IJARAI

International Journal of Advanced Research in Artificial Intelligence

IJARAI

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE

Editorial Preface

From the Desk of Managing Editor...

"The question of whether computers can think is like the question of whether submarines can swim."
— Edsger W. Dijkstra, the quote explains the power of Artificial Intelligence in computers with the changing landscape. The renaissance stimulated by the field of Artificial Intelligence is generating multiple formats and channels of creativity and innovation.

This journal is a special track on Artificial Intelligence by The Science and Information Organization and aims to be a leading forum for engineers, researchers and practitioners throughout the world.

The journal reports results achieved; proposals for new ways of looking at AI problems and include demonstrations of effectiveness. Papers describing existing technologies or algorithms integrating multiple systems are welcomed. IJARAI also invites papers on real life applications, which should describe the current scenarios, proposed solution, emphasize its novelty, and present an in-depth evaluation of the AI techniques being exploited. IJARAI focusses on quality and relevance in its publications.

In addition, IJARAI recognizes the importance of international influences on Artificial Intelligence and seeks international input in all aspects of the journal, including content, authorship of papers, readership, paper reviewers, and Editorial Board membership.

The success of authors and the journal is interdependent. While the Journal is in its initial phase, it is not only the Editor whose work is crucial to producing the journal. The editorial board members , the peer reviewers, scholars around the world who assess submissions, students, and institutions who generously give their expertise in factors small and large— their constant encouragement has helped a lot in the progress of the journal and shall help in future to earn credibility amongst all the reader members.

I add a personal thanks to the whole team that has catalysed so much, and I wish everyone who has been connected with the Journal the very best for the future.

Thank you for Sharing Wisdom!

Editorial Board

Reviewer Board Members

CONTENTS

From the Perspective of Artificial Intelligence: A New Approach to the Nature of Consciousness

Riccardo Manzotti

Department of Linguistics and Philosophy
Massachusetts Institute of Technology
MA, United States

Sabina Jeschke

Institute Cluster IMA/ZLW & IfU
RWTH Aachen University
Aachen, Germany

Abstract—Consciousness is not only a philosophical but also a technological issue, since a conscious agent has evolutionary advantages. Thus, to replicate a biological level of intelligence in a machine, concepts of machine consciousness have to be considered. The widespread internalistic assumption that humans do not experience the world as it is, but through an internal '3D virtual reality model', hinders this construction.

To overcome this obstacle for machine consciousness a new theoretical approach to consciousness is sketched between internalism and externalism to address the gap between experience and physical world. The 'internal interpreter concept' is replaced by a 'key-lock approach'. Here, consciousness is not an image of the external world but the world itself.

A possible technological design for a conscious machine is drafted taking advantage of an architecture exploiting self-development of new goals, intrinsic motivation, and situated cognition. The proposed cognitive architecture does not pretend to be conclusive or experimentally satisfying but rather forms the theoretical the first step to a full architecture model on which the authors currently work on, which will enable conscious agents e.g. for robotics or software applications.

Keywords—consciousness; machine consciousness; multi agent system; genetic algorithms; externalism

I. INTRODUCTION

Even if consciousness is not exactly a 'well-defined' term and generations of philosophers and other scientists have discussed its complex features at length, there is a certain common understanding about its central meaning: Consciousness describes the unique capability of having experiences in terms of perceptions, thoughts, feelings and awareness.1 Obviously, consciousness requires the awareness of the external world. What is still fairly mysterious is the nature of this experience. Although this capability is still very poorly understood and indeed is considered a sort of challenge for the standard picture of the world, it is a plain fact that the conscious human being is one of the outcomes of natural selection. Likewise, it seems undeniable that human beings cope with the most unexpected events by means of conscious

---

1 Consciousness is an 'umbrella term' encompassing a variety of distinct meanings. For this purpose it is important to differentiate between consciousness and self-consciousness. In this paper, to be conscious it is only necessary to be aware of the external world, whereas self-consciousness is an acute sense of self-awareness. In Chapter IV.A. this topic is discussed in more detail.

