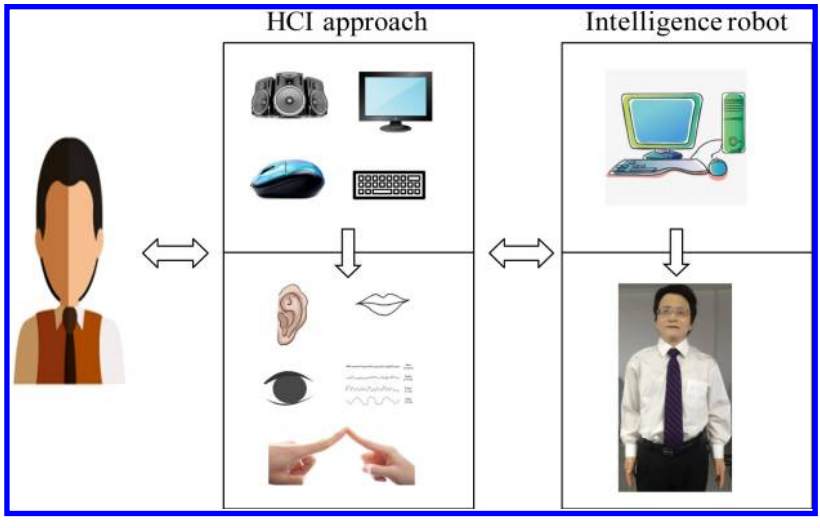

Fig. 2. The transformation of interaction approach and intelligent robot.

we can understand the characteristics of each stages. A situation of tripartite confrontation exists in this field. GUI is still the basis of the HCI platform, due to which, the network user interface is going through a vigorous development with the emergence of large number of network technologies and applications, such as search engines, social media, etc. Simultaneously, due to its characteristics of interaction such as directness, naturalness, and parallelism, natural HMI has shown prominent chances of survival to become the next emerging frontier to be researched and developed in this field. It seems that natural HCI technologies will lead the next generation of interactive technologies. In fact, natural HCI is not a new concept, and it has been in existence for a considerable amount of time and is constantly developing itself since the emergence of the computer. People want to use the simplest and most effective way to control computer to achieve task completion. However, in the present era, people want to use the most direct and natural way for the computer to provide more services, that is, the computer is hoped to be more intelligent.

2.2. *Definition and categorization of intelligent robots*

The definition of intelligence is controversial, and a series of different definitions are provided by experts from different fields, such as from psychology, philosophy, etc. to AI researchers and so on.[16] There are many works[17–21] that have defined intelligence from different aspects.

In general, intelligence is the ability to perceive or learn and understand or reason to form new knowledge and deal with new situations. In more intuitive words, intelligence is the ability of learning to know from nothing to anything, reasoning to understanding, this to the other, generalizing to transferring and from being general to special. It also involves interaction with the surroundings and adaptation to the environments. The robots having the above-mentioned abilities are named as intelligent robots.

According to the degree of their intelligence, robots can be divided into two categories: functional robots and intelligent robots. Before the advent of intelligent robots, robots were primarily referred as functional robots, whose main purpose was to perform actions that humans would not want to do and cannot do on their own. They were treated as tools to improve work efficiency and emancipate humans from manual labor and simple mental labor. These robots possess characteristics, such as high harmfulness, high strength, high speed, and monotonicity, which humans are unable to cope daily. Intelligent robots were invented to meet the demands of human intelligence, such as intelligence quotient (IQ) and emotional quotient (EQ). From the view of intelligence development stage, these robots are categorized as follows: cognitive robots, understanding robots, interactive robots, and autonomous robots. The abilities of acquisition, representation, access, and handling of data and knowledge serve as the main difference between functional robots and intelligent robots. According to the application field, intelligent robots can be divided into

industrial robots, domestic robots, medical robots, military robots, education robots, entertainment robots, etc. Additionally, independent interaction ability and the ability of emotion cognition and expression are also the essential characteristics of intelligent robots.[22,23]

3. Affective Computing

By using the abilities of perception, deduction, and prediction, intelligent robots or computers are involved in a large number of tasks in our daily life. A plethora of evidence demonstrates that the calculation capacity and IQ of intelligent robots have gone far beyond the reach of humans, but it still lacks to confirm that these robots possess human-like intelligence. The key issue is that these robots are not similar to humans from the perspective of emotions, and their EQs have been nil as compared to humans. It is well known that emotion is a necessary factor for communication and interaction between humans. Therefore, people naturally expect intelligent robots to also have the ability of emotional interaction in the HCI process, that is, an intelligent robot should have EQ along with IQ.

In 1985, Marvin Minsky, one of the founders of AI, put forward in the book *The Society of Mind* that "The question is not whether intelligent machines can have emotions, but whether machines can be intelligent without any emotions."[24] Thereafter, emotions gradually became the consensus of the AI professionals as an important part of intelligence. The research of endowing intelligent machines' abilities of understanding, expressing, and reproducing emotions has been widely carried out, which mainly includes affective computing, Kansei engineering, and artificial psychology.

Affective computing will improve a natural HCI environment and expand the application scenarios for intelligent robots. Researchers have carried out extensive innovative works focusing on affective computing and achieved a wealth of research findings. A relatively complete processing procedure and some theoretical systems have also been established, such as mechanism and theoretical modeling of emotions, emotional information acquisition, emotion recognition, emotion understanding, and emotion expression.

In 1997, Picard in MIT put forward the concept of affective computing for the first time. She pointed out that affective computing relates to, arises from, or deliberately influences emotions or other affective phenomena.[25] The purpose of affective computing is to promote the EQ of intelligent robots and equip them with an emotional "heart" so that they can develop the human-like capacities of perception, understanding, and generating a variety of emotional characteristics, and, then, create a natural and harmonious HCI system. Affective computing is the theory basis of realizing natural and harmonious HCI, and is also an extremely challenging research topic in the field of AI.

Nagamachi created Kansei engineering based on the research of affective engineering in 1988.[26] In 1999–2000, researchers had put forward the theory of "Artificial

emotion" and "Artificial psychology".[27,28] Based on information science, artificial psychology is the theory and methodology for intelligent machines stimulating people's psychological activity, such as emotions, volition, and personality.[29]

Emotion is an adaptive physiological expression that humans produce spontaneously when they are buoyed by the external environment in daily life activities. Research results of anatomical and behavioral sciences suggest that emotional activities and expressions are under the control of human brain. Studies have found that emotion is associated with multiple brain regions, including the prefrontal cortex, hypothalamus, and cingulate cortex, and amygdala serves as the center of all emotions.[30,31]

Researchers generally use discrete emotional states model and dimensional model to construct and understand the emotional space. The discrete emotional states model divides emotions into a variety of discrete states, which can be further divided into several different emotional states (e.g., happiness or disgust).[32] The most common classification scheme is dividing it into six emotional states: happiness, sadness, anger, fear, surprise, and disgust. Human emotional states are continuous and dynamical in a natural interaction scene, so the discrete emotional states model is unable to accurately represent the change of human emotions.[33] The dimensional model considers the emotional space as a continuous space composed of different dimensions, which can better characterize and stimulate human emotions.[34] There are two-dimensional valence-arousal model[35] and activation-evaluation model[36] besides the three-dimensional pleasure-arousal-dominance model[37] and arousal-valence-stance model.[38] Reference 39 proposed a new academic system called "Enriching Mental Engineering", which aims to deal with the mental system of human beings. It measures and enriches the mental richness by employing engineering methods. Reference 40 carried out research on affective computing from the view of psychology and proposed a mental state transition network model to dynamically detect human emotions. After that, these researchers conducted a series of experiments involving basic theories, emotional data resources construction, and their applications.[41–44] Table 1 summarizes the above reviewed models. In addition, there are a large amount of literatures available on the applications of affective computing in the field of HCI and intelligent robots, which will be reviewed in the following sections.

Table 1. Emotional states model.

Discrete model	Dimensional model		Dynamic discrete model
	Two-dimensional model	Three-dimensional model	
Discrete states	Valence-Arousal	Pleasure-Arousal-Dominance	Mental State Transition Network
	Activation-Evaluation	Arousal-Valence-Stance	

4. Human–Computer Interaction

In this section, we will review the relevant literatures in the field of HCI by considering the aspect of interactional abilities, such as listening, speaking, reading, writing, visual sense and other senses, possessed by humans. These same activities are desired in an intelligent robot.

4.1. *Listening and speaking*

Auditory sense is one of the most important senses of the human body. It is used for mutual interaction among humans and its main forms include listening and speaking. Listening is used to receive the voices of outside world, and speaking is used to express own ideas and opinions to the outside world. The robot's abilities of listening and speaking aim to imitate the auditory ability of humans in the interaction process, and these two kinds of abilities are carried out via the spoken dialogue system in intelligent robots. Figure 3 shows the framework of a spoken dialogue system. Generally speaking, the spoken dialogue system comprises five modules: automatic speech recognition (ASR), NLU, dialogue management (DM), natural language generation (NLG) and automatic speech synthesis (ASS).

The primary responsibility of the ASR is to transform the continuous time signal of a user's speech into a series of discrete syllable units or words. The primary responsibility of NLG is to analyze the result of speech recognition process and transform the user's dialogue information into a representative form that can be utilized by the dialogue system via syntactic and semantic analysis. DM is used to make a comprehensive analysis based on the result of language understanding, the context of the dialogue, the historical information of the dialogue, etc., to determine the current intention of the user. Thereafter, the response or response strategy is adopted by the system. Then, NLG organizes the appropriate response statement and convert the system's response into the natural language that users can understand. The primary responsibility of ASS is to synthesize the text generated by NLG into the final answering voice and feed it back to the user. A large number of extensive efforts have been actualized in the field of dialogue system, which is divided into two categories, acoustic-based and text-based.

One of the key terminals of the auditory module is ASR, which has changed the way we interact with intelligent agents/systems. The development of ASR benefits from both fields of academic research and industry, including Google, Microsoft,

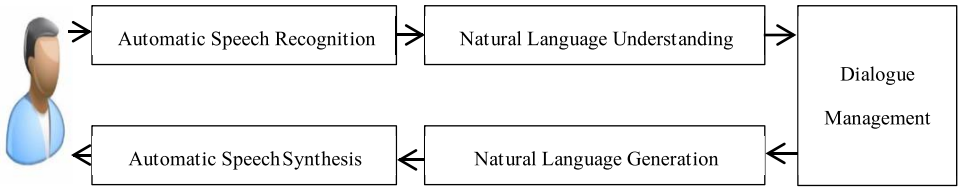

Fig. 3. The framework of spoken dialogue system.

IBM, Baidu, Amazon, iFLYTEK, etc., all of which have developed speech recognition engines. In a traditional solution, hidden Markov models (HMMs) are widely used in speech recognition systems, and most of modern general-purpose speech recognition systems are based on HMMs.[45] HMMs are used in speech recognition process because a speech signal can be viewed as a piecewise stationary signal or a short-time stationary signal. This signal can be considered suitable for the Markov model based on the hypothesis of HMMs that hidden state variables, speech as an observed value, and the transfer among states conform to the hypothesis of HMMs.

Reference 46 introduced the application of the theory of probabilistic functions of a (hidden) Markov chain to actualize ASR for an isolated word. Following that, Ref. 47 described a maximum likelihood approach to the continuous speech recognition process. Also, there are many other works that focus on probability models and HMMs for ASR.[48–50] According to the descriptive ways of observation probability, there are two HMM-based models: CD-GMM-HMM architecture[51] and CD-DNN-HMM architecture.[52]

References 53 and 54 were committed in solving the problem of phoneme recognition and classification by using the neural network models. Reference 55 had reviewed the time delay neural network architectures for speech recognition process. In the last five years, the research works of speech recognition had focused on deep neural networks-based methods, such as CNN,[56,57] LSTM,[6,58] and RNN.[59,60] ASR is also researched as an end-to-end problem, and many works have showed that end-to-end deep learning-based methods had obtained encouraging results.[61–64]

Apart from the speech content, the speech also carries the rich emotions of the speaker. In most natural interactions, we not only need to know the contents of speech, but more importantly, we need to know the emotions present in the speech, which are also an important part of the natural HCI research. Therefore, many researchers have been focusing on the emotional recognition of speech to mine the emotion labels of speech, so that the emotional information can be used by other interactive tasks.[65–67]

ASS is the other terminal of an auditory module. Released in 1975, Multichannel Speaking Automaton was considered as one of the first ASS systems, whereas the Bell Labs system was one of the first multilingual language-independent systems, which made an extensive use of natural language processing methods.[68] The major recognized classification ways of speech synthesis methods are rule-driven methods and data-driven methods according to the design idea[69] (see Fig. 4 for details). The main principle of rule-driven methods is to simulate the physical process of human pronunciation by establishing a series of rules. The resonance peak synthesis method[70] and pronunciation simulation-based synthesis method are rule-driven methods.

