

DGCC: data-driven granular cognitive computing

Guoyin Wang¹ 

Received: 1 May 2017 / Accepted: 19 July 2017 / Published online: 27 July 2017
© Springer International Publishing AG 2017

Abstract 2016 is the 60 anniversary of the birth of artificial intelligence (AI). In the past 60 years, many AI theoretical models have been proposed and great achievements have been accomplished. Most intelligent computing models are inspired by various human/ natural/social intelligence mechanisms. Three main schools of artificial intelligence are formed, that is, symbolism, connectionism and behaviorism. Artificial intelligence is intelligence exhibited by machines. It is applied when a machine mimics cognitive functions that humans associate with other human minds. Achievements of cognitive science could give much inspiration to AI. Cognitive computing is one of the core fields of artificial intelligence. It aims to develop a coherent, unified, universal mechanism inspired by human mind's capabilities. It is one of the most critical tasks for artificial intelligence researchers to develop advanced cognitive computing models. Cognitive computing is the third and most transformational phase in computing's evolution, after the two distinct eras of computing—the tabulating era and the programming era. Inspired by human's granularity thinking, problem solving mechanism and the cognition law of “global precedence”, a new powerful cognitive computing model, data-driven granular cognitive computing (DGCC), is proposed in this paper. It takes data as a special kind of knowledge

expressed in the lowest granularity level of a multiple granularity space. It integrates two contradictory mechanisms, namely, the human's cognition mechanism of “global precedence” which is a cognition process of “from coarser to finer” and the information processing mechanism of machine learning systems which is “from finer to coarser”, in a multiple granularity space. It is also based on the idea of data-driven. The research issues of DGCC to be further addressed are discussed. Based on DGCC, deep learning is neither classified into symbolism, nor connectionism. It is taken as a combination of symbolism and connectionism, and named hierarchical structuralism in this paper. The HD³ characteristics (hierarchical, distributed, data-driven, and dynamical) of the hierarchical structuralism are analyzed. DGCC provides a granular cognitive computing framework for efficient knowledge discovery from big data.

Keywords Granular cognitive computing · Cognitive computing · Granular computing · Data-driven · Hierarchical structuralism · Artificial intelligence · DGCC

1 Introduction

The field of artificial intelligence (AI) research was born at a conference in Dartmouth College in 1956 (Crevier 1993). The attendees, included McCarthy, Newell, Minsky, Samuel and Simon, who became the founders and early leaders of AI research. In the past 60 years, many artificial researchers have tried to develop intelligent computing models with inspiration of various human/natural/social intelligence mechanisms. Three main schools of artificial intelligence are formed; namely, symbolism,

The paper is an extension based on the keynote talk “Data-driven Granular Cognitive Computing” at IJCRS2017 (International Joint Conference on Rough Sets, July 3–7 2017, Olsztyn, Poland).

✉ Guoyin Wang
wanggy@ieec.org

¹ Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Nan'an District, Chongqing, People's Republic of China

connectionism and behaviorism. Many milestone AI achievements have been accomplished.

In the middle 1950s, AI researchers began to explore the possibility that human intelligence could be reduced to symbol manipulation. It is called the symbolism of AI. Newell and Simon introduced the physical symbol system hypothesis in 1976 (Simon 1996; Newell and Simon 1976). Human cognition is taken as a kind of symbolic processing, and the processes of human thinking could be computed by symbol (Simon 1996; Newell and Simon 1976). Logic programming science and tools like Prolog and LISP are its typical representative results. Feigenbaum introduced expert systems, which is a form of AI program that simulates the knowledge and analytical skills of human experts (Jackson 1998; Leondes 2001). It is the first truly successful form of AI software. It has many successful applications in various fields. More and more expert systems are being developed until now.

The connectionism of AI was established by McClelland and Rumelhart in the 1980s (Rumelhart and McClelland 1986). The emergence of the connectionist paradigm largely resulted from various dissatisfactions with the symbolism of AI. It is a set of approaches in the fields of artificial intelligence, cognitive psychology, cognitive science, neuroscience, and philosophy of mind, which models the mental or behavioral phenomena as the emergent processes of interconnected networks of simple units. It relies on the bionics to simulate human intelligence (brain) and considers the intelligence as the result of interconnected neurons' competition and collaboration. Artificial neural network is its most common form (Haykin 1994). There are many typical artificial neural network models, such as, back-propagation neural network (Rumelhart et al. 1986), Hopfield neural network (Hopfield 1982), Kohonen self-organizing maps (Kohonen 1982), Boltzmann machine (Ackley et al. 1985), radial basis function network (Broomhead and Lowe 1988). Artificial neural networks have been applied successfully in a lot of real life areas, such as system identification and control, quantum chemistry, game-playing and decision making, pattern recognition, sequence recognition, medical diagnosis, financial applications, data mining, visualization, etc.

The behaviorism of AI considers that intelligence depends on the perception and behavior. Intelligent behavior can be manifested through interaction between real-world and surrounding environment (Skinner 2011). "Perception-action" model was proposed. In the behaviorism of AI, intelligence may not require knowledge, knowledge representation or knowledge reasoning. It studies and simulates the intelligent behaviors in auto-control processes, such as self-optimizing, self-adaption, self-learning and self-organization. In 1988, Brooks invented the Hexapod Walking Robot, which was

composed of 150 sensors and 23 actuators (Brooks 1991). It is a very successful example of behaviorism in that period. The idea of "Perception-action" has often been applied in auto-control systems.

In the recent 20 years, with the rapid development of high-performance computing, internet, internet-of-things, cloud computing and big data, AI has more and more great achievements on solving some specific problems. Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov on 11 May 1997. In February 2011, IBM's question answering system, Watson, defeated two greatest Jeopardy champions, Rutter and Jennings. In March 2016, AlphaGo defeated Sedol Lee, becoming the first computer Go-playing system to beat a professional Go player. In 2017, the new version of AlphaGo defeated over 60 professional Go players. Several kinds of unmanned intelligent vehicles are running on road successfully nowadays.

The development of artificial intelligence has been accompanied with the development of computer science in the past 60 years. Kelly introduced the following three phases in computing's evolution as shown in Fig. 1 (Kelly III 2015). It would also be very important for artificial intelligence to move into the cognitive era. Cognition inspired artificial intelligence technologies will be developed quickly.

Through studying the basic ideas and typical models of cognitive computing (Kelly III 2015) and granular computing (Yao et al. 2013), with the inspiration of a typical human cognition model, a novel powerful cognitive computing model, data-driven granular cognitive computing (DGCC), is proposed based on the idea of data-driven. It takes data as a special kind of knowledge expressed in the lowest granularity level of a multiple granularity space. Its triangular structure is shown in Fig. 2. It integrates the traditional data-driven bottom-up information computing mechanism of machine learning/data mining systems and an important top-down human cognition law of "global precedence" (Han and Chen 1996; Chen et al. 2003). The triangular structure of DGCC will be further analyzed in the Sect. 4 of this paper.