reflection. Finally, they are extremely sensitive to anything remotely resembling the capability of feeling in other agents. In sum, consciousness appears to be a not negotiable aspect of a highly developed autonomous agents and it cannot be underestimated that the practical advantages may result from its replication within an artificial being [1] [2] [3] [4] [5]. Here, the problem of the physical underpinnings of consciousness rather than the problem of the self is addressed and thus the nature of what it is like to have a certain experience [6] [7] [8] rather than the problem of how the different cognitive processes combine together to form a self. This paper considers how experience may be the result of a physical system interaction with the world – experience rather than the self is the goal of this proposal.

During the recent decades, one got familiar with the conception that one never gets acquainted to the world as it is, but only to a '3D virtual reality model' of the outside world that one's brain switches on as soon as one wakes up. This internal model is taken to be the inner world of consciousness – how the world appears to humans and not what the world really is. To give an example, colour in the external world may be defined, albeit with some simplification, by two physical parameters: wavelength and intensity. For a human being however, light is not just the detection of a certain light frequency on the retina, but a certain experience when detecting that light frequency (say, perceiving red). In his excellent textbook on vision, Stephen Palmer claims: 'Color is a psychological property of our visual experiences when we look at objects and lights, not a physical property of those objects or lights' [9]. Nevertheless, this psychological property is without comprehensive explanation so far. It does not fit to any obvious physical property.

In a nutshell, the current main line of explanation goes as follows:

- An external event

- goes through some kind of internal interpreter (in the human brain)

- and internally produces a certain result (within the human).

This current interpretation can be allocated to the so-called 'internalistic models' – namely those models that take the mind to be a property of what takes place inside the neural system

[10] [11] [12][2]. This view has its strength: putting some kind of 'interpretation layer' between the individual and the outside world allows explaining why humans – all of them build alike – tend to be rather different in their behaviours, reactions and 'feelings'. This argument can be extended to non-human species as well: if the same physical reality 'shows itself' differently to diverse entities, a tentative explanation for the heterogeneous behaviours of these unequal species living in the same environment may be put forward. It also allows an explanation of why humans seem to be capable of consciousness in the absence of obvious external stimuli – in the case of dreams, hallucinations, and afterimages.

However, there are still many open ends, some riddles concerning the conception of consciousness that cannot be answered with the abovementioned picture:

- First, to believe that 'the internal interpreter', and 'neurons will do it "somehow" ', is interesting but doubtful. In the last couple of decades, scientists from all areas have invested a lot of energy in the quest for a neural mechanism capable of producing our everyday conscious experience. Up to now, there is no known law of nature predicting that neural activity should result in one's experience.

- Second, in the current model, consciousness neither fits the physical world nor its properties. To carry it to the extremes, that means that one constantly ignores the 'real world' by overwriting it with some internal 'fantasies'. Of course, this could be the case – but it sounds at least pretty counterintuitive: why should nature take this kind of detour?

- Third: The discrepancy between our immediate experience and the 'world' is more than just 'somewhat regrettable'. If everything one experiences – from pain to colour, from pictures to music – is nothing more than a product of human neurons, then a logical problem occurs: Why should it be easier for neurons to transmute neural firings into music – than for a cello to shape airwaves into music? If the physical world is devoid of qualitative features, why should the brain – which is part of the physical world – be any better in this respect? Why should the brain create meaningful things, but a cello does not? Or, to use an even catchier picture: 'If colours cannot pop out of strawberries, how can they pop out of neurons?'

As anticipated at the beginning of this section, consciousness is not only a scientific conundrum but a practical goal, too. From the Artificial Intelligence community (and the authors admit that they belong to that community), another thought comes up: Whatever consciousness is in detail, it seems to form an important part at least of a human-like intelligence [13] [14] [15] [16] [17] [18] [19]. Therefore, to build artificial systems with certain intelligence, it might be necessary to give them some kind to consciousness too – even if this artificial consciousness might differ very much from the human one, or from other mammals or biological systems. Now, however consciousness might work in biological systems, one may envisage implementing consciousness in totally different ways as part of forthcoming technological systems. Therefore, alternative models to explain consciousness are of the utmost interest, either to explain the 'true nature of human consciousness', or, to allow different approaches for building an artificial/technological agent.