The data-driven synthesis methods mainly include concatenative synthesis method, HMMs-based method, and deep neural networks-based method. The concatenative synthesis method synthesizes sounds by identifying and concatenating the units that best match the specified criterion, further accompanied by prosodic modification.[71] The difficulty and deficiency of this kind of approach is that speech

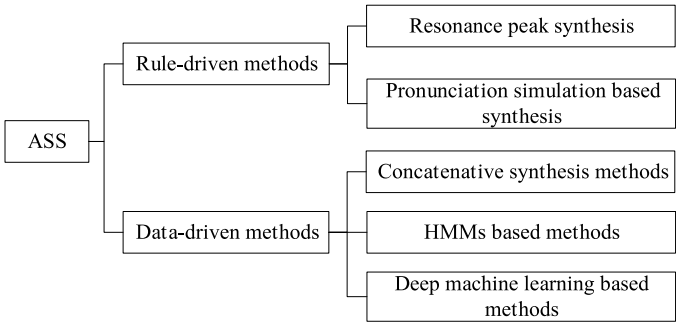

Fig. 4. Classification of speech synthesis methods.

corpus consumes lot of resources and requires a sophisticated design. HMMs- and STRAIGHT-based method overcame this barrier and is suitable for the mobile-embedded platform.[72–74] The deep learning network was first applied in the field of speech recognition, and the recognition rate increased by more than 10%, which greatly attracted the attention of researchers. There are also abundant research achievements in the field of speech synthesis with the use of deep neural networks.[75–83] In a conventional neural networks-based approach, text analysis and acoustic modeling are processed separately. However, Ref. 75 attempted to integrate them together and proposed a novel end-to-end framework to deal with speech synthesis. By combining memory-less modules and stateful recurrent neural networks, the unconditional audio generation in the raw acoustic domain was researched in Refs. 75 and 76. Reference 77 introduced WaveNet to generate the raw audio waveforms and yielded a state-of-the-art performance after applying the same to speech synthesis. References 78–80 have carried out a series of meaningful works in this field based on deep neural networks. References 81 and 82 had focused their works on vocoder-based speech synthesis system to improve the sound quality and real-time performance of speech synthesis. Some research works also aim to synthesize the speech of a specific type or person.[74,83] Reference 83 introduced an emotional speech synthesizer based on the end-to-end neural model, which could be used to generate speech for the given emotion labels. Reference 84 used Variational AutoEncoder (VAE) to synthesize speech to control it in an unsupervised manner. Certain types of speech synthesis tasks, especially emotional speech synthesis task, are of great significance and value, which can affect the content and effect to be expressed, because the effect will be greatly different when the same content is expressed by different emotional semantics.

Although the quality of speech synthesis has steadily improved over the past decades, especially with the rapid development of deep neural network technology, speech synthesis systems remain clearly distinguishable from the natural human speech. The challenges of emotional speech synthesis and natural language processing accompanied by speech synthesis are still in an urgent need to be addressed and solved.

Another research direction of acoustic-based work related to HCI is Voiceprint Recognition (VPR). In the natural HCI scenario, intelligent robots need to know what the interactive person says, and be more natural. Intelligent robots must understand that the identity of the interactive person is also essential to be learned, so that it can adjust its way of speaking according to the speaker's personality. There are two application scenarios of VPR: speaker identification and speaker verification. The former is used to determine one of several peoples who speak a particular speech, whereas the latter is used to confirm whether a speech is spoken by a specified person. In 1995, Reynolds successfully applied the Gaussian mixture model (GMM) to the text-independent VPR task for the first time[85] and established the foundation position of GMM in the acoustic pattern recognition.[86,87] The traditional acoustic features including MFCC, PLP, and PNCC[88] can be used as acoustic features in the VPR task. Also, there are works that have focused on deep learning and i-Vector-based VPR.[89–91]

Actually, the text-related research works between ASR and ASS are at the core of the dialogue system. Many AI companies have launched a series of new information services based on the dialogue system, such as Google's ALLO, Apple's Siri, Microsoft's Cortana, and Baidu's Duer. It can be divided into task-oriented and non-task-oriented dialogue systems based on whether the dialogue system can achieve a specific goal (see Fig. 5 for details). Also, the above-mentioned dialogue systems are both task-oriented and non-task-oriented.

There are pipeline-based methods and end-to-end methods for task-oriented dialogue systems. The pipeline method includes NLU, dialogue state tracking, policy learning, and NLG. NLU is used for topic recognition, intention mining, and semantic annotation in the dialogue system. Topic recognition and intention mining are usually considered as classification tasks, and a number of studies have been published in these fields.[92–97] References 7, 98 and 99 focused on semantic annotation (slot filling), which is a challenge of sequential annotation for words. In recent years, the primary responsibilities of dialogue state tracking are mainly focused on deep neural network-based methods.[100–102] The DSTC: Dialog State Tracking Challenge, which has been held annually since 2013, has given a strong impetus to the study of dialogue state tracking. Deep reinforcement learning is often used in policy learning,[103,104] and other approaches have also been tried in policy learning.[105] NLG

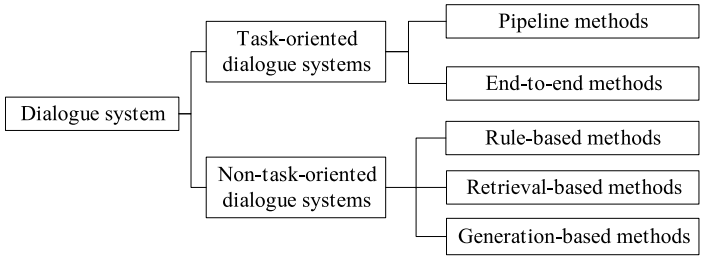

Fig. 5. Classification of dialogue system methods.

is used to generate dialogue responses under the guidance of the dialogue strategy, whose generation ways contain generative models-based methods[106,107] and retrieval-based methods.[108,109] References 110–112 have researched about affective DM, which is one of the cores of dialogue system. Unlike pipeline methods, end-to-end methods had treated the dialogue system learning as the problem of learning a mapping from dialogue histories to system responses, and applied an encoder–decoder model to train the whole system.[113,114]

Non-task-oriented dialogue systems are often called chatbot, whose main purpose is to provide the ability to chat with people in an open domain, and there are rule-based methods, retrieval-based methods, and generation-based methods. In fact, we can think of it as the joint modeling of all modules in the pipeline-based methods. In recent years, the research works in this field are mainly concentrated upon deep neural network methods-based generation models. Referring to the Seq2Seq model of machine translation, multiple end-to-end response systems based on the deep neural network model emerged in 2015.[115–117] Thereafter, the attention mechanism is introduced to generate context-sensitive dialogue responses.[118] In addition, the research works in this field also include deep reinforcement learning-based dialogue generation,[119] dialogue generation model study based on VAE and CVAE[120,121] and dialogue generation study based on GAN.[122,123]

Affective computing and dialogue systems are two emerging and interesting research directions in the field of AI. Many scholars have conducted a lot of research works on these two aspects, respectively; however, the researched content is basically independent and less related to each other. With the gradual perfection of dialogue system and the comprehensive deepening of affective computing, some scholars have begun to explore a new cross-research topic, that is, how to integrate the emotion into the dialogue system to build an emotional dialogue system.[124] Reference 125 combined the affective computing theory with the spoken language dialogue system and proposed to use the spoken language dialogue system as the carrier for the integration of the multi-modal emotion recognition, effective emotional interaction, and the emotion generation and expression of intelligent robots. The generation of emotional dialogue responses is mainly achieved by learning emotional labels.[22,126]

Question answering system, which is focused more on factual questions, can be regarded as a special case of the dialogue system, and it can answer the questions posed by humans with more accurate and concise natural language. There are also plenty of good works in the field. Reference 127 proposed a distantly supervised open-domain question answering (DS-QA) system, which retrieves the relevant text from Wikipedia and extracts the answer by reading comprehension. Reference 128 proposed a denoising DS-QA, which contains a paragraph selector and paragraph reader to make the full use of all informative paragraphs and alleviate the wrong labeling problem in DS-QA. Reference 129 proposed a method of answer extraction for long documents, which separated the answer generation in DS-QA into selecting a target paragraph in document and extracting the correct answer from the target paragraph

by reading comprehension. Reference 130 proposed a Question Condensing Networks (QCN) to utilize the subject–body relationship of community questions.

4.2. *Reading & writing*

Another form of mutual human interaction involves characters, such as reading a book, writing a letter, etc. People express their thoughts, love, for example, by using characters. Then, readers can understand the meaning and thoughts deeply present in characters by reading them. Enduing the intelligent robots with the abilities of reading and writing is still in the category of natural language processing, whose purposes are to enable robots to read human characters and understand human thoughts, and to express their thoughts and ideas by generating a specific character sequence. In the following content, several tasks will be introduced to reflect the robots' abilities of reading and writing (see Table 2 for summaries), including part-of-speech tagging, named entity recognition, text classification, text sentiment analysis, machine translation, machine reading comprehension (MRC), machine writing, etc.

To handle the problems of machine's reading and writing, the first task of the representation of text in the computer needs to be solved. Although in some languages, such as Chinese, word segmentation is needed before the word representation. In traditional statistical natural language processing tasks, text representation is mostly based on the discrete feature vector method, which relies heavily on handcrafted feature engineering (e.g., vector space model (VSM)).[131] Feature engineering is always time consuming and incomplete, and the problem of dimensional explosion also exist in it. With the rise and development of deep learning methods and computing hardware, deep learning methods have been employed and produced state-of-the-art results in many domains, ranging from computer version to speech processing. References 132–134 put forward a neural networks-based method to embed words into the low-dimensional distributional vectors known as Word Embeddings. Word Embeddings is also a statistical method, which follows the distributional hypothesis that words occurring in a similar context tend to have similar meanings. Thus, we can think that Word Embeddings contain syntactical and semantic information, and its major advantage is that they can capture the similarity between words by measuring the similarity between vectors. Distributed representations have been the basis of deep learning-based NLP tasks and have helped achieve encouraging results in a wide range of NLP tasks.[135–137]

POS tagging is the process of marking up a word in a text with a particular part of speech based on both its definition and its context. The difficulty of this problem is that the same word will show different parts of speech in different contexts. Rule-based methods and statistics-based methods are the main approaches in traditional POS tagging and most machine learning methods have achieved accuracy above 95%, whereas recent research works focused on deep learning based-method have been achieving even better accuracy. Reference 137 proposed a deep neural network that learns the character-level representation of words and associates them with

Table 2. Methods for subtasks of reading and writing.

Subtasks	Traditional methods	Deep learning based methods
Word representation	Feature Engineering & VSM[131]	Word Embeddings & neural network-based language model[132–134]
POS Tagging	Rule-based methods & Statistics-based methods	Convolutional neural networks[137] Adversarial neural networks[138] Transfer learning[139] CRF, LSTM, Bi-LSTM, LSTM[140] RNN, GRU, LSTM, and Bi-LSTM[141]
NER	CRF	CRF & CNN,[142] CRF, & RNN[143–145] LSTM, CRF, & Attention[146,147] Transfer learning[148] Semi-supervised method[149] Active learning[150]
Text Classification	Keywords matching-based method Rules-based knowledge engineering Statistical machine learning-based methods (e.g., SVM, KNN)	FastText[152] CNN for text sequence[153] Character-level CNN[154] Hierarchical model[155] Transfer learning method[158]
Text Sentiment Analysis		Autoencoders[161] End-to-end models[172]
Machine Translation	Rules-based methods Statistical machine translation	RNN Encoder–Decoder[174] (Bi)RNN & Attention[176,177] Sequence-to-sequence architecture based on CNN,[178] Attention[179] GANs,[182] NMT without RNN[183]
MRC	Pipelines-based methods	Choose answer from candidates[186,192] No candidates to choose[194] R-NET by MSRA[195] Bi-Directional Attention Flow[196] Deep residual coattention encoder[197] Dynamic-critical reinforcement learning & reattention mechanism[198] Attention-over-attention reader[199] Transfer learning[200] Dynamic Fusion Network[201]
Machine writing	Extraction methods[205–210] Sentences compression[211,212] Sentences fusion[213]	Deliberation Networks (RNN)[215] CNN,[216] GAN[217] Reinforcement Learning[218]

usual word representations to perform POS tagging. Reference 138 presented the method of employing adversarial neural networks to deal with the POS tagging problem for Twitter's text. Transfer learning was introduced to induce automatically a POS tagger for languages that have no labeled training corpus.[139] In Ref. 140, a few

models were utilized to address the Uyghur POS tagger, including CRF, LSTM, Bi-LSTM, LSTM with a CRF layer, and Bi-LSTM networks with a CRF layer. Reference 141 evaluated several sequential deep learning methods, including RNN, GRU, LSTM, and Bi-LSTM for Malayalam tweets, and different experimental parameters.

NER, which is one of the most important bases of NLP tasks, refers to the task of recognizing the entity with a specific meaning in the text, mainly includes name, place's name, institution's name, and proper noun. It also includes two subtasks: entity boundary recognition and entity type determination. CRF is a traditional discriminant probability model recognized as a good algorithm for solving NER problems. Many fusion methods of CRF and neural networks have emerged in this field. Reference 142 is one of the representative works that neural networks were used for NER, in which CRF was fused into CNN. On the basis of similar ideas, Refs. 143–145 have combined RNN and CRF to deal with NER. These papers had proposed novel architectures for combining word embedding with character-level representation, in which attention mechanism was introduced to dynamically extract information from both word- and character-level components.[146,147] Generally speaking, deep learning relies on a large number of annotated samples as training data. In order to solve the limitation caused by the massive annotated data, many literatures have studied the NER methods based on a small amount of annotated data, such as transfer learning,[148] semi-supervised method,[149] and active learning.[150]

Text classification is the technology that automatically marks the text with labels according to a certain or standard classification system. The research of text classification has gone through several stages including keywords matching-based method, rules-based knowledge engineering, statistical machine learning-based methods (e.g., SVM, KNN), and deep learning-based methods. Recently, Ref. 151 is an excellent work that provided an overview of the state-of-the-art elements of text classification. Reference 152 explored a simple but efficient baseline for text classification, fastText, which provides the idea that some tasks can be solved by some extremely simple models. Reference 153 researched convolutional neural networks to deal with text sequence and carried out experiments for sentence-level classification, further achieving compelling results. Following this route, character-level convolutional neural networks are studied for text classification.[154] Reference 155 divided the text into three levels: word, sentence, and document. They constructed a hierarchical model for long text classification by using the hierarchical attention mechanism. Reference 156 proposed deep average networks (DAN) and attentional DAN to actualize the conversational topic classification for the evaluation of the conversational bots. Lai *et al.* introduced recurrent convolutional neural networks for this task and applied a recurrent structure to capture the contextual information, whereas a convolutional neural network was used to construct the representation of text.[157] Much of the success that transfer learning has achieved in computer vision cannot yet be fully transplanted into NLP. Text categorization still requires task-specific modifications and training from scratch. Howard *et al.* proposed an effective

transfer learning method for text classification, known as universal language model fine tuning, and introduced some key techniques for model fine tuning.[158]

Also known as opinion mining and inclination analysis, text sentiment analysis is the process of analyzing the emotions present in the text. Reference 159 gave a macroscopic introduction to the field of sentiment analysis, such as research objects and venues. Reference 160 summarized several major models in the field of deep learning and comprehensively introduced their applications in the task of sentiment analysis. Additionally, they reviewed three levels of granularity research works for sentiment analysis and their subtasks. Reference 161 researched the usage of autoencoders in modeling textual data and sentiment analysis, and tried to address the problems of scalability with the high dimensionality of vocabulary size and task-irrelevant words by introducing a loss function of autoencoders. Also, there are many other leading research works that focused on text sentiment analysis and its applications.[162–170] Although sentiment analysis is treated as a classification problem, sentiment analysis is actually a suitcase research problem that requires dealing with many NLP tasks.[171] Reference 172 proposed a novel tagging scheme to jointly extract entities and relations, which can be seen as the subtasks of sentiment analysis, by using several end-to-end models.