Deep learning (Hinton and Salakhutdinov 2006; Hinton et al. 2006; LeCun et al. 2015), an advanced statistical machine learning model, access to large amounts of data and faster computers has enabled great advances in machine learning and perception in recent 10 years. Based on the idea of DGCC, the intelligence learning mechanism of deep learning is studied and taken as a new artificial

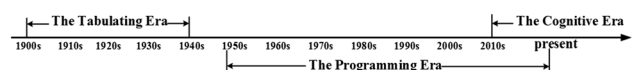


Fig. 1 Three phases in computing's evolution (Kelly III 2015)

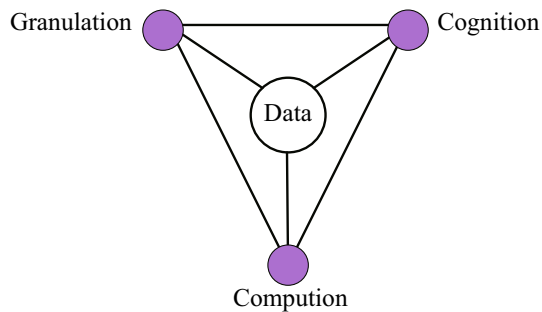


Fig. 2 Triangular structure of DGCC

intelligence mechanism called hierarchical structuralism in this paper. It is also like the idea of hierarchical problem solving in granular computing (Yao 2011).

Data-driven granular cognitive computing provides a granular cognitive computing framework for efficient knowledge discovery from big data. The idea of DGCC has already been applied to solve some real life problems successfully, such as clustering, image processing, time series forecasting, et al (Deng et al. 2016; Yu et al. 2016; Liu et al. 2013; Xu et al. 2016, 2017). In this paper, the key features of DGCC and theoretical issues to be studied for implementing a DGCC are discussed.

2 Cognitive computing: brain/mind inspired computing

Cognitive science is the interdisciplinary scientific study of the mind and cognitive processes. It examines (Thagard 2014)

- what cognition is
- what it does
- how it works

Cognitive science includes research on intelligence and behavior, especially focusing on how information is represented, processed, and transformed within nervous systems (human or other animal) and machines (e.g., computers).

As the cognitive science hexagon shown in Fig. 3 (Miller 2003), researchers in various fields, such as cognitive psychology, philosophy of mind, cognitive linguistics, cognitive anthropology, and artificial intelligence, have had a lot of studies about cognitive science. It is one of the fundamental sciences for cognitive computing. Cognitive computing aims to develop a coherent, unified, universal mechanism inspired by the mind's capabilities (Modha et al. 2011). It is a new type of mind inspired computing with the goal of more accurate models of how the human brain/mind senses, reasons, and responds to

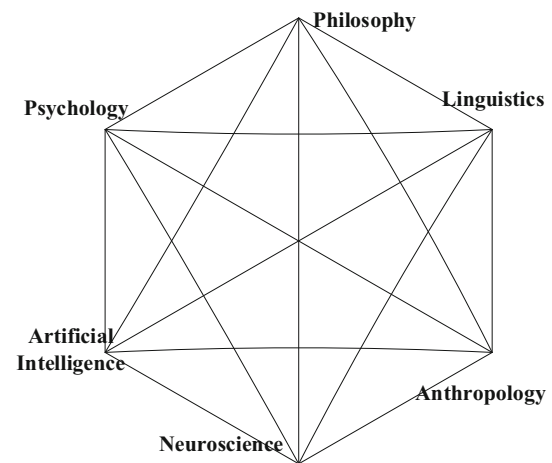


Fig. 3 Cognitive science hexagon (Miller 2003)

stimulus. Cognitive systems should be adaptive, interactive, iterative, stateful, and contextual.

Cognitive computing is based on the scientific disciplines of artificial intelligence and signal processing. Artificial intelligence researchers have developed many intelligent computing models and machine learning models to address complex real-world problems inspired by some specific intelligence observation of brain/mind law, biological law, natural law, and social law. They are close to human's way of reasoning, use non-exact and non-complete knowledge, and could produce control actions in an adaptive way. Fuzzy logic enables a computer to understand natural language and reason in a similar way to human being (Zadeh 1996). Artificial neural networks learn experiential data by operating like the biological/human brain (Rumelhart et al. 1988; Hopfield 1982; Kohonen 1982; Ackley et al. 1985; Broomhead and Lowe 1988). Evolutionary computing is based on the process of natural selection and natural evolution (De Jong 2006). The inspiration of swarm intelligence comes from nature, especially biological systems (Beni and Wang 1993). Artificial immune systems are inspired by theoretical immunology and observed immune functions, principles and models (De Castro and Timmis 2002). Granular computing mimics a way of thinking that relies on the human ability to perceive the real world under various levels of granularity to abstract and consider only those things that serve a specific interest and to switch among different granularities (Yao et al. 2013; Zhang and Zhang 2014). There are also some researchers trying to set up a unified theory of human cognition. They are seeking to implement a unified computational theory of the mind, while not just using different cognitive processes to construct independent cognitive computing systems. It is described by Allen Newell as a single set of mechanisms for all of cognitive behavior (Newell and Simon 1976).

Cognitive computing offers fundamental differences in how intelligent systems are built and interact with humans. Cognitive-based systems, such as IBM Watson, are able to build knowledge and learn, understand natural language, and reason and interact more naturally with human beings than traditional systems (Bellissimo 2015).

In the future, more intelligent cognitive computing systems will be developed inspired by some new cognition laws discovered.

3 Granular computing: human granular thinking and problem solving

Humans often have strong abilities of processing non-numeric information clumps (granules) rather than individual numeric values. Granules are composed of objects that are drawn together by indiscernibility, similarity, and/or functionality among the objects (Zadeh 1997). The original idea of information granulation could be dated back to the nineties of the last century (Zadeh 1997). Information granulation is inherent in human thinking and reasoning processes. It could permeate almost all human endeavors. Granulation—an operation to construct or decompose granules—is one of the key issues of granular computing. The label of Granular Computing (GrC) was suggested by Lin in late 1990s. Granular computing has emerged as one of the fastest growing intelligent computing paradigms in the domain of cognitive intelligence and artificial intelligence (Yao et al. 2013). It is often loosely regarded as an umbrella term to cover theories, methodologies, techniques, and tools that make use of granules in complex problem solving (Zadeh 1997, 2007; Pedrycz 2001, 2006; Lin 2003; Yao 2004, 2005; Bargiela and Pedrycz 2008). Granular computing could be adopted into a structured combination of algorithmic and non-algorithmic information processing that mimics humans by intelligent synthesis of knowledge from information. Bargiela and Pedrycz consider granular computing as a conceptual and algorithmic platform for analyzing and designing human-centric intelligent systems (Bargiela and Pedrycz 2008).

Zadeh considers granular computing as a basis for computing with words which computes with information described in natural language (Zadeh 1996). Fuzzy sets and fuzzy logics could provide a granular representation of uncertain information about a variable of interest.

Skowron uses rough approximations to model syntax, semantics, and operations of information granules (Janowski and Skowron 2007). The two granulation structures, partition with an equivalence relation and cover with a reflexive binary relation, could be generated from rough set points of view. More importantly, multilevel

granulation structures could be induced by hierarchies of the universe and neighborhood systems.

Zhang and Zhang propose a quotient space theory for problem solving inspired by the human thinking ability of perceiving the real world under various levels of granularity to abstract and consider only those things that serve a specific interest and switching among different granularities. A problem space is represented as a triplet of the universe, its structure, and attributes. A problem space with different granularity sizes can be represented by a set of quotient spaces (Zhang and Zhang 2014).