This paper proposes a new hypothesis concerning the nature of conscious experience, to overcome a conceptual war between externalism and internalism (Chapter II). In Chapter III the consequences of this new perspective are discussed. Chapter IV relates the change of perspective to the field of artificial intelligence. Finally, Chapter V summarizes the paper and gives an outlook on the next research steps to be undertaken.

II.   TOWARDS A NEW CONCEPTUALIZATION OF CONSCIOUSNESS

A. *The approach*

This chapter's goal is to flesh out a new concept of consciousness that is directed to overcome the gap between 'experience' and 'physical world'. To reach that goal, one has to undergo several changes in the standard mindset. Below the hardest nuts to crack are mentioned:

1) To bridge the gap between experience and physical world, the external surrounding environment in which a brain and a body are situated have to be more prominent in our concept of consciousness (since the other way around seems even more radical).

2) If the 'externality' becomes more important, the next domino falling is the giving-up of the underlying 'full-bodied' human-centered view of being necessarily located fully and totally inside its body. This assumption is a subliminal driver of the current theories, but it is not based on any empirical evidence – it is something that may be true or false.

3) Also, if 'experience' and 'external world' come closer together, the need for some kind of internal interpreter is dramatically reduced. The transformation of the outside world into an internal representation or a virtual model is getting more and more obsolete.

4) The closer 'experience' and 'external world' get, the less their difference can be. This is not at all a trivial statement. On the contrary, this leads to the most difficult point to grasp: namely, that what people call consciousness 'is' the world people live in. It is not how the world appears, but what the world is.

In Chapter III the consequences of this new model will be discussed. First, however, the authors would like to give the reader the chance to understand HOW such an approach could work in practice with a construction sketch The key idea is that the body and the world are just two pieces of the same physical system and that what the authors call the mind is a physical process that requires both pieces to take place. Body and world

---

[2] Of course, these models do not rule out the importance of the external environment to allow the development of internal structures. Indeed, they consider it as necessary for a healthy brain development to continuously interact with the environment. However, once the required neural connections are in place, the mind is taken to be an internal phenomenon. Dreams and hallucinations are constantly quoted as obvious cases.