Machine translation is a cross-language literacy that automatically translates the source language into the target language. Machine translation consists of experienced rule-based methods and statistical machine translation. In recent years, research works had mainly focused on the neural machine translation (NMT). Reference 173 summarized a successful usage of neural networks in the machine translation system. Cho *et al.* proposed a novel neural network model called RNN encoder–decoder for statistical machine translation and show that the proposed model had the capacity of learning semantic and syntactic meaningful representation of linguistic phrases.[174] This research also involved an empirical evaluation of a novel hidden gated unit. Reference 175 presented a general end-to-end approach to sequence learning for machine translation and suggested that the NMT can achieve results similar to the traditional techniques. Reference 176 proposed the attention mechanism, which achieved state-of-art results for statistical machine translation. Following this research, Ref. 177 explored attention-based NMT architectures, including a global approach and a local one, to improve the NMT performance and achieved remarkable results. Reference 178 introduced a sequence-to-sequence architecture, which was always deployed via RNN and based entirely on CNN, and achieved better accuracy and time efficiency. Different from the previous encoder–decoder architecture, Ref. 179 proposed a neural network architecture that only used attention mechanism, and the experimental results on the machine translation task have showed that the architecture performed well both on quality and training speed. Google team presented Google's NMT system to address some relevant problems such as robustness, accuracy, and speed.[180] Thereafter, their team tried to solve the problem of multilingual translation by using a single NMT model.[181] GANs were also applied to NMT, and Ref. 182 introduced a conditional sequence, GAN, in

which the generator aimed to translate the sentences while the discriminator tried to discriminate the outputs generated by the generator from the sentences translated by a human being. Reference 183 proposed a novel model to produce translation outputs in parallel instead of one after another, so as to reduce the latency occurring during inference.

MRC, also researched as the open domain QA, is the ability of intelligent robots to comprehend a given context and give answers of questions related to the given context. Information retrieval can also be considered as a MRC issue.[184] Many MRC datasets were exposed to train and evaluate MRC, such as Machine Comprehension Test, Children's Book Test, CNN/Daily Mail, The Stanford Question Answering Dataset (SQuAD), and DuReader.[185–191] Traditional research works on MRC always focus on pipeline-based methods consisting of several NLP subtasks. With the popularity of neural network model in NLP tasks, there are a series of works that focus on end-to-end neural networks for the MRC task, in which the answers are chosen from the candidates.[186,192] While a novel end-to-end neural architecture, using match-LSMT[193] and answer pointer, was proposed based on SQuAD, it had no answers of candidates and was thought to be difficult to be dealt with.[194] Thereafter, R-NET introduced by MSRA solved the question via four steps following match-LSTM.[195] Reference 196 presented bi-directional attention flow (BiDAF) network, in which the context is represented at different granularities and BiDAF was used to locate the key context. Reference 197 improved the performance of MRC from the point of objective function and network model, in which a mixed objective combined cross-entropy loss with self-critical policy learning. This research also proposed a DCN that was improved by a deep residual coattention encoder. Reference 198 summarized the advantages and disadvantages of match-LSTM, R-NET, and other previous models, and made significant improvements by the reattention mechanism and dynamic critical reinforcement learning. Another excellent research was Ref. 199, which proposed a novel model known as attention-over-attention reader to address the cloze-style MRC and achieved state-of-art performance in many public datasets. Transfer learning was also introduced into MRC, and a two-stage synthesis network was presented by Ref. 200 to answer the questions in one domain that were provided in a model from another domain. Reference 201 proposed a novel dynamic fusion network model for MRC, in which the attention strategy was chosen flexibly according to the question types. A novel architecture called QANet consisted of local convolution, and global self-attention was proposed to improve the speed of training and reasoning.[8] The experiment results have showed that the proposed model achieved a greater increase in speed along with equivalent accuracy with recurrent models. Reference 202 extended the paragraph-level MRC to the documents level where the documents are given as context and a novel objective function was introduced to produce a global answer. Reference 203 proposed a meaningful assumption that if the MRC models could combine textual evidence from multiple contexts, then the scope of this model would be extended. Based on this novel task, the literature produced datasets and validated some methods.

By machine writing, we mean generating text, not writing calligraphy by robots. In essence, all of the above-mentioned methods such as the dialogue system, QA system, and machine translation belong to the category of machine writing, and the difference lies in the different application premises and application scenarios. Machine writing will be one of the most important ways for intelligent robots to express themselves and also one of the most important means of natural HCI. There are many other forms of machine writing tasks such as text summarization, news writing, image description, etc. Reference 204 first proposed the technology of text summarization, which refers to analysis background documents, summarization of the main points of documents, along with extraction or generation of short summaries relative to the original documents. Traditional machine learning-based text summarization mainly adopted the extraction method, in which the summary was accomplished in two steps, which were sorting sentences by importance[205–207] and sentences' arrangement.[208–210] To make the generated summaries more compact, sentences compression and sentences fusion are commonly employed in the text summarization system. Sentences compression can be seen as a sentence-level text summarization, through which a long sentence is summarized to a short one, and several ways were employed in this direction.[211,212] Sentence fusion technology combines sentences and overlapping content to get a single one so as to reduce repeatability in the generated summary.[213] For the generative text summary methods, the sentences in the abstract are not extracted and rewritten based on the original text, but are generated based on semantic information.[9,214] Similar research works also include methods based on various deep neural network models, such as RNN, CNN, GAN, etc., the hidden layers among which can be regarded as abstract semantic information.[215–218] The researchers also conducted some interesting applications based on the technique, such as academic summaries[219,220] and student course feedback summarization.[221]

Another area of research in machine writing is automatic text generation based on data, which have been widely used in many fields, such as weather report, news report generation, and biography domains.[222–224] In recent years, many Chinese scholars have used the text generation technology to create Chinese poems of specific subjects or emotion and achieved prominent results.[225–227] Another data-based machine writing field is image caption generation, whose task is to generate texts describing the content of the given image. Apparently, this task has led to a series works of joint modeling of image semantic annotation and NLG.[228–230] Automatic music generation is also an interesting research avenue, which is related to artistic creation. A great amount of novel deep learning methods was proposed to address this challenge.[231–233]

4.3. *Visual sense*

Vision is the most important sense in human beings, and more than 80% of the information received from the outside world is obtained through vision. Machine

vision, or computer vision, is a science that studies how to make a machine "see" like humans. This implies to use the camera to replace the human eye to obtain images and use the computer to replace the human brain to process images, so that the machines can be made for gaining a high-level understanding of images to simulation functions that the human visual system possesses. In the process of interpersonal communication, human beings recognize and judge the object's identity, expression, physical behavior, etc., through vision, and consider this as the basis of interaction. In the following sections, we will briefly summarize these contents (see Table 3 for an outline).

Identification is a technique used in computer vision to determine one's identity and characteristics. The most common identification methods are biometrics-based methods, such as face recognition, iris recognition, and fingerprint recognition. The research of biometric recognition has a long history, and its development can be divided into four stages.[234] Although the biometric identification based on the traditional methods has achieved satisfactory results, with the rise of deep learning, the

Table 3. A brief summary of visual senses for HCI.

Tasks	Subtasks	Representative works and review works
Identification	Face detection	Faster R-CNN[235]
		Faceness-Net[236]
		Face detection for low-quality images[237]
	Face alignment	Face Alignment,[238]
		Facial feature point detection[239]
	Face recognition	Shallow representations-based methods[240,241]
		Deep learning-based methods[242]
	Iris recognition	Machine learning-based methods[243]
		Long-range iris recognition[244]
	Fingerprint recognition	Fingerprint recognition for young children[245]
		Fingerprint recognition at crime scenes[246]
		Review works[247,248]
	Others	Age and gender recognition[249,250]
Facial expression recognition	—	CNN based methods[252,253]
		Multi-modality feature fusion-based method[254]
		Expression recognition based on static images[255]
		Micro-Expression Recognition[256–258]
Facial expressions generation	—	Interactive GAN-based method[260]
		3D facial expression generation[261]
		Humanoid robot expression generation[23]
		Three-dimensional speaking characters[262]
		Expression generation natural description[264,265]
Posture or gestures recognition	—	Driving posture recognition[266]
		Weighted fusion method for gesture recognition[267]
		Posture recognition for hazard prevention[268]
		Emotional body gesture recognition[269]
		Gesture recognition in video[271]
		Hand gesture recognition[272]

deep neural network methods are introduced into the field to seek a better recognition performance.

The first step of face recognition is the detection of face with an aim to determine whether faces exist on a given image or not. If these faces exist, the location and size of faces are also determined. A number of studies have focused on this area.[235,236] Reference 237 provided a review of face detection for low-quality images. Face alignment is the process of marking out the important organs, such as eyes, nose, and mouth, in the image with feature points, and Refs. 238 and 239 had reviewed the research progresses in this field systematically. There is a lot of research works in the field of face recognition. References 240 and 241 summarized the previous research works based on shallow representations, whereas Ref. 242 focused on the literatures of deep learning-based face recognition. In addition to the face recognition function like humans, the intelligent robots also have the abilities of identification that human beings do not have. Because of the stability of biometrics such as iris and fingerprints, these biometric features are often used for identification. Reference 243 surveyed the iris recognition literatures based on machine learning methods, whereas Ref. 244 focused on long-range iris recognition research works that can extend the application range of this technology. Jain *et al.* researched the fingerprint recognition question of young children, which did not get enough attention as much as the research of Ref. 245. An automated latent fingerprint recognition algorithm was proposed for the comparison of latents found at the crime scenes.[246] References 247 and 248 are the latest reviews conducted in this field. Besides, there are other works that are related to identification, which focused on age and gender recognition.[249,250]

Facial expression recognition refers to the recognition of the states of expression contained in the image from a given static image or dynamic video sequence, so as to determine the psychological emotions of the identified object.[251] Reference 252 proposed a neural network-based expression recognition method to improve the generalizability of model, which consisted of two convolutional layers with each followed by max pooling and, then, four inception layers. Reference 253 proposed another CNN-based expression recognition scheme, which was combined together with specific image pre-processing steps to address the questions of limited training samples and the uncertainty of sampling during training. A multi-modality feature fusion-based framework was proposed for face recognition in videos to improve the system's robustness.[254] While expression recognition based on static images was also researched by the authors, Ref. 255 proposed a novel method to train an expression recognition network based on the static images.[255] Micro-expression recognition, which is regarded as a harder problem, was also researched by a large amount of research works.[256–258]

Corresponding to facial recognition, this study provides the automatic generation of facial expressions. Its content generated various emotional expressions of a given facial image or a specific text. This research is considered important as it can be seen as a feedback in the HMI. In Ref. 259, a chaotic feature which extracted associative memory was proposed to stimulate the human brain in generating the facial

expressions. Reference 260 proposed an interactive GAN-based method for generating facial behaviors in a dyadic interaction scene. A novel point clouds-based method was introduced for 3D facial expression generation in Ref. 261. Additionally, there are research works in this field combined with robotics and bionics to generate or imitate expression on robots or virtual faces. References 23 and 262 researched the automatic facial expression learning methods for a humanoid robot to generate vivid expressions and increase the interactivity of the humanoid robot. Reference 263 developed a free software and API that can generate dynamic facial expressions for the three-dimensional speaking characters. Reference 264 investigated a novel problem of generating images from some natural description and proposed a CAVE-based method to address this problem. A similar research work was conducted in Ref. 265.

The detection and recognition of posture, gestures, and eye movements are of great significance in the interactive process, and there are a lot of research works that need to be performed in this area. A CNN-based method for driving posture recognition was introduced in Ref. 266 to detect the driver's fatigue and inattention. Reference 267 researched the problem of gesture recognition with a weighted fusion method of D-S evidence theory by fusing Kinect and surface Electromyogram (EMG) signals. An ergonomic posture recognition technique was discussed in Ref. 268, which aimed to prevent construction hazard by a using an ordinary 2D camera. Reference 269 defined a framework for automatic emotional body gesture recognition and reviewed the related research results in this field. Besides, multi-modal approaches for improved emotion recognition were also discussed in both this work and Ref. 270. An end-to-end architecture incorporating temporal convolutions and bidirectional recurrence was proposed in Ref. 271 for gesture recognition in videos. A novel approach and a real-time system for static hand gesture recognition were introduced in Ref. 272, which could vastly improve the accuracy and speed of recognition. Research works on vision-based gesture recognition were reviewed by Ref. 273, which also included the discussion of the technical aspects of the whole pipeline and the challenges in this field. It is very useful to recognize and track eye movements during HCI, and it can be used to detect the direction of human attention. Reference 274 introduced an approach integrating eye movement recognition, and tracking and application scenarios were designed to evaluate the proposed method. A robust online saccade recognition algorithm was proposed, which involved the integration of electrooculography (EOG) and video signals. The experiments results proved that the multimodal fusion technology was helpful in improving the accuracy of eye movement recognition.[275]

Optical character recognition (OCR) is the process of converting typewritten or handwritten characters present in an image into the format that the computer can identify and edit, which is one of the most important ways of interaction. Reference 276 surveyed the OCR systems based on soft computing methods for different languages, such as English, French, German, Latin, whereas the methods of feature extraction of OCR was summarized in Ref. 277. The method for improving the OCR

performance of low-quality images was studied in Ref. 278. Reference 279 proposed a CNN-based method to learn the features of Chinese characters. Then, it addressed the problem of Chinese characters in completely automated public Turing test to tell computers and humans apart, which is increasingly used in many web applications for security reasons. Another aspect of OCR is to train the computer to automatically write characters or generate images with the character, which is also challenging and interesting. Reference 280 proposed a RNN-based framework to train a discriminative model and a generative model for recognition of Chinese characters and generation of Chinese characters, respectively. Reference 281 proposed a novel RNN-based model in order to overcome the challenge of handwritten character generation.