Formal concept analysis could be adopted to automatically derive ontology from a set of objects (Wille 1982). The granular structure of concept lattices in formal concept analysis is useful for knowledge reduction. A granular reduction of a formal context is a minimal attribute set preserving the object granules of the concept lattice obtained from the full attribute set. Based on a reduction of a formal decision context, knowledge in the sense of granular IF–THEN rules hidden in the decision context can be generated (Yao et al. 2013; Chou et al. 2009).

As shown in Fig. 4 (Yao 2016b), Yao views granular computing as a complementary and dependent triangle, which integrates three important perspectives, namely, philosophy of structured thinking, methodology of structured problem solving, and mechanism of structured information processing (Yao 2016b). He proposed a trisecting-and-acting model to explain three way decisions (3WD), a class of effective ways and heuristics commonly used in human problem solving and information processing, in terms of two basic tasks. One task is to divide a universal set into three pair-wise disjoint regions called a trisection or a tri-partition of the universal set. The other task is to act upon objects in one or more regions by developing appropriate strategies (Yao 2016a).

Wang proposes a bidirectional cognitive computing model (BCC) between concept's extension and intension based on a qualitative and quantitative mapping model for expressing and processing of uncertain concepts developed by Li (Liu et al. 2013; Wang and Xu 2012; Wang et al.

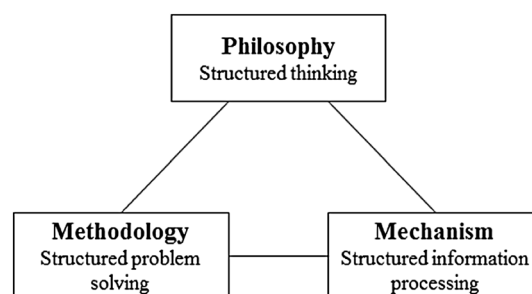


Fig. 4 The granular computing triangle (Yao 2016b)

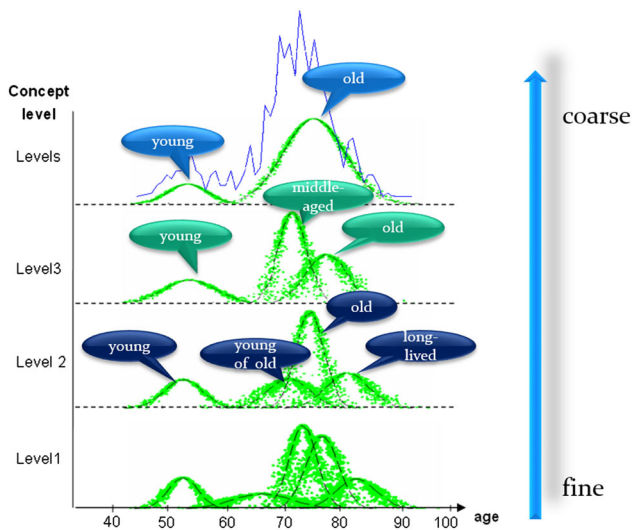


Fig. 5 A concept tree generated at different granularity levels (Liu et al. 2013)

2014; Li et al. 1995). It uses three parameters, namely expected value, entropy and hyper entropy, to describe the intension of a concept, while a set of samples to describe its corresponding extension. Forward transformation algorithm implements the transformation from the intension of a concept to its extension, while backward transformation algorithm vice versa. A multiple granularity concept generation model was developed for generating hierarchical concept trees as shown in Fig. 5 (Liu et al. 2013).

Xu and Wang study the problem of adaptive multiple granular hierarchical clustering. Rodriguez and Laio develop an effective clustering approach based on fast searching and finding of density peaks (Rodriguez and Laio 2014). Xu and Wang further improve it and develop an adaptive hierarchical clustering approach to generate a hierarchical tree as shown in Fig. 6 (Xu et al. 2016, 2017).

The original data could be clustered into 2, 4 or 5 clusters in a hierarchical structure.

There are also some similar granular computing models developed to simulate and implement human granular thinking and problem solving, such as, interactive granular computing (Skowron et al. 2016; Wilke and Portmann 2016), granular neural network (Song and Wang 2016), granular clustering (Peters and Weber 2016; Yu et al. 2016; Xu et al. 2016), etc.

4 Data-driven granular cognitive computing: integration of mind-driven and data-driven computing

In classical intelligent information systems, original data are collected from environment at first usually, useful information is extracted through analyzing the input data then, and it is used to solve some problem at last. There is a common characteristic of traditional machine learning, data mining and knowledge discovery models. That is, knowledge is always transformed (extracted) from data. It is a unidirectional transformation from finer granularity to coarser granularity as shown in Fig. 7.

There is a human cognition law called “global precedence” developed by Chen, a cognitive scientist, in 1980s (Han and Chen 1996; Chen et al. 2003; Chen 1982). As shown in Fig. 8a (Han and Chen 1996; Navon 1977), there are four large characters (the global level) made out of two small characters (the local level). People always recognize the large characters in the global level at first and then the small characters in the local level. It is easy to draw and recognize a people, as shown in Fig. 8b, through his/her caricature, which has just a few lines, without analyzing detailed pixels. It is also easy for a people to recognize his/

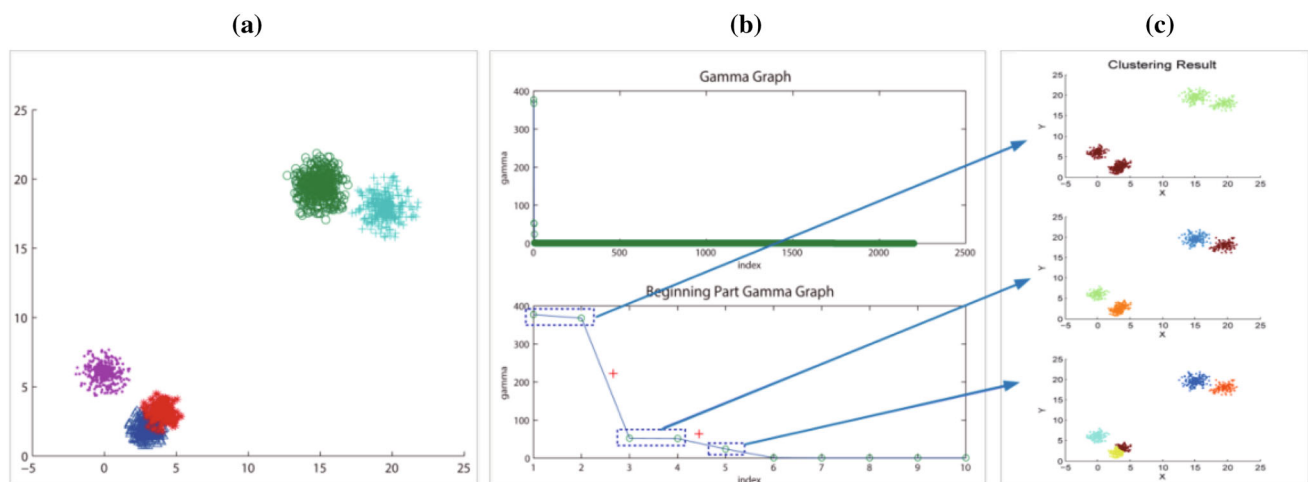


Fig. 6 Multiple granularity clustering based on density peaks (Xu et al. 2016)