are interlocked gear wheels and the consciousness turns them. A schematic description of how the coupling between body and world works is sketched in Figures 1, 2 and 3.
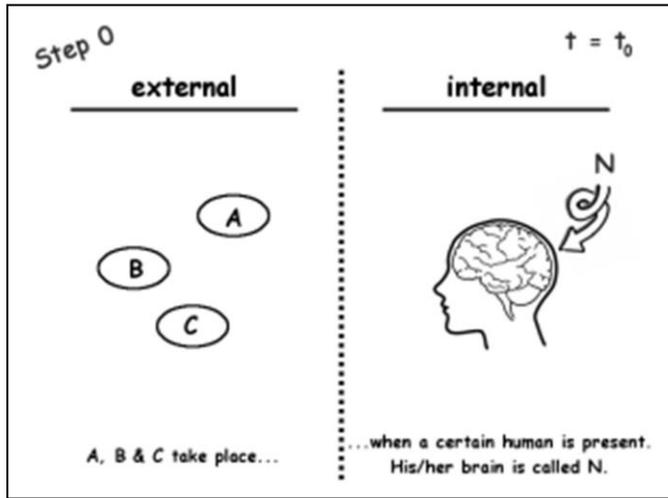
Fig. 1. Step 0 - before perception, there is no external object, as one perceives it. There are smaller and scattered physical phenomena, which are not the target of any normal experience.
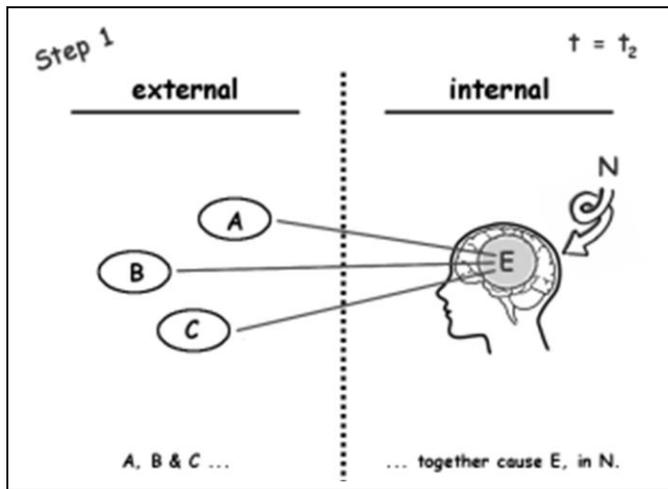
Fig. 2. Step 1 - because of the presence of a certain neural structure inside a body with the proper sensor apparatus, the scattered external phenomena produces a joint effect inside one's brain.
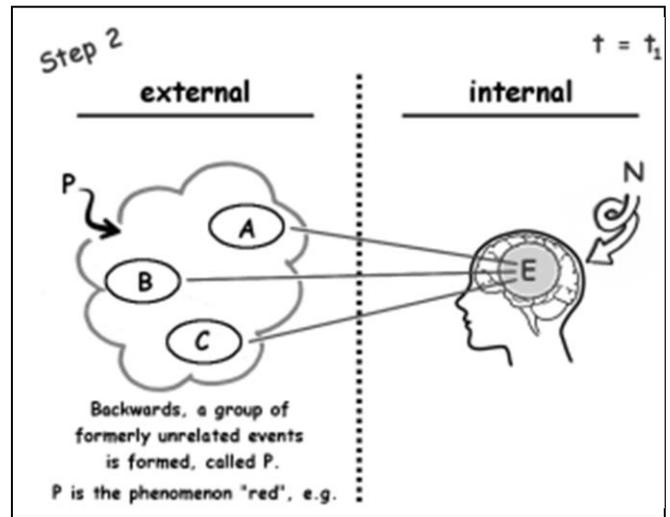
Fig. 3. Step 2 - events that are responsible for the occurrence of a joint effect are a joint cause – they become a whole. One's experience is the process P which takes place thanks to both: to one's brain and by A, B, and C.

Below the different steps are explained in detail:

- In the outside[3] world, there are physical scattered events. Let us call them A, B, C. They are located in time and space. A, B, and C do not have anything in common – each take place on their own.

  o For instance, A may be a certain light ray with a certain frequency emitted at a certain time in a certain location.

- In a nearby body, a healthy brain (call it N), through the causal connection provided by sensorial paths, is affected by these three events A, B, and C. This is not an obvious step. What now happens is that the neural structure – thanks to various neural learning mechanisms – allows the fusion of A, B, and C to create the causal circumstances[4] that would allow the fusion itself to act as a cause for further interaction. A whole is born where P is called the fusion of A, B, and C.

- E is the joint effect of P and takes place inside the brain N.

---

[3] Outside and inside are used to refer to physical events inside and outside one's body.

[4] In causal terms, one may distinguish between 'the cause and that without which the cause would not be a cause' ([20], p. 119). The former may be taken to be an event actually occurring while the latter may be just a state of affairs. The latter may be formalized in terms of conditions G such that $P \wedge G \rightarrow E$ (which may unfolded in three conditionals $P \wedge \neg G \rightarrow \neg E$, $\neg C \wedge G \rightarrow \neg E$, and $\neg C \wedge \neg G \rightarrow \neg E$.

o As a simple analogy consider the 'distributed key' of an atomic rocket launch system: to launch it, two different 'keys' have to be turned at the same time (in movies, usually one of them always refuses to work!). However, considering the matter more carefully, there are not two keys, there is just one key in two pieces, and to be even more precise, the two pieces are not even keys since, when alone they do not unlock anything.