4.4. *Other senses*

Reference 282 discussed the influences of physiological signals on cognition, and there are a large number of signal sources that could be detected and processed by some special equipment and, later, used for interaction, for example representation and detection of states of human emotions. From the aspect of original source of emotions, affective computing can be divided into two categories: external nonphysical performance-based affective computing (e.g., facial expressions, text, body gestures, and speech) and inherent physiological information-based affective computing, such as electroencephalography (EEG), electrocardiogram (ECG), and EMG.

Reference 283 proposed the EEG-based method for the recognition of human intentions, which can be used for brain–computer interface, by employing both cascade and parallel convolutional recurrent neural network models. Reference 284 explored the feasibility of wireless EEG signals to assess the memory workload levels in special tasks, and the experimental results indicated that the proposed project can be used for mental workload identification when humans are engaged in cognitive activities. EEG signals are also applied to emotion detection tasks.[285,286] Both EEG signals and facial expressions were used for continuous emotion detection in Ref. 287, and the relationship between them was analyzed. More literatures based on physiological signal emotion recognition are presented in Ref. 288, which is a newly published review in this field. Reference 289 summarized the application of deep learning and reinforcement learning to several different biological datasets and discussed the future development perspectives. In Ref. 290, sleep apnea features were extracted from capacitively coupled ECG signals to monitor sleep apnea. Reference 291 researched ECG used for healthcare monitoring by employing residential wireless sensor networks.

The EMG signal is also widely used in the man–machine control system. An upper limb rehabilitation training system combined with portable accelerometers and EMG was designed and developed for children with cerebral palsy to capture their functional movements and address the problems of in-home training.[292] In Ref. 293, an EMG- and AdaBoost-based movements recognition method was introduced into a robotic hand–eye system for grasping and manipulation of control strategy.

In Ref. 294, high-density surface EMG signals were decomposed from the forearm muscles in the non-isometric wrist motor tasks of normally limbed and limb-deficient individuals, which could be used for prosthesis control with the help of the decoded neural information. Reference 295 proposed an optimal control framework based on EMG for the design of physical human–robot interaction in the application of rehabilitation. In Ref. 296, natural EMG signals were collected in a natural manner by introducing a physical haptic feedback mechanism, and an interface was designed for human adaptive impedance, extracted from the transfer of EMG signals. An algorithmic framework is proposed in Ref. 297 for EMG-based gesture recognition, and a prototype system along with an application program was developed to realize the gesture-based real-time interaction.

An EOG-based eye-movement tracking system was proposed for HCI in Ref. 298. Reference 299 developed a real-time eye-writing recognition system based on EOG, and users can write predefined 29 symbolic patterns (26 lower case alphabet characters and 3 functional input patterns representing space, backspace, and enter keys) with their volitional eye movements. Blood volume pulse (BVP) signal is a weak physiological signal formed by the periodic contraction and expansion of the heart, which leads to the periodic changes in the blood volume of the face. Therefore, the BVP signal is often used to detect the heart rate and breathing rate.[300,301] Galvanic skin response (GSR) was used in Refs. 302 and 303 to design GSR-based sensors for the detection of stress states and prediction of performance under stressful conditions. GSR applied to sentiment classification was also studied.[304] Tactile ability is essential for intelligent robots to interact with humans in a HCI environment. Electronic devices having tactile ability were designed in Refs. 305 and 306 to address this challenge. Tactile sensors were also used for object recognition in Refs. 307 and 308. Additionally, methods and technologies for the implementation of large-scale robot tactile sensors were researched in Ref. 309. In addition, WiFi can also be deployed in the HCI system for the implementation of the functions such as motion detection, activity recognition, and sleep monitoring.[310–312]

5. Intelligent Robots

Intelligent robots are an updated version of the traditional robots in both software and hardware systems. By upgrading the software, the intelligent robots have higher levels of brains, which bestow them with a comprehensive improvement in perception, reasoning, and decision-making. With the hardware upgrade, the intelligent robots have more perfect bodies so that they can better imitate human behaviors on the basis of completing delicate works and toilsome works. In combination with the both improvements, the intelligent robots can execute human commands or think independently to complete certain tasks, learn, and improve them autonomously. They can also interact with human beings in a friendly manner.

Motion elements are the centralized embodiment of robot positioning, obstacle recognition, navigation, and other functions in an unstructured environment,

thereby reflecting the autonomous ability of an intelligent robot to adapt to the complex environment. Reference 313 developed a simple and highly mobile hexapod robot RHex, who can traverse solid, broken, and obstructed ground without any topographic induction or active control. Boston Dynamics has developed two four-legged robots, rough terrain robots,[314] and small four-legged robots[315] that mimic the mobility, autonomy, and speed of living creatures. The robots can move flexibly in various terrains such as steep, rutted, rocky, wet, muddy, and snowy outdoor terrains. ATLAS is a two-legged humanoid robot developed by Boston dynamics,[316] which can realize the dynamic planning, control, and state estimation of the two-legged robot. The robot can operate reliably in complex environments and can regain its balance even after slipping on snow, or it can get up if it is pushed down deliberately.

Another important element of an intelligent robot is the control element, which can perceive human's control intention in various ways and execute relevant actions according to commands. It is often used to assist the control of prostheses for patients with paralysis. Spinal cord injuries, stroke in the brain stem, and other diseases make it impossible for patients with paralysis to control their limbs autonomously. The prosthesis with controllable capability can detect and execute the patient's intention via signal sources, such as neural interface and physiological signals, so as to realize the patient's control of the prosthesis. Reference 317 exhibited the abilities of people with chronic tetraplegia to perform three-dimensional stretching and grasping motions by using a robotic arm controlled through the neural interface. This literature also showed that it is possible for tetraplegic patients to reconstruct the useful multi-dimensional neural controls from complex devices directly even after years of central nervous system injuries. References 318 and 319 researched on the controlling of robotic arm by modeling the multi-channel EEG signals and motion state together. Using pneumatic artificial muscles and inflatable sleeves, Ref. 320 developed a robotic arm with seven degrees of freedom (DOFs), which were combined with elements and positive qualities of rigid and soft robotics. Brain–computer interfaces (BCIs) were employed in Ref. 321 to stimulate the muscle and control of robotic arm for reaching and grasping movements in people with tetraplegia.

The above-mentioned robots generally have solid bodies, complex structures, and limited DOFs, whereas the soft robots can achieve continuous deformation and, therefore, have infinite DOFs. References 322 and 323 conducted research on soft robots. The development of 3D printing technology and materials science have greatly benefitted researcher works on soft robots, owing to which they have shown a significant progress and achieved the tasks of grabbing, human–robot collaboration, etc.

The interactive elements of intelligent robots are studied and practiced by a large number of researchers, and Sec. 4 introduces a great amount of research works and technologies focused on these interactions. In fact, scientists have developed several intelligent platforms and robots with the rudimentary ability of natural HCI. For example, MIT affective computing research team launched the Tega and Jibo platforms successively in 2016, which have certain emotional computing and

perception abilities. In 2014, Microsoft launched the interactive platform Xiaobing, which can understand the emotional context to a certain extent. In 2015, the Turing robot team released an AI robot operating system with multimodal interaction mode, that is Turing OS. Turing OS simulates human-to-human interaction, giving the robot a wealth of input and output modes, including text, voice, action, environment, etc. IBM teamed up with Japan's SoftBank in 2016 to develop Pepper, an "emotional" robot that responds to the parts of the spoken language in limited settings. ABC Robot is a leading multi-modal human-computer interaction platform of Baidu. The platform can realize multimodal HCI such as speech recognition, semantic understanding, face recognition, gesture recognition, and multi-sensor fusion.

The Ren team from Hefei University of Technology in China studied the emotion computing system on the platform of a humanoid robot; constructed a heart state transfer network, which combined universality and individuality, for mental health problems; developed a multi-modal emotional response model based on the established heart state transfer network; and established an evaluation system for coping strategies. The emotional robot platform and its cloud system developed by the team mainly have the functions of character identity and emotion recognition, gesture and voice interaction, intelligent emotional conversation and chat, and emotional interaction. Emotional robots can be used at home and in medical settings for people of different ages (especially for the elderly) and the assisted rehabilitation of specific conditions (autism and depression).

The above content roughly belongs to the intelligent system of intelligent robots. In fact, more research works are focused on the hardware system of robots, such as actuator, driving device, sensing device, control system, etc. However, these studied contents are not within the scope of this review. For more literatures about intelligent robot systems, see Ref. 1. The authors of that research had reviewed the current research works on intelligent robot systems and prospected the future development trend in this field.

6. Challenges for HCI and Intelligence Robots

HCI and intelligent robot technologies have broad development prospects in various industries. However, although there are many achievements in these two fields, but there is still a large space needed for the expansion of the intelligence level grow. Future intelligent robots and their interaction technologies need to be developed in the following aspects.

6.1. *Technologies of multimodal fusion perception and human-like intelligent perception*

Human beings express their emotions and intentions through multiple signals, such as language, pronunciation, and intonation, facial expressions and gestures, as well as

some physiological signals, such as blood pressure and heartbeat. Most of the existing perception methods are focused on the single mode, whereas the correlation between the multiple modes is ignored. Therefore, multimodal databases, multimodal data hierarchical fusion perception, and human-like intelligent perception technologies based on this database will become an important direction for research.

Existing mainstream approach for multimodal fusion perception is dependent on large-scale neural network and big data. In addition, we also can be provided good references by the group decision making and multiple criteria decision making in management to study the decision-making process in the process of multimodal fusion perception.[324–328]

6.2. *Mechanism of multimodal cooperative analysis and intelligent reasoning*

At present, research works mainly start from the external appearance of HCI and adopt the traditional engineering methods. Then, they focus on the research and implementation of perception theory and technology. However, at the real thinking level, cooperative analysis and intelligent reasoning mechanism of multi-source data have not yet been formed yet. The cooperative representation of multimodal heterogeneous emotional data, the deep adaptive cooperative semantic understanding mechanism, and the efficient reasoning mechanism integrating ontology knowledge and containing knowledge should be the major topics of research in this field.

6.3. *Technologies of emotion creation and natural HCI*

Emotion is a very important factor in the process of natural HCI, further acting as the key for its establishment. Emotion is still a stumbling block on the path of natural HCI and will restrict the further development of intelligent robots. The existing interaction platform does not integrate multi-channel information and their corresponding feedback mechanism, and cannot achieve an emotional interaction. The methods mapping human emotions to the emotions of machines and the dynamic feedback mechanism of the emotional loop are the possible ways to realize the creation of emotions and natural HCI.

6.4. *Mental health perception and calculation based on an emotional interaction*

Psychology studies have demonstrated that emotional state is the important indicator of mental health, and behaviors such as language, voice, facial expressions, and gestures in the process of interaction always convey emotions. These interaction behaviors have become important ways for people to express their feelings and a visible indicator of the states of psychological health. Sometimes, these behaviors even resemble a variety of psychological crisis. A sudden low voice, for example, may simply be a sign of poor physical health (such as a cold), but not a sign of poor mental

health. However, if the voice is low and the expression is painful at the same time, and the content of negative emotions is included in the ordinary voice text, the mental health of the patient can be judged to be in a bad state, which needs to be adjusted and dealt with. It is a great challenge to accurately perceive and calculate people's mental health state through multi-source signals. Additionally, ways to guide and improve people's psychological state in the process of interaction also serve as another great challenge.

6.5. *Deep understanding of natural language and personalized interaction*

Deep understanding of natural language and personalized interaction are also difficult challenges faced by HCI. First, the combination of scene, historical interactive information, pragmatics, and, even, emotions and then, interaction with a deep understanding of semantics could be natural and efficient. In the personalized interaction, the intelligent robot can adjust the interaction method and strategize neatly according to the scene, interaction object, interaction state, etc.

6.6. *Human–machine integration and intelligent human–machine interface technology*

The intelligence degree of an intelligent robot is growing higher and higher, and the human beings are becoming more and more dependent on the intelligent robot technology. The ways to promote the integration of human and robot will become an important research avenue in this field. In order to better adapt to the application of different users and different tasks, improving the harmony of human–robot interaction and intelligent man–machine interface will become an effective way of human–machine integration.

7. Conclusions

Ever since computers were born, there have been various interactions between people and computers in order to make computers more responsive to the humans' needs. The continuous improvement in human demand and curiosity drives the development of HCI technology and intelligent robot technologies. A large amount of research works has been carried out to make HCI more natural and harmonious and, at the same time, make robots more intelligent and adaptable. With the rapid development of AI technology in recent years, it provides unprecedented development opportunities for the research of these two technologies. This paper summarizes the development status of HCI from the aspect of interaction abilities and introduced the related technologies of an intelligent robot. Thereafter, the challenges for these two fields in the future development and possible research approaches are expounded.

Acknowledgments

This research has been partially supported by National Natural Science Foundation of China under Grant Nos. 61432004 and U1613217.