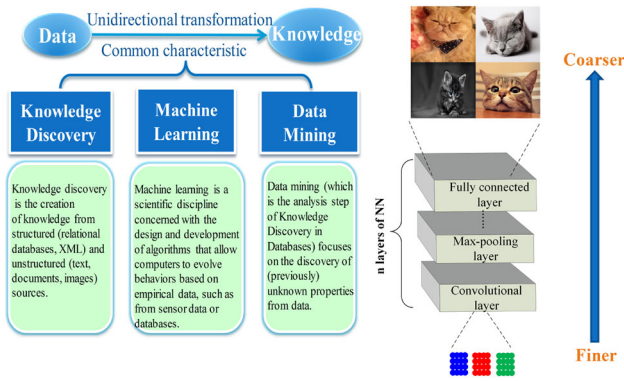
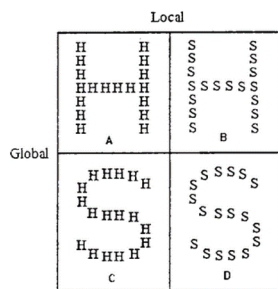
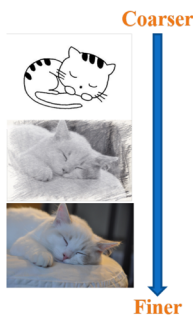


Fig. 7 Unidirectional transformation from finer granularity to coarser granularity



(a) Global precedence (Han and Chen 1996; Navon 1977)



(b) From coarser to finer

Fig. 8 Human cognition: from coarser granularity levels to finer levels

her old friend in a far distance at glance through its outline without detailed information. It shows the cognition law of the information processing in human visual perception. It is a cognition process from coarser granularity to finer granularity.

It is obvious that there is contradiction between the unidirectional transformation mechanism “from finer granularity to coarser granularity” of traditional intelligent information systems with the global precedence law of human cognition. It would be a great challenge to integrate them into a machine learning model. Here, we propose a new model, DGCC, for this purpose. The Triangular structure of DGCC is shown in Fig. 2. In this triangular

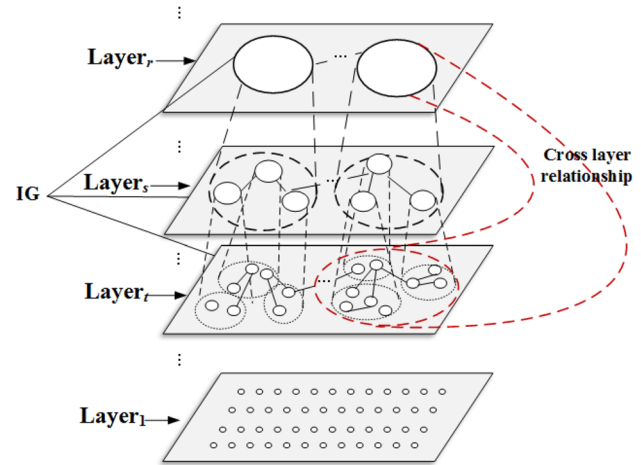


Fig. 9 A general multiple granularity structure for DGCC (Wang et al. 2016)

structure, computation emphasizes the data science which includes all efficient computing models and methods for processing big data; cognition emphasizes the smart understanding of big data and the intelligent interaction between users and information systems; granulation emphasizes the multiple granularity thinking and modeling for dealing with big data. Computation, cognition and granulation are implemented in a data-driven way. Wang developed a general multiple granularity structure for DGCC as shown in Fig. 9 (Wang et al. 2016).

Data-driven granular cognitive computing has the following key features.

- In DGCC, data are considered to be knowledge in the lowest granularity level, and knowledge is considered to be the abstraction of data in different granularity layers.
- There could be relationship both between nodes (concepts) in a same granularity layer, and between nodes (concepts) in different layers.
- Nodes in different granularity layers could take action jointly and simultaneously in a parallel way, while not just sequentially.

There are many theoretical issues to be studied for implementing a DGCC model.

1. Multiple granularity representation of data, information and knowledge.

Traditionally, the relationship of data, information and knowledge is shown in Fig. 9. In this understanding, data are in the bottom layer, information in the middle layers, while knowledge in the high layers. In DGCC, data are considered to be the knowledge represented in the lowest granularity layer. In another way, data could be viewed as the extension of concepts (knowledge in a higher granularity layer), a concept could be viewed

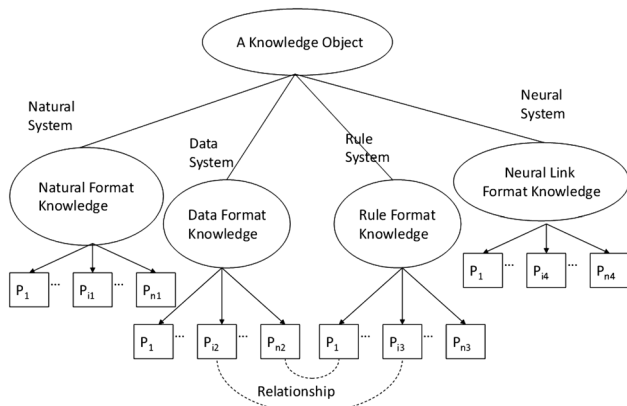


Fig. 10 Knowledge transformation framework of 3DM (Wang and Wang 2009)

as the intension (abstraction) of some data. The idea of considering data as a format for encoding knowledge was introduced in our early work about domain-oriented data-driven data mining (3DM), which is shown in Fig. 10 (Wang and Wang 2009). Data, information and knowledge will be encoded in a hierarchical multiple granularity space together. It would be more complex than any traditional data space, information space or knowledge space. A general multiple granularity structure shown in Fig. 9 needs to be set up for expressing data, information, and knowledge.

2. Integration of the human cognition of “from coarser to finer” and the information processing of “from finer to coarser”.
There have already been some studies about bottom-up and top-down processing in psychology and cognitive science (Lindsay and Norman 1977). Yao studied three top-down approaches, bottom-up approaches, and middle-out approaches (Yao 2016b). In DGCC, two kinds of transformation operators, namely, upward operators and downward operators are needed to simulate and implement such human intelligence mechanisms. An upward operator transforms the data/information/knowledge in a low granularity layer to a high granularity layer, while a downward operator transforms the data/information/knowledge in a high granularity layer to a low granularity layer. It could be viewed as an extension of the backward transformation and forward transformation algorithms of the qualitative and quantitative mapping model (Liu et al. 2013; Wang and Xu 2012; Wang et al. 2014; Li et al. 1995). Downward operators mimic the human cognition of “from coarser to finer”, while upward operators mimic the information processing of “from finer to coarser”.
3. Transformation of the uncertainty of big data in a multiple granularity space.

Generally speaking, concepts (information and knowledge) in a higher granularity layer would be more uncertain than the ones in a lower granularity layer. A concept in a higher granularity layer would be the abstraction of some objects (data or concepts in a lower granularity layer). Exceptionally, some concepts in a lower granularity layer could also be more uncertain than the ones in a higher granularity layer, since they are partial description of some aspects of an object. Thus, the transformation of the uncertainty of big data among a multiple granularity space will be very complex.