- Because the fusion P of A, B, and C causes E, these three originally independent events 'become' a whole in any practical and sense – things are defined in causal terms and not in ideal terms. P has not existed until E occurred. E is inside the body, while P remains outside of it.

  o Going back to the Rocket launcher analogy, before being put into the 'lock', both 'key pieces' are as unrelated as A, B, C. After being inserted, they, together with the lock, from a group and form a 'lock-and-key principle' which is a construction partially internal and partially external. The separation between them becomes purely conceptual – physically and causally they are a whole.

- Finally, the hardest bullet to swallow, namely the temporal order. No, the authors have NOT confused the indices: E takes place at a time t_2, while P occurred at a time t_1, with $t_1 < t_2$. Why that – and does that mean that E is changing its past? In some sense yes. At least it is changing the causal role of the past.

The last point merits some further consideration. It is not the intention to invoke some sort of retro-causation that moves backward in time. Rather, the point shows something that should not come as a surprise: Physical phenomena are extended in time. This means that they get to completion within time. Therefore, when something begins to unfold, its nature is not wholly defined until it reaches some natural ending in causal terms. Nothing goes backward in time – the past is, of course, past.

However, what the past was may well be defined by what happens in the present. Using the above time indexes, there is no need to suppose that anything is going from t_2 to t_1. However, there is no harm either in considering that what the world at time t_1 was (that is, A, B, and C) changed after t_2 (that is, the fusion P takes place). If one considers a physical phenomenon as something extended in time, then it may well be the case that what happens along such an extension redefines the structure of the phenomenon.

To understand the temporal structure of the proposed causal sequence one has to take a closer look. At time $t_1 < t_2$, the actual cause P has not yet happened. Until t_2, P has not yet produced any effect. Thus, from a physical standpoint, P has not yet existed. One may put the situation in these terms – until t_2 P's existence is not causally any different from P's absence. Then E takes place at t_2. Things have changed. P is now the actual cause of an event E that might have not happened. Yet E happened and thus P was its actual cause. Has this temporal sequence any effect to the temporal order of perceived events

[21] [22]? Not necessarily, as a matter of fact, the subjective temporal order of events depends on how subsequent cognitive processes exploit them. Furthermore, here the crucial issue is the internal physical and causal structure of a perceptual act rather than how the temporal order of different perceptual acts is experienced. Just to dispel any possible misunderstanding, neither the time t_2 is not the subjective time, nor the time t_1 is the objective time. Both t_1 and t_2 are physical times and they refer to when certain causal processes take place. The interval t_1-t_2 has no mandatory impact as to which order P is in temporal relation with other perceived events.

As a further example, let's get back to the case that was mentioned at the onset – namely conscious perception[5] of colour. If one applies the approach just sketched, a colour is a collection of scattered and otherwise separate physical properties until they produce a joint effect in one's brain (E)[6]. When they do so, the scattered wavelengths can be considered as a set of external phenomena (A, B, C, …), namely the colour red. What is hard to grasp is how to step from some scattered wavelengths to the impression of a colour. The answer is that it happens in exactly the same way as one comes from a bundle of 'whatsoever-pieces' to a key – the components themselves do not constitute a whole (in respect to colour) but the sum of them does. The pieces merge into one key IF AND ONLY IF they have the opportunity to do so in causal and actual terms, meaning that there is a 'suitable lock' around (certain interactions with the eye-brain system). Then the whole may take place – without that lock, nothing may happen. By doing so, the scattered events make the colour red happen. In this account, the colour red is the causal fusion of the set of incoming wavelengths. It is neither an internal impression nor a mental ink. Red is an external whole whose occurrence is made possible by causal coupling with the neural event (the joint effect E). A colour-blind person would not have the 'suitable lock' – and therefore would not be conscious of the phenomenon that normally sighted subjects call colour. In physical terms, if there was only a colour-blind person in a certain environment, the combination of physical phenomena getting to an end in a normally-sighted subject would never be able to produce a causal joint effect.