References

1. T. M. Wang, Y. Tao and H. Liu, Current researches and future development trend of intelligent robot: A review, *International Journal of Automation and Computing* **15**(5) (2018) 525–546.
2. A. H. Abdelaziz, Comparing fusion models for DNN-based audiovisual continuous speech recognition, *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* **26**(3) (2018) 475–484.
3. S. Ren, K. He, R. Girshick, X. Zhang and J. Sun, Object detection networks on convolutional feature maps, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(7) (2017) 1476–1481.
4. T. Young, D. Hazarika, S. Poria and E. Cambria, Recent trends in deep learning based natural language processing, *IEEE Computational Intelligence Magazine*, **13**(3) (2018) 55–75.
5. K. Tirri and P. Nokelainen, *Measuring Multiple Intelligences and Moral Sensitivities in Education* (SensePublishers, 2011).
6. J. Kim, M. El-Khamy and J. Lee, Residual LSTM: Design of a deep recurrent architecture for distant speech recognition (2017), arXiv preprint arXiv:1701.03360.
7. Y. N. Chen, W. Y. Wang, A. Gershman and A. Rudnicky, Matrix factorization with knowledge graph propagation for unsupervised spoken language understanding, *ACL-IJCNLP* **1** (2015) 483–494.
8. A. W. Yu, D. Dohan, M. T. Luong *et al.*, QANet: Combining local convolution with global self-attention for reading comprehension (2018), arXiv preprint arXiv:1804.09541.
9. F. Liu, J. Flanigan, S. Thomson, N. Sadeh and N. A. Smith, Toward abstractive summarization using semantic representations (2018), arXiv preprint arXiv:1805.10399.
10. J. H. Carlisle, Evaluating the impact of office automation on top management communication, in *Proc. June 7–10, 1976, National Computer Conference and Exposition* (1976), pp. 611–616.
11. S. K. Card, T. P. Moran and A. Newell, The keystroke-level model for user performance time with interactive systems, *Communications of the ACM* **23**(7) (1980) 396–410.
12. T. T. Hewett, R. Baecker *et al.*, ACM SIGCHI curricula for human-computer interaction, *ACM* (1992), doi: 10.1145/2594128.
13. G. D'Amico, A. D. Bimbo, F. Dini *et al.*, Natural human-computer interaction, in *Multimedia Interaction and Intelligent User Interfaces* (London, Springer, 2010), pp. 85–106.
14. H. A. Wang and F. Tian, Natural and efficient human computer interaction, in *10000 Selected Problems in Sciences* (Science Press, Beijing, 2011), pp. 625–627.
15. X. Cao, What is the natural characteristics of natural user interface, *Commun China Computer Federation* **11** (2011) 14–18.
16. S. Legg and M. Hutter, A collection of definitions of intelligence, *Frontiers in Artificial Intelligence and Applications* **157** (2007) 17–24.
17. L. S. Gottfredson, Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography, *Intelligence* **24**(1) (1997) 13–23.
18. U. Neisser, G. Boodoo, T. J. Bouchard *et al.*, Intelligence: Knowns and unknowns, *American Psychologist* **51**(2) (1996) 77–101.

19. G. Matthews, M. Zeidner and R. D. Roberts, *The Science of Emotional Intelligence: Knowns and Unknowns* (Oxford University Press, New York, NY, 2007).
20. H. Gardner, Frames of mind: The theory of multiple intelligences, *Quarterly Review of Biology* **4**(3) (1985) 19–35.
21. S. Legg and M. Hutter, Universal intelligence: A definition of machine intelligence, in *Minds and Machines*, Vol. 17(4) (Kluwer Academic Publishers, 2007), pp. 391–444.
22. H. Zhou, M. Huang, T. Zhang, X. Zhu and B. Liu, Emotional chatting machine: Emotional conversation generation with internal and external memory (2017), arXiv preprint arXiv:1704.01074.
23. F. Ren and Z. Huang, Automatic facial expression learning method based on humanoid robot XIN-REN, *IEEE Transactions on Human-Machine Systems* **46**(6) (2016) 810–821.
24. M. L. Minsky, *The Society of Mind* (Simon & Schuster Press, 1988).
25. R. W. Picard, *Affective Computing* (MIT Press, 1997).
26. M. Nagamachi, *Kansei/Affective Engineering* (CRC Press, 2011).
27. X. Tu, Artificial emotion, *The Paper Assembly of the 10th Annual CAAI* (Guangzhou, China, 2000).
28. Z. Wang and L. Xie, Artificial psychology-an attainable scientific research on the human brain, *IPMM*, Vol. 2 (Honolulu, 1999), pp. 1067–1072.
29. Z. Wang, Artificial psychology and artificial emotion, *CAAI Transactions on Intelligent Systems* **1**(1) (2006) 38–43.
30. J. E. Ledoux, Emotion circuits in the brain, *Annual Review of Neuroscience* **23**(23) (1999) 155–184.
31. R. N. Cardinal, J. A. Parkinson, J. Hall *et al.*, Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex, *Neuroscience & Biobehavioral Reviews* **26**(3) (2002) 321–352.
32. M. S. Hossain and G. Muhammad, Audio-visual emotion recognition using multi-directional regression and Ridgelet transform, *Journal on Multimodal User Interfaces* **10**(4) (2016) 325–333.
33. P. Shaver, J. Schwartz, D. Kirson *et al.*, Emotion knowledge: further exploration of a prototype approach, *Journal of Personality & Social Psychology* **52**(6) (1987) 1061.
34. M. A. Nicolaou, S. Zafeiriou and M. Pantic, Correlated-spaces regression for learning continuous emotion dimensions, *ACM International Conference on Multimedia* (2013), pp. 773–776.
35. L. F. Barrett, Discrete emotions or dimensions? The role of valence focus and arousal focus, *Cognition & Emotion* **12**(4) (1998) 579–599.
36. S. K. A. Kamarol, M. H. Jaward, H. Kälviäinen *et al.*, Joint facial expression recognition and intensity estimation based on weighted votes of image sequences, *Pattern Recognition Letters* **92**(C) (2017) 25–32.
37. A. Mehrabian, Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament, *Current Psychology* **14**(4) (1996) 261–292.
38. C. Breazeal, Function meets style: Insights from emotion theory applied to HRI, *IEEE Transactions on Systems Man & Cybernetics Part C* **34**(2) (2004) 187–194.
39. F. Ren, C. Quan and K. Matsumoto, Enriching mental engineering, *International Journal of Innovative Computing, Information and Control* **9**(8) (2013) 3271–3284.
40. H. Xiang, P. Jiang, S. Xiao *et al.*, A model of mental state transition network, *IEEJ Transactions on Electronics Information & Systems* **127**(3) (2007) 434–442.
41. F. Ren, Affective information processing and recognizing human emotion, *Electronic Notes in Theoretical Computer Science* **225** (2009) 39–50.

42. C. Quan and F. Ren, A blog emotion corpus for emotional expression analysis in Chinese, *Computer Speech and Language* **24**(4) (2010) 726–749.
43. F. Ren and K. Matsumoto, Semi-automatic creation of youth slang corpus and its application to affective computing, *IEEE Transactions on Affective Computing* **7**(2) (2016) 176–189.
44. F. Ren and N. Liu, Emotion computing using Word Mover's Distance features based on Ren_CECps, *PLoS ONE* **13**(4) (2018), doi: 10.1371/journal.pone.0194136.
45. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* **77**(2) (1989) 257–286.
46. S. E. Levinson, L. R. Rabiner and M. M. Sondhi, An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition, *Bell Labs Technical Journal* **62**(4) (1982) 1035–1074.
47. L. R. Bahl, F. Jelinek and R. L. Mercer, A maximum likelihood approach to continuous speech recognition, *Readings in Speech Recognition* **5**(2) (1983) 179–190.
48. L. Bahl, P. Brown, P. De Souza and R. Mercer, Maximum mutual information estimation of hidden Markov model parameters for speech recognition, in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86*, Vol. 11 (1986), pp. 49–52.
49. K. F. Lee and H. W. Hon, Speaker-independent phone recognition using hidden Markov models, *IEEE Transactions on Acoustics Speech & Signal Processing* **37**(11) (1989) 1641–1648.
50. X. Huang, Y. Ariki and M. Jack, *Hidden Markov Models for Speech Recognition* (Edinburgh University Press, Edinburgh, 1990).
51. M. Gales and S. Young, The application of hidden Markov models in speech recognition, *Foundations and Trends in Signal Processing* **1**(3) (2007) 195–304.
52. G. E. Dahl, D. Yu, L. Deng and A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing* **20**(1) (2012) 30–42.
53. A. Waibel, T. Hanazawa, G. Hinton *et al.*, Phoneme recognition using time-delay neural networks, *IEEE Transactions on Neural Networks* **1**(2) (1990) 393–404.
54. A. Graves and J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM networks, *IEEE International Joint Conference on Neural Networks*, Vol. 4. (2005), pp. 2047–2052.
55. M. Sugiyama, H. Sawai and A. H. Waibel, Review of TDNN (time delay neural network) architectures for speech recognition, *IEEE International Sympoisum on Circuits and Systems*, Vol. 1 (1991), pp. 582–585.
56. T. N. Sainath, A. R. Mohamed, B. Kingsbury *et al.*, Deep convolutional neural networks for LVCSR, *IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 8614–8618.
57. M. Neumann and N. T. Vu, Attentive convolutional neural network based speech emotion recognition: A study on the impact of input features, signal length, and acted speech (2017), arXiv preprint arXiv:1706.00612.
58. H. Sak, A. Senior and F. Beaufays, Long short-term memory recurrent neural network architectures for large vocabulary speech recognition, *Computer Science* (2014), http://arxiv.org/abs/1402.1128v1.
59. A. Graves, Sequence transduction with recurrent neural networks, *Computer Science* **58**(3) (2012) 235–242.
60. S. Kim and I. Lane, Recurrent models for auditory attention in multi-microphone distant speech recognition, in *17th Annual Conference of the International Speech Communication Association* (2016), pp. 3838–3842.

61. D. Amodei, S. Ananthanarayanan *et al.*, Deep speech 2: End-to-end speech recognition in english and mandarin, *International Conference on Machine Learning*, June 2016, pp. 173–182.
62. A. Hannun, C. Case, J. Casper, B. Catanzaro *et al.*, Deep speech: Scaling up end-to-end speech recognition (2014), arXiv preprint arXiv:1412.5567.
63. K. Rao, H. Sak and R. Prabhavalkar, Exploring architectures, data and units for streaming end-to-end speech recognition with RNN-transducer, *Automatic Speech Recognition and Understanding Workshop (ASRU)*, December 2017, pp. 193–199.
64. N. Carlini and D. Wagner, Audio adversarial examples: Targeted attacks on speech-to-text (2018), arXiv preprint arXiv:1801.01944.
65. V. Chernykh and P. Prikhodko, Emotion recognition from speech with recurrent neural networks (2017), arXiv preprint arXiv:1701.08071.
66. R. Xia and Y. Liu, A multi-task learning framework for emotion recognition using 2d continuous space, *IEEE Transactions on Affective Computing* **8** (2017) 3–14.
67. F. Tao and G. Liu, Advanced LSTM: A study about better time dependency modeling in emotion recognition, in *ICASSP 2018* (Calgary, Canada, April 2018), pp. 1–6.
68. R. Sproat, Multilingual text-to-speech synthesis: The Bell Labs approach, *Computational Linguistics* **3**(4) (1997) 761–764.
69. Y. Tabet and M. Boughazi, Speech synthesis techniques. A survey, *International Workshop on Systems, Signal Processing and Their Applications* (2011), pp. 67–70.
70. C. Khorinphan, S. Phansamdaeng and S. Saiyod, Thai speech synthesis with emotional tone: Based on Formant synthesis for Home Robot, *Student Project Conference* (2014), pp. 111–114.
71. D. Schwarz, Current research in concatenative sound synthesis, *International Computer Music Conference* (2005), pp. 42–45.
72. H. Banno, H. Hata, M. Morise *et al.*, Implementation of realtime STRAIGHT speech manipulation system: Report on its first implementation (Applied Systems), *Acoustical Science & Technology* **28**(3) (2007) 140–146.
73. X. Gonzalvo, S. Tazari, C. A. Chan *et al.*, Recent advances in Google real-time HMM-driven unit selection synthesizer, in *17th Annual Conference of the International Speech Communication Association* (2016), pp. 1–5.
74. K. Tokuda, The HMM-based speech synthesis system (HTS), *Ieice Technical Report Natural Language Understanding & Models of Communication* **107**(406) (2007) 301–306.
75. W. Wang, S. Xu and B. Xu, First step towards end-to-end parametric TTS synthesis: Generating spectral parameters with neural attention, *Interspeech* (2016), pp. 2243–2247.
76. S. Mehri, K. Kumar, I. Gulrajani *et al.*, SampleRNN: An unconditional end-to-end neural audio generation model (2016), arXiv preprint arXiv:1612.07837.
77. A. V. D. Oord, S. Dieleman, H. Zen *et al.*, WaveNet: A generative model for raw audio, in *SSW* (September 2016), pp. 1–15.
78. S. O. Arik, M. Chrzanowski, A. Coates *et al.*, Deep voice: Real-time neural text-to-speech (2017), arXiv preprint arXiv:1702.07825.
79. S. Arik, G. Diamos, A. Gibiansky *et al.*, Deep voice 2: Multi-speaker neural text-to-speech (2017), arXiv preprint arXiv:1705.08947.
80. W. Ping, K. Peng, A. Gibiansky *et al.*, Deep voice 3: Scaling text-to-speech with convolutional sequence learning (2018).
81. Y. Agiomyrgiannakis, Vocaine the vocoder and applications in speech synthesis, *IEEE International Conference on Acoustics, Speech and Signal Processing* (2015), pp. 4230–4234.