4. Multiple granularity joint computing model and problem solving mechanism.
Data, information and knowledge are encoded in a multiple granularity space together. They could be used in problem solving simultaneously in a parallel way. As shown in Fig. 11, decisions in a manufacturing industry group are being made at several different layers simultaneously every day. Decisions in different layers might be either dependent or independent. Mechanisms for joint computing and decision making in a multiple granularity space is required.
5. Dynamical evolution mechanism in a multiple granularity knowledge space.
In real life applications, most systems are dynamical ones. The data, information and knowledge of an intelligent information system would also be dynamical, while not statical. Figure 12 shows the dynamical evolution process of a multiple granularity knowledge space. Some dynamical evolution mechanisms need to be developed to deal with the dynamical data, information and knowledge in a multiple granularity knowledge space.
6. Effective progressive variable granularity computing method.
The information and knowledge in different granularity layers are abstract representations of objects. Usually, coarser answers could be generated in a higher granularity layer with less time cost, while finer

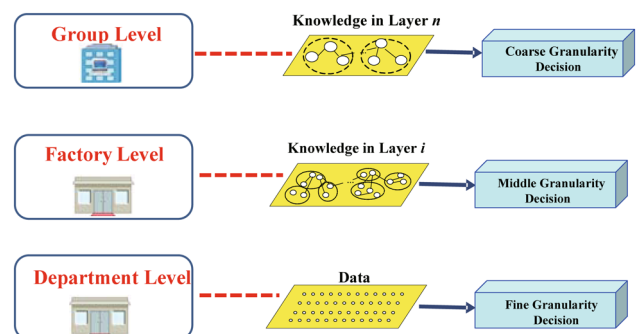


Fig. 11 Joint decision making in a multiple granularity space

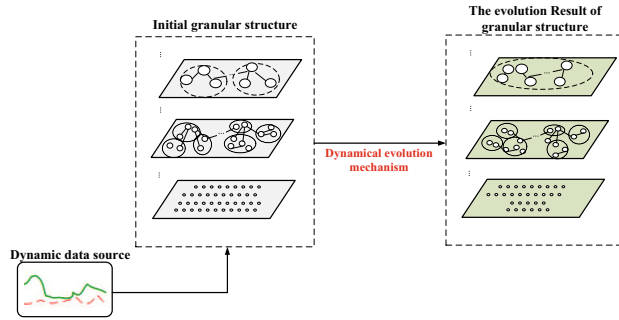


Fig. 12 Dynamical evolution process of a multiple granularity knowledge space

answers in a lower granularity layer with more time cost. Thus, effective progressive variable granularity computing method should be developed. Some kinds of coarser answers are generated in a higher granularity layer at first, and more detailed answers will be available in lower granularity layers later.

7. Calculation goes ahead of perception.

In some real life application systems, not all input information (data) is available simultaneously in the beginning. It would be better to make a draft decision according to the partial inputs available currently, while not to wait for all inputs. In some problem solving tasks, we do not need all inputs. In this case, to take efficient actions, an answer (decision) in a lower granularity layer will be generated based on partial and local inputs at first, and then an improved answer (decision) could be generated in a higher granularity layer after more inputs are available. In some special cases, it is impossible to have all inputs. There are only a few limited inputs available. A decision (answer) could only be generated according to the limited inputs in a lower granularity layer.

8. Distributed multiple granularity machine learning method.

Since data, information and knowledge are encoded in a multiple granularity space together, a parallel and distributed learning process would be possible. It is not needed to learn layer by layer.

9. Multiple granularity mechanism of associative memory with forgetting.

Up to now, the information storage mechanism of computers is a mechanical one. Information (data, knowledge) could be either stored in a memory system or not. It will be unavailable after being removed. However, it is not such a case in human brain. Human brain can forget an object and recall it later in some degree, even in a degree of 100%. There is an association mechanism in human brain. The bidirectional cognitive computing model in (Liu et al. 2013)

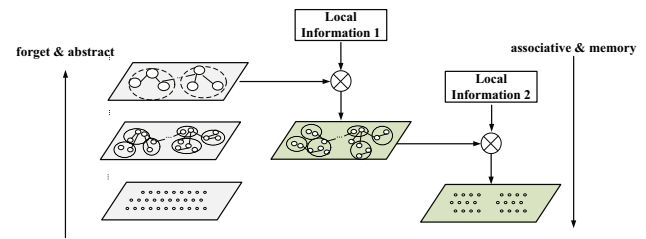


Fig. 13 Associative memory with forgetting

might be used to implement such an association mechanism of human brain in a multiple granularity space. Upward operators could simulate a forgetting process through transforming information in a lower granularity layer to some abstracted information in a higher granularity layer, while downward operators could simulate an associating (recalling) process through transforming information in a higher granularity layer to some detailed information in a lower granularity layer. Figure 13 shows the idea of associative memory with forgetting.

5 Hierarchical structuralism: a new mechanism for artificial intelligence

Deep learning opened a revolution of artificial intelligence. Various deep learning architectures (Bengio 2009; Schmidhuber 2015), like convolutional deep neural networks (LeCun et al. 1989), deep belief networks (Hinton et al. 2006) and recurrent neural networks (Goller and Kuchler 1996) have been applied to a lot of fields such as image recognition, speech recognition, natural language processing and bioinformatics where they have produced state-of-the-art results on various tasks (Goller and Kuchler 1996; Deng et al. 2013; Deng and Yu 2014; Cireşan et al. 2013; Mesnil et al. 2015; Krizhevsky et al. 2012; Chicco et al. 2014). It becomes one of the main streams that push the development of artificial intelligence in the whole world.

Deep Learning architecture built from artificial neural networks (ANN) could date back to the Neocognitron in 1980 (Fukushima 1980). The biggest challenge of artificial neural network study is how to train a network with multiple layers. In 1989, LeCun applied the standard back propagation algorithm to a deep neural network with the purpose of recognizing handwritten ZIP codes (LeCun et al. 1989). In 1995, Hinton trained a network containing six fully connected layers and several hundred hidden units using the wake–sleep algorithm (Hinton et al. 1995). However, the time cost to train these networks was too high, making it impractical for general use.

Most artificial neural network researchers usually used networks with only three layers in real applications before deep learning was developed since it is very difficult and time consuming to train a traditional artificial neural network with multiple layers. The inner structure of a traditional artificial neural network is considered as a black box. The logical structure or knowledge structure of an artificial neural network is unknown even it has been trained to deal with a problem very well. There is no observable, understandable structure or feature in a trained network. The more hidden layers an artificial neural network has, the more difficult and time cost to train it.

In 2006, Hinton and Salakhutdinov effectively pre-trained a many-layered feed forward neural network one layer at a time, treated each layer in turn as an unsupervised restricted Boltzmann machine, and then fine-tuned it using supervised back propagation (Hinton 2007). The linking structure relation between adjacent layers is considered in the training process. It solved the problem of training an artificial neural network with multiple layers (even over 200 layers). It is viewed as the resurgence of artificial neural networks. It has become part of many state-of-the-art systems in various disciplines in recent years, particularly automatic image recognition and speech recognition.