To recap, a causal notion of fusion may thus be put forward – any group of events X_i is fused if and only if there are the causal conditions for a further event E to take place. The idea is that a fusion takes place only if it is the actual cause of some event. Thus, a fusion is not an abstract entity, but a physical occurrence with its own causal efficacy. For any group of events, there is a fusion, whenever they produce together an effect. The structure of neural networks embedded into one's nervous and sensor systems are ideal in this respect – they are the causal circumstances that allow complex events in the environment to be the actual causes of some bodily events, hereafter integrated into the agent's behaviour. Thus, the fusion

---

[5] As the distinguished neuroscientist Semir Zeki once said, there is no such a thing as unconscious perception. However, in this case the authors prefer to be redundant rather than misunderstood.

[6] These properties may be quite an inhomogeneous set of actual physical properties such as the reflected colour spectrum, percentage of certain components, contrastive ratios among different areas, and so forth. For the sake of the example, just a set of wavelengths is considered.

takes place because of some neural event, which is the effect of external groups of events. What is fused, though, is not inside the body, but it is the group of external events. The actual cause, too, remains outside of the body.

Coming back to the steps 1-4 noted in Chapter II the main changes with respect to the standard view are recapped:

- The gap between experience and world … is gone since in the upper model, the external events and the internal perceptions become one.

- The human-centred view … is gone, since in this model, experience is driven partly internal und partly external to a physical body and it is constituted by physical events. The experience is internal to the physical system that underpins it, and it is external to one's body. The body is, of course, nothing but as a subset of a larger physical superset of processes taking place in time.

- Here, the word external does not mean being "projected" but just being physical outside of one's neural structure. Such a notion very strongly suggests that the physical underpinnings of mental states are made of physical events taking place outside of one's body – a standpoint sometimes dubbed phenomenal externalism or externalism about phenomenal content. The goal is to single out a physical event identical with one's experience without having to resort to any mentalistic notion such as content, character, interpretation, projection, reference, and so forth.

- The internal interpreter … is gone or at least not necessary any longer. It was enough to relocate our insights as to what the physical underpinnings of one's mind are. The consciousness of 'seeing red' (instead of seeing several scattered wavelengths) is the result of the fitting between key-parts and lock. Red is not a meaning associated to some internal representation – red is a physical phenomenon in one's physical environment.

- Consciousness 'is' the world people live in: to make an example, to see red is to be united with an external collection of physical phenomena, since experience takes places as a temporally and causally extended phenomenon that includes internal and external components.

In comparison to the traditional view previously mentioned, a new approach is fleshed out (key differences in italics):

- External events & a neural event
  - *form a key-lock-system,*
  - which is, therefore, *partly internal - partly external*.
- The external events produce a certain neural activity,
  - an event inside the human body
- and, as a result,
  - *an external actual cause has occurred.*

B. The consequences overall

For a moment, before raising the inevitable objections, consider this view as a tentative scientific hypothesis as to the physical nature of consciousness – a scientific hypothesis insofar as it puts forward a falsifiable hypothesis as to what the mind is. If this hypothesis had any merit, a few conceptual advantages are immediately obvious:

- First, **the hard problem of consciousness**: The hard problem of consciousness, introduced by David Chalmers [6], addresses the problem of explaining how and why one has qualia and phenomenal experiences such as pain, colours, taste etc. (incl. 'Why does awareness of sensory information exist at all?' 'And why is there a subjective component to experience?'). The presented approach sweeps away the premises on which the hard problem of consciousness is based on (and thus the hard problem itself). In short, the hard problem is based on the dustbin model of the conscious mind [23] [24] in which a set of features is relocated that have been eschewed from the physical world. The idea is that – according to Chalmers – the hard problem is not an unavoidable chasm in the structure of nature, but a false issue created by assuming wrong premises. The approach presented here addresses such premises and indeed suggests a different picture. The mind and the world would no longer be two incommensurable and indeed autonomous domains, but the same one under two different perspectives.