82. M. Morise, F. Yokomori and K. Ozawa, WORLD: A vocoder-based high-quality speech synthesis system for real-time applications, *IEICE Transactions on Information & Systems* **99**(7) (2016) 1877–1884.
83. Y. Lee, A. Rabiee and S. Y. Lee, Emotional end-to-end neural speech synthesizer (2017), arXiv preprint arXiv:1711.05447.
84. K. Akuzawa, Y. Iwasawa and Y. Matsuo, Expressive speech synthesis via modeling expressions with variational autoencoder (2018), arXiv:1804.02135.
85. D. A. Reynolds and R. C. Rose, Robust text-independent speaker identification using Gaussian mixture speaker models, *IEEE Transactions on Acoustics Speech Signal Processing* **3** (1995) 72–83.
86. Y. U. Xian, H. E. Song, Y. Peng and W. Zhou, Pattern matching of voiceprint recognition based on GMM, *Communications Technology*, **48**(1) (2015) 97–101.
87. J. Zhang and X. M. Chen, A research of improved algorithm for GMM voiceprint recognition model, *Control and Decision Conference* (2016), pp. 5560–5564.
88. P. Kenny, Joint factor analysis of speaker and session variability: Theory and algorithms, *CRIM*, (2005).
89. Y. Liu, Y. Qian, N. Chen *et al.*, Deep feature for text-dependent speaker verification, *Speech Communication* **73** (2015) 1–13.
90. N. Dehak, P. J. Kenny, R. Dehak *et al.*, Front-end factor analysis for speaker verification, *IEEE Transactions on Audio Speech & Language Processing* **19**(4) (2011) 788–798.
91. O. Ghahabi and J. Hernando, Deep belief networks for i-vector based speaker recognition, *IEEE International Conference on Acoustics, Speech and Signal Processing* (2014), pp. 1700–1704.
92. P. Haffner, G. Tur and J. H. Wright, Optimizing SVMs for complex call classification, *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings*, Vol. 1. (2003), pp. I-632–I-635.
93. R. Sarikaya, G. E. Hinton and A. Deoras, Application of deep belief networks for natural language understanding, *IEEE/ACM Transactions on Audio Speech & Language Processing* **22**(4) (2014) 778–784.
94. A. Ezen-Can and K. E. Boyer, Unsupervised classification of student dialogue acts with query-likelihood clustering, In *Educational Data Mining*, July 2013
95. F. Jiang, X. Chu, X. U. Sheng *et al.*, A macro discourse primary and secondary relation recognition method based on topic similarity, *Journal of Chinese Information Processing*, **32**(1) (2018) 43–50.
96. Y. Du, W. Zhang and T. Liu, Topic augmented convolutional neural network for user interest recognition, *Journal of Computer Research & Development*, **55**(1) (2018) 188–197.
97. F. Ren and H. Yu, Role-explicit query extraction and utilization for quantifying user intents, *Information Sciences*, **329**(C) (2016) 568–580.
98. K. Yao, G. Zweig, M. Y. Hwang *et al.*, Recurrent neural networks for language understanding, *Interspeech* (2013) 2524–2528.
99. G. Mesnil, Y. Dauphin, K. Yao *et al.*, Using recurrent neural networks for slot filling in spoken language understanding, *IEEE/ACM Transactions on Audio Speech & Language Processing* **23**(3) (2015) 530–539.
100. N. Mrkšić, D. O. Séaghdha, B. Thomson *et al.*, Multi-domain dialog state tracking using recurrent neural networks (2015), arXiv preprint arXiv:1506.07190.
101. H. Shi, T. Ushio, M. Endo *et al.*, A multichannel convolutional neural network for cross-language dialog state tracking, *Spoken Language Technology Workshop (SLT)*, December 2016, pp. 559–564.

102. A. Rastogi, D. Hakkani-Tür and L. Heck, Scalable multi-domain dialogue state tracking, *Automatic Speech Recognition and Understanding Workshop* (2018), pp. 561–568.
103. G. Weisz, P. Budzianowski, P. H. Su *et al.*, Sample efficient deep reinforcement learning for dialogue systems with large action spaces, *IEEE/ACM Transactions on Audio Speech & Language Processing* (2018), https://arxiv.org/abs/1802.03753.
104. B. Peng, X. Li, J. Gao *et al.*, Deep Dyna-Q: Integrating planning for task-completion dialogue policy learning (2018), arXiv:1801.06176, 2018.
105. Z. Zhang, M. Huang, Z. Zhao *et al.*, Memory-augmented dialogue management for task-oriented dialogue systems (2018), arXiv:1805.00150.
106. T. H. Wen, M. Gasic, N. Mrksic *et al.*, Semantically conditioned lstm-based natural language generation for spoken dialogue systems (2015), arXiv preprint arXiv:1508.01745.
107. L. Yu, W. Zhang, J. Wang *et al.*, Seqgan: Sequence generative adversarial nets with policy gradient, *AAAI Conference on Artificial Intelligence* (San Francisco, California, USA, 2017), pp. 2851–2858.
108. R. Yan, Y. Song and H. Wu, Learning to respond with deep neural networks for retrieval-based human-computer conversation system, in *Proc. 39th International ACM SIGIR Conf.Research and Development in Information Retrieval* (2016), pp. 55–64.
109. M. Wang, Z. Lu, H. Li and Q. Liu, Syntax-based deep matching of short texts (2015), arXiv preprint arXiv:1503.02427.
110. F. Ren, Y. Wang and C. Quan, A novel factored POMDP model for affective dialogue management, *Journal of Intelligent & Fuzzy Systems* **31**(1) (2016) 127–136.
111. F. Ren, Y. Wang and C Quan, TFSM-based dialogue management model framework for affective dialogue systems, *IEEJ Transactions on Electrical and Electronic Engineering* **10**(4) (2015) 404–410.
112. Y. Wang, F. Ren and C Quan, A new factored POMDP model framework for affective tutoring systems, *IEEJ Transactions on Electrical and Electronic Engineering* **13**(11) (2018) 1603–1611.
113. T. H. Wen, D. Vandyke, N. Mrksic *et al.*, A network-based end-to-end trainable task-oriented dialogue system (2016), arXiv preprint arXiv:1604.04562.
114. B. Liu, G. Tur, D. Hakkani-Tur *et al.*, Dialogue learning with human teaching and feedback in end-to-end trainable task-oriented dialogue systems (2018), arXiv:1804.06512.
115. O. Vinyals and Q. Le, A neural conversational model (2015), arXiv preprint arXiv:1506.05869.
116. A. Sordoni, M. Galley, M. Auli *et al.*, A neural network approach to context-sensitive generation of conversational responses (2015), arXiv preprint arXiv:1506.06714.
117. L. Shang, Z. Lu and H. Li, Neural responding machine for short-text conversation, in *Proc. ACL-IJCNLP* (2015), pp. 1577–1586.
118. V. K. Tran and L. M. N guyen, Neural-based natural language generation in dialogue using RNN encoder-decoder with semantic aggregation (2017), arXiv:1706.06714v2, 2017.
119. J. Li, W. Monroe, A. Ritter *et al.*, Deep reinforcement learning for dialogue generation (2016), arxiv.org/abs/1606.01541.
120. T. Zhao, R. Zhao and M. Eskenazi, Learning discourse-level diversity for neural dialog models using conditional variational autoencoders (2017), arXiv:1703.10960.
121. I. V. Serban, A. Sordoni, R. Lowe *et al.*, A hierarchical latent variable encoder-decoder model for generating dialogues (2016), arXiv:1605.06069.
122. J. Li, W. Monroe, T. Shi *et al.*, Adversarial learning for neural dialogue generation (2017), arxiv.org/abs/1701.06547v1.

123. J. Guo, S. Lu, H. Cai *et al.*, Long text generation via adversarial training with leaked information (2017), arxiv.org/abs/1709.08624v1.
124. E. André, L. Dybkjær, W. Minker *et al.*, *Affective Dialogue Systems, Tutorial and Research Workshop, ADS 2004* (Kloster Irsee, Germany, 2004).
125. D. Ma, S. Li, X. Zhang *et al.*, Interactive attention networks for aspect-level sentiment classification, *Twenty-Sixth International Joint Conf. Artificial Intelligence* (2017), pp. 4068–4074.
126. X. Sun, X. Peng and S. Ding, Emotional human-machine conversation generation based on long short-term memory, *Cognitive Computation* **10**(3) (2018) 389–397.
127. D. Chen, A. Fisch, J. Weston *et al.*, Reading Wikipedia to answer open-domain questions, *Meeting of the Association for Computational Linguistics* (2017), pp. 1870–1879.
128. Y. Lin, H. Ji, Z. Liu and M. Sun, Denoising distantly supervised open-domain question answering, *Meeting of the Association for Computational Linguistics* (2018), pp. 1736–1745.
129. E. Choi, D. Hewlett, J. Uszkoreit *et al.*, Coarse-to-fine question answering for long documents, *Meeting of the Association for Computational Linguistics* (2018), pp. 209–220.
130. W. Wu, X. Sun and H. Wang, Question condensing networks for answer selection in community question answering, *Meeting of the Association for Computational Linguistics* (2018), pp. 1746–1755.
131. G. Salton, A. Wong and C. S. Yang, A vector space model for automatic indexing, *Communications of the ACM* **18**(11) (1975) 613–620.
132. Y. Bengio, R. Ducharme, P. Vincent and C. Jauvin, A neural probabilistic language model, *Journal of Machine Learning Research* **3** (2003) 1137–1155.
133. T. Mikolov, M. Karafiát, L. Burget, J. Cernocký and S. Khudanpur, Recurrent neural network based language model, in *Proc. Interspeech*, Vol. 2 (2010), pp. 1045–1048.
134. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, Distributed representations of words and phrases and their compositionality, in *Proc. Advances Neural Information Processing Systems* (2013), pp. 3111–3119.
135. E. Cambria, S. Poria, A. Gelbukh and M. Thelwall, Sentiment analysis is a big suitcase, *IEEE Intelligent Systems* **32**(6) (2017) 74–80.
136. C. N. Dos Santos and V. Guimaraes, Boosting named entity recognition with neural character embeddings (2015), arXiv preprint, arXiv:1505.05008.
137. C. D. Santos and B. Zadrozny, Learning character level representations for part-of-speech tagging, in *Proc. 31st Int. Conf. Machine Learning* (2014), pp. 1818–1826.
138. T. Gui, Q. Zhang, H. Huang *et al.*, Part-of-speech tagging for twitter with adversarial neural networks, *Conf. Empirical Methods in Natural Language Processing* (2017), pp. 2411–2420.
139. S. Meftah, N. Semmar, O. Zennaki *et al.*, Using transfer learning in part-of-speech tagging of english tweets, *The Language and Technology Conference* (2017), pp. 236–240.
140. M. Maimaiti, A. Wumaier, K. Abiderexiti *et al.*, Bidirectional long short-term memory network with a conditional random field layer for uyghur part-of-speech tagging, *Information* **8**(4) (2017) 157.
141. S. Kumar, M. A. Kumar and K. P. Soman, Deep learning based part-of-speech tagging for malayalam twitter data (Special Issue: Deep Learning Techniques for Natural Language Processing), *Journal of Intelligent Systems* (2018), doi: https://doi.org/10.1515/jisys-2017-0520.
142. R. Collobert, Natural language processing from scratch (2011), arXiv:1103.0398,2011.

143. G. Lample, M. Ballesteros, S. Subramanian *et al.*, Neural architectures for named entity recognition (2016), arXiv:1603.01360.
144. X. Ma and E. Hovy, End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF (2016), arXiv:1603.01354.
145. J. P. Chiu and E. Nichols, Named entity recognition with bidirectional LSTM-CNNs, *Computer Science* (2015), arXiv:1511.08308.
146. M. Rei, G. K. Crichto and S. Pyysalo, Attending to characters in neural sequence labeling models (2016), arXiv:1611.04361.
147. A. Bharadwaj, D. Mortensen, C. Dyer *et al.*, Phonologically aware neural model for named entity recognition in low resource transfer settings, *Conf. Empirical Methods in Natural Language Processing* (2016), pp. 1462–1472.
148. Z. Yang, R. Salakhutdinov and W. W. Cohen, transfer learning for sequence tagging with hierarchical recurrent networks (2017), arXiv:1703.06345.
149. J. Xu, X. Sun, H. He *et al.*, Cross-domain and semi-supervised named entity recognition in chinese social media: A unified model, *IEEE/ACM Transactions on Audio Speech & Language Processing*, **26**(11) (2018) 2142–2152.
150. Y. Shen, H. Yun, Z. C. Lipton *et al.*, Deep active learning for named entity recognition (2018), arXiv:1707.05928.
151. M. M. Mirończuk and J. Protasiewicz, A recent overview of the state-of-the-art elements of text classification, *Expert Systems with Applications* **106** (2018) 36–54.
152. A. Joulin, E. Grave, P. Bojanowski *et al.*, Bag of tricks for efficient text classification (2016), arXiv:1607.01759.
153. Y. Kim, Convolutional neural networks for sentence classification (2014), arXiv preprint arXiv:1408.5882.
154. X. Zhang, J. Zhao and Y. Lecun, Character-level convolutional networks for text classification (2016), arXiv:1509.016262015.
155. Z. Yang, D. Yang, C. Dyer *et al.*, Hierarchical attention networks for document classification, in *Proc. 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2016), pp. 1480–1489.
156. F. Guo, A. Metallinou, C. Khatri *et al.*, Topic-based evaluation for conversational bots (2018), arXiv preprint arXiv:1801.03622.
157. S. Lai, L. Xu, K. Liu and J. Zhao, Recurrent convolutional neural networks for text classification, *AAAI*, Vol. 333, (2015), pp. 2267–2273.
158. J. Howard and S. Ruder, Universal language model fine-tuning for text classification (2018), arXiv:1801.06146.
159. M. V. Mäntylä, D. Graziotin and M. Kuutila, The evolution of sentiment analysis — A review of research topics, venues, and top cited papers, *Computer Science Review* **27** (2018) 16–32.
160. L. Zhang, S. Wang and B. Liu, Deep learning for sentiment analysis: A survey, *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery* (2018).
161. S. Zhai and Z. M. Zhang, Semisupervised autoencoder for sentiment analysis, *Thirtieth AAAI Conf. Artificial Intelligence* (AAAI Press, 2016), pp. 1394–1400.
162. C. Quan and F. Ren, Textual emotion recognition for enhancing enterprise computing, *Enterprise Information Systems* **10**(4) (2016) 422–443.
163. R Fuji and K. Matsumoto, Emotion analysis on social big data, *ZTE Communications* **15**(S2) (2017) 30–37.
164. C. Quan and F. Ren, Feature-level sentiment analysis by using comparative domain corpora, *Enterprise Information Systems* **10**(5) (2016) 505–522.
165. F. Ren and L. Wang, Sentiment analysis of text based on three-way decisions, *Journal of Intelligent and Fuzzy Systems* **33**(1) (2017) 245–254.