Deep learning could also be taken as a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data using a deep graph with multiple processing layers of linear or non-linear transformations (Goodfellow et al. 2016). It makes better representations and creates models to learn these representations from large-scale unlabeled data. For example, the architecture of a convolutional neural network (CNN) in Fig. 14 also shows the layered abstraction structure from the input of data and the output of cognition.

Deep learning has the following key characteristics (Deng and Yu 2014).

- Uses a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each

successive layer uses the output from the previous layer as input.

- Is based on the (unsupervised) learning of multiple levels of features or representations of the data. Higher level features are derived from lower level features to form a hierarchical representation.
- Is part of the broader machine learning field of learning representations of data.
- Learns multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

In fact, there were some artificial neural network researchers who had also implemented such ideas in their many-layered neural networks in 1990s, before Geoffrey Hinton developed deep learning.

Wang developed a neuro-fuzzy network (FCN) for the bucket motion control with 9 layers in 1995 (Wang 1992; Wang and Kim 1995). As shown in Fig. 15, it is a structured network composed of 3 structured sub networks, namely, pattern recognition network, fuzzy reasoning network and control synthesis network.

Jang developed an adaptive-network-based fuzzy inference system (ANFIS) in 1993 (Jang 1993), which is a neural network with five layers to implement a Takagi–Sugeno–Kang (TSK) fuzzy inference system. As shown in Fig. 16, in an ANFIS, each neuron in the 1st layer calculates a fuzzy membership, each neuron in the 2nd layer multiplies its incoming signals, each neuron in the 3rd layer calculates a normalized firing strength, each neuron in the 4th layer is the multiplication of the output of layer 3 and a parameterized inputs ($p_i + q_i + r_i$), the single node in the last layer computes the summation of all incoming signals. Neuro-fuzzy systems are a typical example of granular neural networks (Song and Wang 2016). Perhaps ANFIS is the best known neuro-fuzzy system (Song and Wang 2016). Many neuro-fuzzy systems have been proposed now (Pedrycz and Aliev 2009; Lee and Lee 1974; Zhang and Kandel 1998; Frayman and Wang 1998; Jang

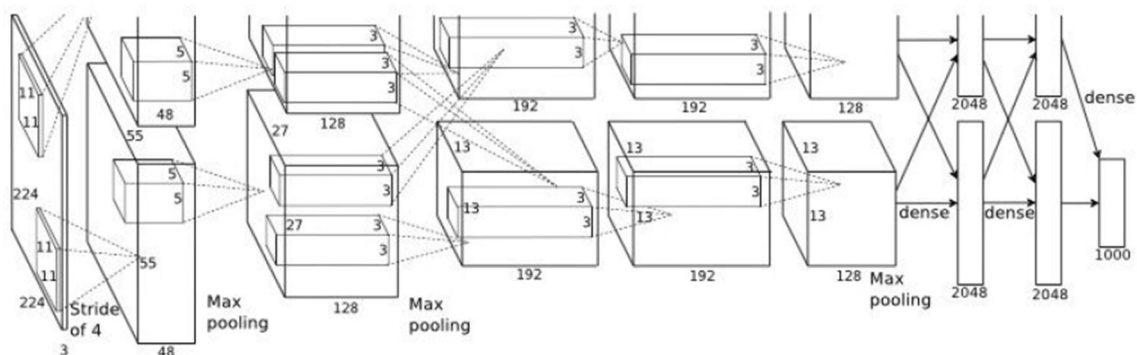


Fig. 14 Architecture of a CNN (Krizhevsky et al. 2012)

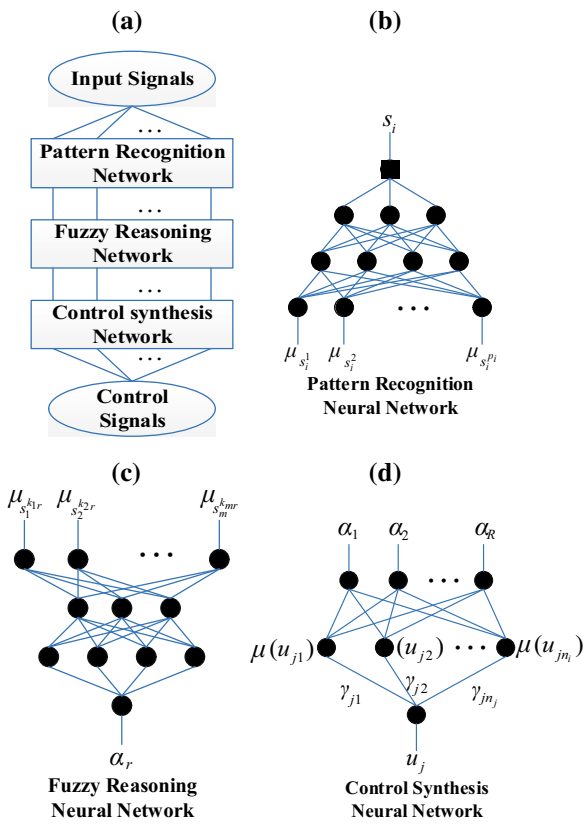


Fig. 15 Structured neuro-fuzzy network (FCN) (Wang 1992; Wang and Kim 1995)

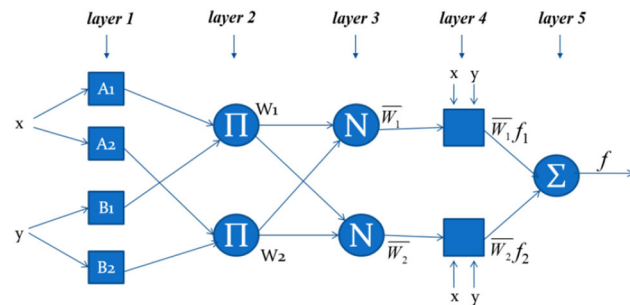
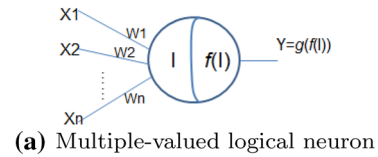


Fig. 16 ANFIS (Jang 1993)

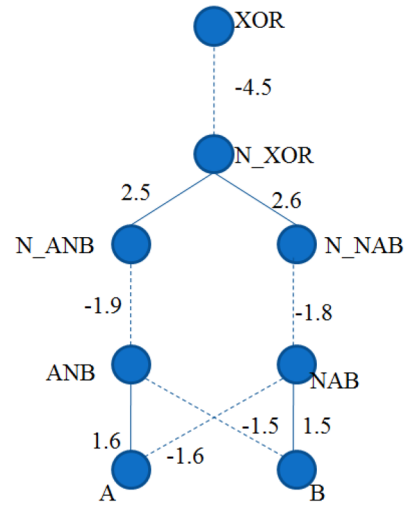
1993; Malekzadeh-A and Akbarzadeh-T 2004; Chen et al. 2011).

Wang developed a triple-valued or multiple-valued logic neural network model (TMLNN) in 1996 (Wang 1996; Wang and Shi 1996, 1998). As shown in Fig. 17, each neuron of this network is a triple-valued or multiple-valued logic neuron. A logic structure is constructed in a TMLNN. A TMLNN with 5 layers could implement any logic function.