- Second, the **mind-body-problem**: overt and covert dualism would finally be overturned. Dualism is not just the straw man often depicted as the traditional substance dualism contrasting matter and soul or body and mind. There are also forms of dualism that suggest a juxtaposition between cognition and the brain [25] [26], sometimes dubbed Cartesian Materialism [27] [28]. There is no longer the need to differentiate the way in which things look to subjects and the way in which things are. There is just a flow of physical phenomena causally interconnected.

- Third, **exclusiveness**: being conscious of something is a 'private' event – but in contrast to the traditional interpretation, the privacy is no longer created by an internal individual interpreter. It is no longer an exclusive and unbridgeable metaphysical privacy. Rather, it is the kind of privacy that prevents two individuals from eating the same piece of cake. It is a notion akin to that considered by the philosopher Mark Johnston who considers privateness as the impossible to receive the same anti-flu shot [29]. The exclusiveness follows from the fact that the pieces fuse into one key only if there is a 'suitable lock' – the suitable brain of a conscious agent - around. The causal interaction between internality and external world links the observed object and its observer. Of course, in presence of two similar groups of events, two similar brains let similar fusions occur.

- Fourth, **location of consciousness**: it is possible to physically locate the (conscious) mind into the physical

world. The location is not some inside neural activity though. However, it is possible to pinpoint a certain physical process and consider whether such a process is identical to one's own experience of, say, a red patch. It is thus possible to resurrect the theory of identity in terms of broader physical processes and not just in terms of neural processes. The fact that consciousness takes place partially outside the body is not in contrast with the impression one may have to be located inside the own body. Nothing in our experience points to where our experience takes place, only to what our experience is. If someone cuts a finger, he or she does not feel a pain inside the brain, but rather a pain in the finger. By the same token, it is not necessary that the process has to be located within our body.

- Fifth, *the misperception issue*[7] – namely, the fact that apparently one may experience things that are not physically present, as it happens in the case of hallucinations or dreams e.g. – has to be dealt with differently. They are no longer the result of a somewhat 'hyper-creative' internal interpreter, but of an unusual connection with real features in one's environment since this component has been removed from the picture. First, it is important to realize that our dreams are just 'boring' recombinations of the basic components of our past, albeit reshuffled in possibly original ways, they are chimeric but not innovative [30] [31] [32] [33]. Second, it is important to realize that ALL perceptions require a temporal lag between the object and the neural activity, due to the velocity of information transportation. Combining these two insights leads to a possible and fairly simple explanation approach, namely, that dreams and hallucinations may just happen to be cases of very long and reshuffled perception of one's world. So, tentatively, this approach suggests that the stuff dreams (and consciousness) are made of is the same stuff the world is made of.

- Sixth, **tabula rasa**. According to this view there is no mental content distinct from a physical event (that may be part of one's body, of course). Thus, one may experience a red apple or an itch in the elbow, but one cannot experience a pure mental content that one may experience. This is a very physical view that rules out any immaterial or purely mental content. By the same

---

[7] Whenever it was necessary to point to the autonomy of the mental with respect to the physical domain, the issue of misperception has been the battering ram of both philosophers and scientists. Dream and hallucinations appear as formidable evidence in favour of an inner world. However, this approach promises to locate in the physical surrounding a physical cause for ANY experience. All cases of conscious experience ought to be revisited as cases of (admittedly unusual) perception. The approach presented here honestly stands or falls on whether it will succeed to show that – perhaps surprisingly – whenever there is consciousness there is a physical phenomenon, which is the content of one's experience. The authors cannot do justice here to the problem of misperception by and large. However, one can flesh out a template of the strategy – namely to address each purported case of misperception and to revise it in terms of actual perception. (One of the authors is actually working on such an account for most cases of misperception, from hallucination to illusions, from aftereffects to direct brain stimulations.)

token, at the very beginning, organisms cannot have any experience since, by definition, they have not yet been in contact with any physical phenomenon. This does not prevent, of course, that either new born infants or foetuses may have consciousness as long as 1) they have a working neural system and 2) they perceive external events through parts of their bodies (or their mothers' bodies). However, the approach presented here rules out any innate or purely mental content of experience.