166. F. Ren and J. Deng, Background knowledge based multi-stream neural network for text classification, *Applied Sciences* **8**(12) (2018), https://doi.org/10.3390/app8122472.
167. F. Ren, X. Kang and C Quan, Examining accumulated emotional traits in suicide blogs with an emotion topic model, *IEEE Journal of Biomedical and Health Informatics* **20**(5) (2016) 1384–1396.
168. X. Kang, F. Ren and Y. Wu, Exploring latent semantic information for textual emotion recognition in blog articles, *IEEE/CAA Journal of Automatics Sinica* **5**(1) (2018) 204–216.
169. F. Ren and Y. Wu, Predicting user-topic opinions in Twitter with social and topical context, *IEEE Transactions on Affective Computing* **4**(4) (2013) 412–424.
170. F. Ren and X. Kang, Employing hierarchical Bayesian networks in simple and complex emotion topic analysis, *Computer Speech and Language* **27**(4) (2013) 943–968.
171. E. Cambria, S. Poria, A. Gelbukh *et al.*, Sentiment analysis is a big suitcase, *IEEE Intelligent Systems* **32**(6) (2018) 74–80.
172. S. Zheng, F. Wang, H. Bao *et al.*, Joint extraction of entities and relations based on a novel tagging scheme (2017), arXiv:1706.05075.
173. H. Schwenk, Continuous space translation models for phrase-based statistical machine translation, *COLING 2012: Posters* (2012), pp. 1071–1080.
174. K. Cho, B. V. Merrienboer, C. Gulcehre *et al.*, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in *EMNLP* (2014), pp. 1–15.
175. I. Sutskever, O. Vinyals and Q. V. Le, Sequence to sequence learning with neural networks, in *Advances in Neural Information Processing Systems*, Vol. 4 (2014), pp. 3104–3112.
176. D. Bahdanau, K. Cho and Y. Bengio, Neural machine translation by jointly learning to align and translate, in *ICLR* (2015), pp. 1–15.
177. M. T. Luong, H. Pham and C. D. Manning, Effective approaches to attention-based neural machine translation, in *EMNLP* (2015), pp. 1–11.
178. J. Gehring, M. Auli, D. Grangier, D. Yarats and Y. Dauphin, Convolutional sequence to sequence learning (2017), arXiv preprint arXiv:1705.03122v2.
179. A. Vaswani, N. Shazeer, N. Parmar *et al.*, Attention is all you need, in *Advances in Neural Information Processing Systems*, Vol. 10 (2017), pp. 5998–6008.
180. Y. Wu, M. Schuster, Z. Chen *et al.*, Google's Neural machine translation system: Bridging the gap between human and machine translation (2016), arXiv:1609.08144.
181. M. Johnson, M. Schuster, Q. V. Le *et al.*, Google's multilingual neural machine translation system: Enabling zero-shot translation (2017), arXiv:1611.04558.
182. Z. Yang, W. Chen, F. Wang *et al.*, Improving neural machine translation with conditional sequence generative adversarial nets, *NAACL2018* (2018), pp. 1356–1365.
183. J. Gu, J. Bradbury, C. Xiong *et al.*, Non-autoregressive neural machine translation, *ICLR* (2018), pp. 1–13.
184. F. Ren and D. B. Bracewell, Advanced information retrieval, *Electronic Notes in Theoretical Computer Science* **225**(1) (2009) 303–317.
185. M. Richardson, C. J. Burges and E. Renshaw, Mctest: A challenge dataset for the open-domain machine comprehension of text, in *Proc. 2013 Conference on Empirical Methods in Natural Language Processing* (2013), pp. 193–203.
186. F. Hill, A. Bordes, S. Chopra *et al.*, The Goldilocks Principle: Reading children's books with explicit memory representations, in *ICLR* (2016), pp. 1–13.
187. K. M. Hermann, T. Kocisky, E. Grefenstette *et al.*, Teaching machines to read and comprehend, in *Advances in Neural Information Processing Systems*, Vol. 4 (2015), pp. 1693–1701.

188. P. Rajpurkar, J. Zhang, K. Lopyrev *et al.*, SQuAD: 100,000+ Questions for Machine Comprehension of Text (2016), arXiv:1606.05250.
189. A. Miller, A. Fisch, J. Dodge *et al.*, Key-value memory networks for directly reading documents (2016), arXiv:1606.03126.
190. T. Nguyen, M. Rosenberg, X. Song *et al.*, MS MARCO: A human generated machine reading comprehension dataset (2016), arXiv:1611.09268.
191. W. He, K. Liu, J. Liu *et al.*, DuReader: A Chinese machine reading comprehension dataset from real-world applications, *ACL* (2018), pp. 1–10.
192. W. Yin, S. Ebert and H. Schütze, Attention-based convolutional neural network for machine comprehension (2016), arXiv preprint arXiv:1602.04341.
193. S. Wang and J. Jiang, Learning Natural Language Inference with LSTM (2016), arXiv:1512.08849.
194. S. Wang and J. Jiang, Machine comprehension using Match-LSTM and answer pointer, in *ICLR* (2017), pp. 1–15.
195. Microsoft Asia Natural Language Computing Group. R-net: Machine reading comprehension with self-matching networks (2017).
196. M. Seo, A. Kembhavi, A. Farhadi *et al.*, Bidirectional attention flow for machine comprehension, *ICLR* (2017), pp. 1–13.
197. C. Xiong, V. Zhong and R. Socher, DCN+: Mixed objective and deep residual coattention for question answering (2017), arXiv:1711.00106.
198. M. Hu, Y. Peng, Z. Huang *et al.*, Reinforced mnemonic reader for machine reading comprehension (2017), arXiv preprint arXiv:1705.02798.
199. Y. Cui, Z. Chen, S. Wei *et al.*, Attention-over-attention neural networks for reading comprehension (2017), arXiv:1607.04423.
200. D. Golub, P. Huang, X. He *et al*, Two-stage synthesis networks for transfer learning in machine comprehension (2017), arXiv:1706.09789.
201. Y. Xu, J. Liu, J. Gao *et al.*, Dynamic fusion networks for machine reading comprehension (2018), arXiv:1711.04964v2.
202. C. Clark and M. Gardner, Simple and effective multi-paragraph reading comprehension (2017), arXiv:1710.10723.
203. J. Welbl, P. Stenetorp and S. Riedel, Constructing datasets for multi-hop reading comprehension across documents (2018), arXiv:1710.06481.
204. H. P. Luhn, The automatic creation of literature abstracts, *Ibm Journal of Research & Development* **2**(2) (1958) 159–165.
205. D. Shen, J. Sun, H. Li *et al.*, Document summarization using conditional random fields, *IJCAI*, Vol. 7, (2007), pp. 2862–2867.
206. Y. Ouyang, W. Li, S. Li and Q. Lu, Applying regression models to query-focused multi-document summarization, *Information Processing & Management* **47**(2) (2011) 227–237.
207. Z. Cao, F. Wei, L. Dong *et al.*, Ranking with recursive neural networks and its application to multi-document summarization, in *Twenty-Ninth AAAI Conf. Artificial Intelligence* (2015), pp. 2153–2159.
208. D. Bollegala, N. Okazaki and M. Ishizuka, A bottom-up approach to sentence ordering for multi-document summarization, *Information Processing & Management* **46**(1) (2010) 89–109.
209. C. Li, X. Qian and Y. Liu, Using supervised bigram-based ILP for extractive summarization, *ACL* (2013), pp. 1004–1013.
210. H. Lin and J. Bilmes, A class of submodular functions for document summarization, in *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (2011), pp. 510–520.

211. X. Qian and Y. Liu, Fast joint compression and summarization via graph cuts, *EMNLP* (2013), pp. 1492–1502.
212. C. Li, Y. Liu, F. Liu *et al.*, Improving multi-documents summarization by sentence compression based on expanded constituent parse trees, in *EMNLP* (2014), pp. 691–701.
213. L. Bing, P. Li, Y. Liao *et al.*, Abstractive multi-document summarization via phrase selection and merging, *Computational Linguistics* **31**(4) (2015) 505–530.
214. S. Dohare, H. Karnick and V. Gupta, Text summarization using abstract meaning representation (2017), arXiv:1706.01678.
215. Y. Xia, F. Tian, L. Wu *et al.*, Deliberation networks: Sequence generation beyond one-pass decoding, *NIPS* (2017), pp. 1–11.
216. J. Gehring, M. Auli, D. Grangier *et al.*, A convolutional encoder model for neural machine translation (2016) arXiv:1611.02344.
217. K. Lin, D. Li, X. He *et al.*, Adversarial ranking for language generation (2018), arXiv:1705.11001.
218. R. Paulus, C. Xiong and R. Socher, A deep reinforced model for abstractive summarization (2017), arXiv preprint arXiv:1705.04304.
219. A. Abu-Jbara and D. Radev, Coherent citation-based summarization of scientific papers, in *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (2011), pp. 500–509.
220. S. Wang, X. Wan and S. Du, Phrase-based presentation slides generation for academic papers, *AAAI* (2017), pp. 196–202.
221. W. Luo, F. Liu, Z. Liu *et al.*, Automatic summarization of student course feedback, *Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2018), pp. 80–85.
222. E. Goldberg, R. Kittredge and A. Polguère, Computer generation of marine weather forecast text, *Journal of Atmospheric and Oceanic Technology* **5**(4) (2009) 473–483.
223. J. Zhang, J. Yao and X. Wan. Toward constructing sports news from live text commentary, *ACL* **16** (2016) 1361–1371.
224. R. Lebret, D. Grangier and M. Auli, Neural text generation from structured data with application to the biography domain, *EMNLP* (2016), pp. 1–11.
225. Z. Wang, W. He, H. Wu *et al.*, Chinese poetry generation with planning based neural network, *COLING* (2016), pp. 1051–1060.
226. J. Zhang, Y. Feng, D. Wang *et al.*, Flexible and creative Chinese poetry generation using neural memory, *ACL* (2017), pp. 1364–1373.
227. L. Xu, L. Jiang, C. Qin *et al.*, How images inspire poems: Generating classical Chinese poetry from images with memory networks (2018), arXiv:1803.02994.
228. X. Chen and C. L. Zitnick, Mind's eye: A recurrent visual representation for image caption generation, *CVPR* (2015), pp. 2422–2431.
229. P. Anderson, X. He, C. Buehler *et al.*, Bottom-up and top-down attention for image captioning and VQA (2017), arXiv preprint arXiv:1707.07998.
230. S. Liu, Z. Zhu, N. Ye *et al.*, Improved image captioning via policy gradient optimization of spider, in *Proc. IEEE Int. Conf. Comp. Vis*, Vol. 3 (2017), pp. 873–881.
231. S. Lee, U. Hwang, S. Min *et al.*, Polyphonic music generation with sequence generative adversarial networks (2018), arXiv:1710.11418.
232. H. Dong and Y. Yang, Convolutional generative adversarial networks with binary neurons for polyphonic music generation, *ISMIR* (2018), pp. 1–13.
233. J. C. García and E. Serrano, Automatic music generation by deep learning, *Int. Symp. Distributed Computing and Artificial Intelligence* (Springer, 2018), pp. 284–291.
234. M. Wang and W. Deng, Deep face recognition: A survey (2018), arXiv:1804.06655.

235. H. Jiang and E. Learned-Miller, Face detection with the faster R-CNN, *IEEE Int. Conf. Automatic Face & Gesture Recognition* (2017), pp. 650–657.
236. S. Yang, P. Luo, C. Loy *et al.*, Faceness-Net: Face detection through deep facial part responses, *IEEE Transactions on Pattern Analysis & Machine Intelligence* **40**(8) (2017) 1845–1859.
237. Y. Zhou, D. Liu and T. Huang, Survey of face detection on low-quality images, *IEEE Int. Conf. Automatic Face & Gesture Recognition* (2018), pp. 769–773.
238. X. Jin and X. Tan, Face alignment in-the-wild: A survey, *Computer Vision & Image Understanding* **162**(9) (2017) 1–22.
239. N. Wang, X. Gao, D. Tao *et al.*, Facial feature point detection: A comprehensive survey, *Neurocomputing* **275** (2018) 50–65.
240. X. Zou, J. Kittler and K. Messer, Illumination invariant face recognition: A survey, in *BTAS* (2017), pp. 1–8.
241. R. Jafri and H. R. Arabnia, A survey of face recognition techniques, *Jips* **5**(2) (2009) 41–68.
242. M. Wang and W. Deng, Deep face recognition: A survey (2018), arXiv:1804.06655.
243. M. D. Marsico, A. Petrosino and S. Ricciardi, Iris recognition through machine learning techniques: A survey, *Pattern Recognition Letters* **82** (2016) 106–115.
244. K. Nguyen, C. Fookes, R. Jillela *et al.*, Long range Iris recognition: A survey, *Pattern Recognition* **72** (2017) 123–143.
245. A. K. Jain, S. S. Arora, K. Cao *et al.*, Fingerprint recognition of young children, *IEEE Transactions on Information Forensics & Security* **12**(7) (2017) 1501–1514.
246. K. Cao and A. K. Jain, Automated latent fingerprint recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence* **41**(4) (2018) 788–800.
247. U. A. Soni and M. M. Goyani, A survey on state of the art methods of fingerprint recognition (2018), http://ijsrset.com/paper/3542.pdf.
248. J. Parvathy, A survey on fingerprint identification techniques (2018), doi: 10.22214/ijraset.2018.1267.
249. Z. Q. Wang and I. Tashev, Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks, *IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (2017), pp. 5150–5154.
250. R. Ranjan, V. M. Patel and R. Chellappa, HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2017), doi: 10.1109/TPAMI.2017.2781233.
251. F. Ren and Z. Huang, Facial expression recognition based on AAM-SIFT and adaptive regional weighting, *IEEJ Transactions on Electrical and Electronic Engineering* **10**(6) (2015) 713–722.
252. A. Mollahosseini, D. Chan and M. H. Mahoor, Going deeper in facial expression recognition using deep neural networks, in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, March 2016, pp. 1–10.
253. A. T. Lopes, E. D. Aguiar, A. F. D. Souza *et al.*, Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order, *Pattern Recognition* **61** (2017) 610–628.
254. J. Chen, Z. Chen, Z. Chi *et al.*, Facial expression recognition in video with multiple feature fusion, *IEEE Transactions on Affective Computing* **9**(1) (2018) 38–50.
255. H. Ding, S. K. Zhou and R. Chellappa, FaceNet2ExpNet: Regularizing a deep face recognition net for expression recognition, *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition* (2017), pp. 118–126.