It is easy to train an FCN, ANFIS and TMLNN with a low time cost. They have very clear logical structures. Although FCN, ANFIS and TMLNN were developed and applied in some fields in 1990s, about 10 years earlier than



(a) Multiple-valued logical neuron



(b) Multiple-valued Exclusive OR (XOR) Neural Network

Fig. 17 Multiple-valued logic neural network (Wang 1996; Wang and Shi 1996, 1998)

deep learning developed in 2000s, they did not have general use. In 1990s, both the computing power of a computer system and the training data were very limited. It was impossible to use them to solve large scale complex real life problems such as image recognition and speech recognition at that time.

From the above discussion, it could be found that FCN, ANFIS and TMLNN have the same theoretical idea with deep learning. They all exploit the idea of hierarchical explanatory factors where higher layer, more abstract concepts are learned from the lower layer ones. It is a kind of multi-granularity knowledge/information representation structure discussed in Sect. 4. In these multi-level artificial neural network models, the inner structure of a network is no longer a black box. Neurons in each layer have distinct logic meaning. The linking weights between neurons correspond to their logic relationship. Thus, this kind of artificial neural networks could be considered as an implementation way of symbolism systems. They could be considered as some kind of logic reasoning networks of symbolism systems. We can call them hierarchical structuralism. It is also a special case of DGCC. It has the following HD³ characteristics.

- *Hierarchical* The knowledge and information are encoded in a hierarchical system. The inner structure of a hierarchical system is understandable while not a black box. Nodes in different layers of a hierarchical

system correspond to some distinct logic meaning like concepts. It is similar to symbolism in some degree on this point.

- *Distributed* The knowledge and information are encoded in a distributed manner. It is similar to connectionism in some degree on this point.
- *Data-driven and training based* A hierarchical system is set up based on training. The hierarchical structure is not man made. It is trained in a data-driven manner.
- *Dynamical* The inner structure of a hierarchical system could be dynamically adjusted in an adaptive and evolutionary way.

6 Conclusion

It is a natural way to develop intelligent computing models with inspiration of natural/brain/social cognition laws. Inspired by human's granularity thinking, problem solving mechanism and the cognition law of "global precedence", a new powerful cognitive computing model, DGCC, is proposed in this paper. It could integrate two contradict mechanisms, namely, human's cognition mechanism of "global precedence" which is a cognition process of "from coarser to finer" and the "from finer to coarser" machine learning mechanism, in a multiple granularity space. It is a multiple granularity representation of data, information and knowledge. It could transform the uncertainty of big data in a multiple granularity space, implement multiple granularity joint computing and problem solving, simulate the dynamical knowledge evolution mechanism. Both progressive variable granularity computation and calculation going ahead of perception could be realized. Multiple granularity mechanism of associative memory with forgetting could also be possible in DGCC.

After examining the principles of different intelligent cognitive computing models, such as deep learning, logic neural network, fuzzy set, rough set, quotient space theory, cloud model (a qualitative and quantitative mapping model), formal concept analysis, three way decisions and clustering, a hierarchical structuralism is proposed for artificial intelligence based on DGCC, which is a combination of symbolism and connectionism. The HD³ characteristics of the hierarchical structuralism are discussed.

Data-driven granular cognitive computing could provide a granular cognitive computing framework for efficient knowledge discovery from big data. The idea of DGCC has already been applied in some real life problems successfully, such as clustering, image processing, time series forecasting. The important theoretical issues of DGCC discussed in this paper would help researchers to develop big data intelligent computing algorithms and systems.

Acknowledgements This work has been supported by the National Key Research and Development Program of China under Grants 2016QY01W0200 and 2016YFB1000905, the National Natural Science Foundation of China under Grants 61572091 and 61533020.

References

- Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for boltzmann machines. *Cognit Sci* 9(1):147–169
- Bargiela A, Pedrycz W (2008) Toward a theory of granular computing for human-centered information processing. *IEEE Trans Fuzzy Syst* 16(2):320–330
- Bellisimo J (2015) What's the future of cognitive computing? <http://www.forbes.com/sites/ibm/2015/02/23/whats-the-future-of-cognitive-computing-ibm-watson>
- Bengio Y (2009) Learning deep architectures for AI. *Found Trends Mach Learn* 2(1):1–127
- Beni G, Wang J (1993) Swarm intelligence in cellular robotic systems. *Robots and biological systems: towards a new bionics*. Springer, Berlin, pp 703–712
- Brooks RA (1991) Intelligence without representation. *Artif Intell* 47(1–3):139–159
- Broomhead DS, Lowe D (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks. Tech. rep, DTIC Document
- Chen L (1982) Topological structure in visual perception. *Science* 218(4573):699–700
- Chen L, Zhang S, Srinivasan MV (2003) Global perception in small brains: topological pattern recognition in honey bees. *Proc Natl Acad Sci* 100(11):6884–6889
- Chen ZF, Aghakhani S, Man J, Dick S (2011) Ancfis: a neurofuzzy architecture employing complex fuzzy sets. *IEEE Trans Fuzzy Syst* 19(2):305–322
- Chicco D, Sadowski P, Baldi P (2014) Deep autoencoder neural networks for gene ontology annotation predictions. In: *Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics*. ACM, New York, pp 533–540
- Chou GF, Ma JM, Yang HZ (2009) Mathematic model of concept granular computing system. *Sci China Ser F-Inf Sci* 39(12):1239–1247
- Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J (2013) Mitosis detection in breast cancer histology images with deep neural networks. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Berlin, pp 411–418
- Crevier D (1993) *AI: the tumultuous search for artificial intelligence*. Basic Books, New York
- De Castro LN, Timmis J (2002) *Artificial immune systems: a new computational intelligence approach*. Springer Science & Business Media, Berlin
- De Jong K (2006) *Evolutionary computation: a unified approach*. MIT Press, New York
- Deng L, Yu D (2014) Deep learning: methods and applications. *Found Trends Signal Process* 7(3–4):197–387
- Deng L, Hinton GE, Kingsbury B (2013) New types of deep neural network learning for speech recognition and related applications: an overview. In: *2013 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, New York, pp 8599–8603
- Deng WH, Wang GY, Zhang XR, Xu J, Li GD (2016) A multi-granularity combined prediction model based on fuzzy trend forecasting and particle swarm techniques. *Neurocomputing* 173:1671–1682