256. X. Ben, X. Jia, R. Yan *et al.*, Learning effective binary descriptors for micro-expression recognition transferred by macro-information, *Pattern Recognition Letters* **107** (2018) 50–58.
257. Y. J. Liu, J. K. Zhang, W. J. Yan *et al.*, A main directional mean optical flow feature for spontaneous micro-expression recognition, *IEEE Transactions on Affective Computing* **7**(4) (2016) 299–310.
258. X. Huang, S. J. Wang, G. Zhao *et al.*, Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection, *The Workshop on Computer Vision for Affective Computing at ICCV* (IEEE Computer Society, 2015), pp. 1–13.
259. I. Nejadgholi, S. A. Seyyedsalehi and S. Chartier, A brain-inspired method of facial expression generation using chaotic feature extracting bidirectional associative memory, *Neural Processing Letters* **46**(3) (2017) 943–960.
260. B. Nojavanasghari, Y. Huang and S. Khan, Interactive generative adversarial networks for facial expression generation in dyadic interactions, (2018) arXiv:1801.09092.
261. M. Xue, S. Tokai and H. Hase, Point clouds based 3D facial expression generation, *Int. Conf. Mechanical Design* (Springer, Singapore, 2017), pp. 467–484.
262. N. Liu, and F. Ren, Emotion classification using a CNN_LSTM based model for smooth emotional synchronization of the humanoid robot REN-XIN, *PLoS ONE*, **14**(5) (2019), https://doi.org/10.1371/journal. pone.0215216.
263. R. Amini, C. Lisetti and G. Ruiz, HapFACS 3.0: FACS-based facial expression generator for 3D speaking virtual characters, *IEEE Transactions on Affective Computing* **6**(4) (2017) 348–360.
264. X. Yan, J. Yang, K. Sohn *et al.*, Attribute2Image: Conditional image generation from visual attributes, *European Conference on Computer Vision* (Springer International Publishing, 2016) pp. 776–791.
265. A. Dash, J. C. B. Gamboa, S. Ahmed *et al.*, TAC-GAN - text conditioned auxiliary classifier generative adversarial network (2017), arXiv:1703.06412.
266. C. Yan, B. Zhang and F. Coenen, Driving posture recognition by convolutional neural networks, *Int. Conf. Natural Computation* (IEEE, 2016), pp. 680–685.
267. Y. Sun, C. Li, G. Li *et al.*, Gesture recognition based on Kinect and sEMG signal fusion, *Mobile Networks & Applications* **23** (2018) 797–805.
268. X. Yan, H. Li, C. Wang *et al.*, Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion, *Advanced Engineering Informatics* **34** (2017) 152–163.
269. F. Noroozi, C. A. Corneanu, D. Kamińska *et al.*, Survey on emotional body gesture recognition (2018) arXiv:1801.07481.
270. H. Wang, P. Wang, Z. Song *et al.*, Large-scale multimodal gesture recognition using heterogeneous networks, *IEEE Int. Conf. Computer Vision Workshop* (IEEE, 2018), pp. 3129–3137.
271. L. Pigou, A. V. D. Oord, S. Dieleman *et al.*, Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video, *International Journal of Computer Vision* **126**(2-4) (2018) 430–439.
272. S. F. Chevtchenko, R. F. Vale and V. Macario, Multi-objective optimization for hand posture recognition, *Expert Systems with Applications* **92** (2017) 170–181.
273. B. K. Chakraborty, D. Sarma, M. K. Bhuyan *et al.*, Review of constraints on vision-based gesture recognition for human–computer interaction, *IET Computer Vision* **12**(1) (2018) 3–15.
274. J. Xiahou, H. He, K. Wei *et al.*, Integrated approach of dynamic human eye movement recognition and tracking in real time, *Int. Conf. Virtual Reality and Visualization* (IEEE, 2017), pp. 94–101.

275. X. Ding, Z. Lv, C. Zhang *et al.*, A robust online saccadic eye movement recognition method combining electrooculography and video, *IEEE Access* **5** (2017) 17997–18003.
276. A. Chaudhuri, K. Mandaviya, P. Badelia *et al.*, *Optical Character Recognition Systems for Different Languages with Soft Computing*, Vol. 352 (Springer, 2017) ISBN 978-3-319-50251-9.
277. A. Marial and J. Jos, Feature extraction of optical character recognition: Survey, *International Journal of Applied Engineering Research* **12**(7) (2017) 1129–1137.
278. M. Brisinello, R. Grbic, M. Pul *et al.*, Improving optical character recognition performance for low quality images, in *Proc. ELMAR-2017* (IEEE, 2017), pp. 167–171.
279. D. Lin, F. Lin, Y. Lv *et al.*, Chinese character CAPTCHA recognition and performance estimation via deep neural network, *Neurocomputing* **288** (2018) 11–19.
280. X. Y. Zhang, F. Yin, Y. M. Zhang *et al.*, Drawing and recognizing chinese characters with recurrent neural network, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40**(4) (2018) 849–862.
281. K. Han and H. H. Chang, Digits generation and recognition using RNNPB (2018).
282. H. D. Critchley and S. N. Garfinkel, The influence of physiological signals on cognition, *Current Opinion in Behavioral Sciences* **19** (2018) 13–18.
283. D. Zhang, L. Yao, X. Zhang *et al.*, Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface, *AAAI Conf. Artificial Intelligence* (2018), pp. 1–8.
284. S. Wang, J. Gwizdka and W. A. Chaovalitwongse, Using wireless EEG signals to assess memory workload in the, n-back task, *IEEE Transactions on Human-Machine Systems* **46**(3) (2016) 424–435.
285. T. Chen, S. Ju, X. Yuan, M. Elhoseny, F. Ren and M. Fan, Emotion recognition using empirical mode decomposition and approximation entropy, *Computers and Electrical Engineering* **72** (2018) 383–392.
286. F. Ren, Y. Dong and W. Wang, Emotion recognition based on physiological signals using brain asymmetry index and echo state network, *Neural Computing and Applications* (2018), https://doi.org/10.1007/s00521-018-3831-4.
287. M. Soleymani, S. Asghari-Esfeden, Y. Fu *et al.*, Analysis of EEG signals and facial expressions for continuous emotion detection, *IEEE Transactions on Affective Computing* **7**(1) (2016) 17–28.
288. L. Shu, J. Xie, M. Yang *et al.*, A review of emotion recognition using physiological signals, *Sensors* **18**(7) (2018), doi.org/10.3390/s18072074.
289. M. Mahmud, M. S. Kaiser, A. Hussain *et al.*, Applications of deep learning and reinforcement learning to biological data, *IEEE Transactions on Neural Networks & Learning Systems* **29**(6) (2018) 2063–2079.
290. I. D. Castro, C. Varon, T. Torfs *et al.*, Evaluation of a multichannel non-contact ECG system and signal quality algorithms for sleep apnea detection and monitoring, *Sensors* **18**(2) (2018), doi.org/10.3390/s18020577.
291. N. Dey, A. S. Ashour, F. Shi *et al.*, Developing residential wireless sensor networks for ECG healthcare monitoring, *IEEE Transactions on Consumer Electronics* **63**(4) (2018) 442–449.
292. L. Liu, X. Chen, Z. Lu *et al.*, Development of an EMG-ACC-based upper limb rehabilitation training system, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(3) (2017) 244–253.
293. Y. Hu, Z. Li, G. Li *et al.*, Development of sensory-motor fusion-based manipulation and grasping control for a robotic hand-eye system, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **47**(7) (2017) 1169–1180.

294. T. Kapelner, F. Negro, O. C. Aszmann and D. Farina, Decoding motor unit activity from forearm muscles: Perspectives for myoelectric control, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(1) (2018) 244–251.
295. T. Teramae, T. Noda and J. Morimoto, EMG-based model predictive control for physical human–robot interaction: Application for assist-as-needed control, *IEEE Robotics and Automation Letters* **3**(1) (2018) 210–217.
296. C. Yang, C. Zeng, P. Liang *et al.*, Interface design of a physical human-robot interaction system for human impedance adaptive skill transfer, *IEEE Transactions on Automation Science and Engineering* **15**(1) (2018) 329–340.
297. Z. Lu, X. Chen, Q. Li *et al.*, A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices, *IEEE Transactions on Human-Machine Systems* **44**(2) (2017) 293–299.
298. L. Y. Deng, C. L. Hsu, T. C. Lin *et al.*, EOG-based human–computer interface system development, *Expert Systems with Applications* **37**(4) (2010) 3337–3343.
299. K. R. Lee, W. D. Chang, S. Kim *et al.*, Real-time "eye-writing" recognition using electrooculogram, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(1) (2017) 37–48.
300. H. Monkaresi, N. Bosch, R. Calvo *et al.*, Automated detection of engagement using video-based estimation of facial expressions and heart rate, *IEEE Transactions on Affective Computing* **8**(1) (2017) 15–28.
301. A. Procházka, M. Schätz, O. Vyšata *et al.*, Microsoft kinect visual and depth sensors for breathing and heart rate analysis, *Sensors* **16**(7) (2016) 996–1006.
302. M. V. Villarejo, B. G. Zapirain and A. M. Zorrilla, A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee, *Sensors* **12**(5) (2012) 6075–6101.
303. C. Mundell, J. P. Vielma, T. Zaman, Predicting performance under stressful conditions using galvanic skin response (2016), arXiv preprint arXiv:1606.01836.
304. X. Sun, T. Hong, C. Li and F Ren, Hybrid spatiotemporal models for sentiment classification via galvanic skin response, *Neurocomputing* **358** (2019) 385–400.
305. C. G. Núñez, W. T. Navaraj, E. O. Polat *et al.*, Energy—autonomous, flexible, and transparent tactile skin, *Advanced Functional Materials* **27**(18) (2017), doi.org/10.1002/adfm.201606287.
306. T. Paulino, P. Ribeiro, M. Neto *et al.*, Low-cost 3-axis soft tactile sensors for the human-friendly robot Vizzy, *IEEE Int. Conf. Robotics and Automation* (IEEE, 2017), pp. 966–971.
307. M. Kaboli, D. Feng, K. Yao *et al.*, A tactile-based framework for active object learning and discrimination using multimodal robotic skin, *IEEE Robotics and Automation Letters* **2**(4) (2017) 2143–2150.
308. H. Liu, Y. Yu, F. Sun *et al.*, Visual–tactile fusion for object recognition, *IEEE Transactions on Automation Science and Engineering* **14**(2) (2017) 996–1008.
309. A. Schmitz, P. Maiolino, M. Maggiali *et al.*, Methods and technologies for the implementation of large-scale robot tactile sensors, *IEEE Transactions on Robotics* **27**(3) (2011) 389–400.
310. Y. Gu, J. Zhan, Y. Ji, J. Li, F. Ren and S Gao, MoSense: An RF-based motion detection system via off-the-shelf WiFi devices, *IEEE Internet of Things Journal* **4**(6) (2017) 2326–2341.
311. Y. Gu, F. Ren and J. Li, PAWS: Passive human activity recognition based on WiFi ambient signals, *IEEE Internet of Things Journal* **3**(5) (2016) 796–805.
312. Y. Gu, J. Zhan, J. Li, Y. Ji, X. An and F Ren, Sleepy: Wireless channel data driven sleep monitoring via commodity WiFi devices, *IEEE Transactions on Big Data* (2018), 10.1109/TBDATA.2018.2851201.

313. U. Saranli, M. Buehler and D. E. Koditschek, RHex: A simple and highly mobile hexapod robot, *The International Journal of Robotics Research* **20**(7) (2001) 616–631.
314. M. Raibert, K. Blankespoor, G. Nelson *et al.*, Bigdog, the rough-terrain quadruped robot, *IFAC Proceedings Volumes* **41**(2) (2008) 10822–10825.
315. M. P. Murphy, A. Saunders, C. Moreira *et al.*, The littledog robot, *The International Journal of Robotics Research* **30**(2) (2011) 145–149.
316. S. Kuindersma, R. Deits, M. Fallon *et al.*, Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot, *Autonomous Robots* **40**(3) (2016) 429–455.
317. L. R. Hochberg, D. Bacher, B. Jarosiewicz *et al.*, Reach and grasp by people with tetraplegia using a neurally controlled robotic arm, *Nature* **485** (2012) 372–375.
318. G. Santhanam, S. I. Ryu, B. M. Yu *et al.*, A high-performance brain–computer interface, *Nature* **442**(7099) (2006) 195–198.
319. J. Dobson, Remote control of cellular behaviour with magnetic nanoparticles, *Nature Nanotechnology* **3**(3) (2008) 139–143.
320. P. Ohta, L. Valle, J. King *et al.*, Design of a lightweight soft robotic arm using pneumatic artificial muscles and inflatable sleeves, *Soft Robotics* **5**(2) (2018) 204–215.
321. A. B. Ajiboye, F. R. Willett, D. R. Young *et al.*, Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: A proof-of-concept demonstration, *The Lancet* **389**(10081) (2017) 1821–1830.
322. OctopusGripper, https://www.festo.com/group/en/cms/12745.htm.
323. R. Deimel and O. Brock, A novel type of compliant and underactuated robotic hand for dexterous grasping, *The International Journal of Robotics Research* **35**(1–3) (2016) 161–185.
324. G. Kou, Y. Lu, Y. Peng *et al.*, Evaluation of classification algorithms using MCDM and rank correlation, *International Journal of Information Technology & Decision Making* **11**(01) (2012) 197–225.
325. G. Li, G. Kou and Y. Peng, A group decision making model for integrating heterogeneous information, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **48**(6) (2018) 982–992.
326. H. Zhang, G. Kou and Y. Peng, Soft consensus cost models for group decision making and economic interpretations, *European Journal of Operational Research* **227**(3) (2019) 964–980.
327. G. Kou, D. Ergu, C. Lin *et al.*, Pairwise comparison matrix in multiple criteria decision making, *Technological & Economic Development of Economy* **22**(5) (2016) 738–765.
328. G. Kou, Y. Peng and G. Wang, Evaluation of clustering algorithms for financial risk analysis using MCDM methods, *Information Sciences* **275**(11) (2014) 1–12.