- Frayman Y, Wang LP (1998) Data mining using dynamically constructed recurrent fuzzy neural networks. Springer, Berlin
- Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36(4):193–202
- Goller C, Kuchler A (1996) Learning task-dependent distributed representations by backpropagation through structure. In: IEEE international conference on neural networks, 1996, vol 1. IEEE, New York, pp 347–352
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, New York
- Han SH, Chen L (1996) The relationship between global properties and local properties-global precedence. *Adv Psychol Sci* 4(1):36–41
- Haykin S (1994) Neural networks: a comprehensive foundation. Macmillan College Publishing Company, Prentice Hall PTR
- Hinton GE (2007) Learning multiple layers of representation. *Trends Cognit Sci* 11(10):428–434
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- Hinton GE, Dayan P, Frey BJ, Neal R (1995) The wake-sleep algorithm for unsupervised neural networks. *Science* 268:1158–1161
- Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 79(8):2554–2558
- Jackson P (1998) Introduction to expert systems, 3rd edn. Addison-Wesley Longman Publishing Co. Inc, Boston
- Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23(3):665–685
- Jankowski A, Skowron A (2007) Toward rough-granular computing. Springer, Berlin
- Kelly JE III (2015) Computing, cognition and the future of knowing. Dr Kelly III John IBM Research. Cognitive Computing IBM Corporation, USA
- Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biol Cybern* 43(1):59–69
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
- LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
- Lee SC, Lee ET (1974) Fuzzy sets and neural networks. *J Cybern* 4(2):83–103
- Leondes CT (2001) Expert systems: the technology of knowledge management and decision making for the 21st century. Academic Press, New York
- Li DY, Meng HJ, Shi XM (1995) Membership clouds and membership cloud generators. *J Comput Res Dev* 32(6):15–20
- Lin TY (2003) Granular computing. In: International workshop on rough sets, fuzzy sets, data mining, and granular-soft computing. Springer, Berlin, pp 16–24
- Lindsay PH, Norman DA (1977) Human information processing: an introduction to psychology. Academic Press, New York
- Liu YC, Li DY, He W, Wang GY (2013) Granular computing based on gaussian cloud transformation. *Fundamenta Informaticae* 127(1–4):385–398
- Malekzadeh AA, Akbarzadeh TM (2004) Complex-value adaptive neuro fuzzy inference system-canfis. In: Presented at the proceedings of World Automation Congress. Seville, Spain
- Mesnil G, Dauphin Y, Yao K, Bengio Y, Deng L, Hakkani-Tur D, He X, Heck L, Tur G, Yu D et al (2015) Using recurrent neural networks for slot filling in spoken language understanding. *IEEE/ACM Tran Audio Speech Lang Process (TASLP)* 23(3):530–539
- Miller GA (2003) The cognitive revolution: a historical perspective. *Trends Cognit Sci* 7(3):141–144
- Modha DS, Ananthanarayanan R, Esser SK, Ndirango A, Sherbondy AJ, Singh R (2011) Cognitive computing. *Commun ACM* 54(8):62–71
- Navon D (1977) Forest before trees: the precedence of global features in visual perception. *Cognit Psychol* 9(3):353–383
- Newell A, Simon HA (1976) Computer science as empirical inquiry: symbols and search. *Commun ACM* 19(3):113–126
- Pedrycz W (2001) Granular computing: an emerging paradigm. Physica-Verlag GmbH, Heidelberg
- Pedrycz W (2006) Granular computing: an overview. Applied soft computing technologies: the challenge of complexity. Springer, Berlin, pp 19–34
- Pedrycz W, Aliev RA (2009) Logic-oriented neural networks for fuzzy neurocomputing. *Neurocomputing* 73(1):10–23
- Peters G, Weber R (2016) Dcc: a framework for dynamic granular clustering. *Granul Comput* 1(1):1–11
- Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. *Science* 344(6191):1492–1496
- Rumelhart DE (1986) McClelland. Parallel distributed processing, explorations in the microstructures of cognition
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by backpropagating errors. *Nature* 323(99):533–536
- Rumelhart DE, Hinton GE, Williams RJ (1988) Learning representations by back-propagating errors. *Cognit Model* 5(3):1
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Simon AH (1996) The sciences of the artificial. MIT Press, New York
- Skinner B (2011) About behaviorism. Knopf Doubleday Publishing Group, Vintage
- Skowron A, Jankowski A, Dutta S (2016) Interactive granular computing. *Granul Comput* 1(2):95–113
- Song ML, Wang YB (2016) A study of granular computing in the agenda of growth of artificial neural networks. *Granul Comput* 1(4):247–257
- Thagard P (2014) Cognitive science. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, fall, 2014th edn. Stanford University, Metaphysics Research Lab, USA
- Wang FY (1992) Knowledge structure in neural nets using fuzzy logic. In: Jamshidi M (ed) Robotics and manufacturing: recent trends in research. Education and applications. ASME Press, New York
- Wang FY, Hm Kim (1995) Implementing adaptive fuzzy logic controllers with neural networks: a design paradigm. *J Intell Fuzzy Syst* 3(2):165–180
- Wang GY (1996) Study of the neural network models and algorithms in an integrated intelligent system. PhD thesis, Xian Jiaotong University
- Wang GY, Shi HB (1996) Three valued logic neural network. In: Proc. of int. conf. on neural information processing, Hong Kong, pp 1112–1115
- Wang GY, Shi HB (1998) TMLNN: triple-valued or multiple-valued logic neural network. *IEEE Trans Neural Netw* 9(6):1099–1117
- Wang GY, Wang Y (2009) 3DM: domain-oriented data-driven data mining. *Fundamenta Informaticae* 90(4):395–426
- Wang GY, Xu CL (2012) Cloud model—a bidirectional cognition model between concept’s extension and intension. In: International conference on advanced machine learning technologies and applications. Springer, Berlin, pp 391–400
- Wang GY, Xu CL, Li DY (2014) Generic normal cloud model. *Inf Sci* 280:1–15

- Wang GY, Yang J, Xu J (2016) Granular computing: from granularity optimization to multi-granularity joint problem solving. *Granul Comput* doi:10.1007/s41066-016-0032-3
- Wilke G, Portmann E (2016) Granular computing as a basis of human–data interaction: a cognitive cities use case. *Granul Comput* 1(3):181–197
- Wille R (1982) Restructuring lattice theory: an approach based on hierarchies of concepts. *Ordered sets*. Springer, Berlin, pp 445–470
- Xu J, Wang GY, Deng WH (2016) Denpehc: density peak based efficient hierarchical clustering. *Inf Sci* 373:200–218
- Xu J, Wang GY, Li TR, Deng WH, Gou GL (2017) Fat node leading tree for data stream clustering with density peaks. *Knowl Based Syst* 120:99–117
- Yao JT, Vasilakos AV, Pedrycz W (2013) Granular computing: perspectives and challenges. *IEEE Trans Cybern* 43(6):1977–1989
- Yao YY (2004) A partition model of granular computing. *Transactions on rough sets I*. Springer, Berlin, pp 232–253
- Yao YY (2005) Perspectives of granular computing. In: 2005 IEEE international conference on granular computing, vol 1. IEEE, New York, pp 85–90
- Yao YY (2011) Artificial intelligence perspectives on granular computing. *Granular computing and intelligent systems*. Springer, Berlin, pp 17–34
- Yao YY (2016a) Three-way decisions and cognitive computing. *Cognit Comput* 8(4):543–554
- Yao YY (2016b) A triarchic theory of granular computing. *Granul Comput* 1(2):145–157
- Yu H, Zhang C, Wang GY (2016) A tree-based incremental overlapping clustering method using the three-way decision theory. *Knowl Based Syst* 91:189–203
- Zadeh LA (1996) Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A. Zadeh. World Scientific Publishing Co. Inc, River Edge
- Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 90(2):111–127
- Zadeh LA (2007) Granular computing—computing with uncertain, imprecise and partially true data. In: Proc 5th int sym spat data qual (ISSDQ 2007)
- Zhang L, Zhang B (2014) Quotient space based problem solving: a theoretical foundation of granular computing. Elsevier Science, Amsterdam
- Zhang YQ, Kandel A (1998) Compensatory genetic fuzzy neural networks and their applications. World Scientific Publishing Co., Inc, Singapore