III. CONSCIOUSNESS OPENING NEW PERSPECTIVES FOR ARTIFICIAL INTELLIGENCE

Can the outlined approach help in shaping and devising an architecture capable of consciousness? The authors believe that it can, because it suggests a causal structure of consciousness and thus something that may help in singling out relevant architectures in an artificial agent. For instance, the approach suggests that being conscious is not a matter of either having the right internal code [34] [35], or having a central global dashboard [36] [37], or processing information in a certain way [38]. The advantage of this proposal is that it allows for rather precise indications as to why the causal coupling between the environment and the agent ought to be realized. Of course, by itself, the approach does not provide a complete picture of how to implement an intelligent agent. Many other aspects – often already addressed and partially implemented in AI and robotics – must flank what is here suggested. In sum, the suggested approach to consciousness does not aim to be alternative to other approaches in AI or in robotics, rather it aims to tune them in a way that should be productive for consciousness.

A. From machine intelligence to machine consciousness

It may be useful to make a comparison between current attempts to implement intelligence and consciousness. The understanding of what intelligence is – or what it is not – fills book. The notion derives from the Latin verb 'intellegere' ('understanding', more literally 'to have the choice between', 'to read between' [39] [40]). As a scientific term originated in psychology, the concept of intelligence addresses the cognitive capabilities of an entity, usually a human being. In these general terms, the notion partially overlaps with the psychological notion of consciousness. This is also partially due to the fact that both notions (consciousness and intelligence) are mongrel concepts that encompass several vague and not entirely coherent aspects.

By and large, an agent with intelligence is often considered as divided into three central parts following each other:

1) recognition of external changes
- having **sensory** components in order to receive stimuli from the external environment

2) information processing
- being capable of processing the sensory data together with internal knowledge in order to adapt behaviour, **cognition**

3) reaction
- having the capability to interact with external environment, realized by **actuators**

Hereby, the Latin verb 'cognoscere' translates into 'conceptualize, recognize'. Cognition comprises the processes of information processing within an intelligent actor ('he sensory input is transformed, reduced, elaborated, stored, recovered, and used). Cognitive processes are divided into conscious and unconscious ones, e.g., by far not all learning processes are conscious. From this argumentation chain – *from intelligence to cognition to consciousness* – it follows that consciousness plays an important role in the understanding of intelligence.

Even if the majority of research done in the field of intelligence is directed towards human intelligence, the upper description states clearly that intelligence is NOT a primacy of humans. Obviously, many animals have a certain form of intelligence – proof is already given by observing your pet cat – even if it may differ from the human. Interestingly, the scientific status of 'consciousness' in animals continues to be hotly debated even if it is obvious that most animals have a phenomenal consciousness including a sense of pain, colour recognition, temperature etc. As mentioned above, the confusion is partly due to the variety of conceptions of consciousness. Researchers from different fields include very different aspects into the concept: a) phenomenal consciousness, b) the capability of thinking (thinking, remembering, planning, expecting), c) self-consciousness (awareness of oneself), d) consciousness of uniqueness (of oneself and of others) etc. Whereas phenomenal consciousness is probably part of most animals, it is still unclear if at least highly developed animals as mammals dispose of additional types of consciousness [5]. So the research space may be unfolded according to two broad criteria; one related to the kind of agents (animal, human or machine) and the other related to the kind of cognition involved (sensori-motor skill, symbolic capability aka traditional intelligence, linguistic capability, consciousness).

So which interim conclusions can be deduced? Machine consciousness lies in the promising middle ground between the extremes of biological chauvinism (i.e., only brains are conscious) and liberal functionalism (i.e., any behaviourally equivalent functional systems is conscious) [41]. One of the most central concepts behind 'intelligence' and perhaps the most difficult aspect to grasp is clearly not restricted to humans. From that it follows quite naturally that when building a technological system with a somewhat 'authentic intelligence', consciousness will have to play its part. Phenomenal consciousness – that is. It remains to be seen whether new concepts to realize this aspect will lead to insights into other components of consciousness.

B. *Weak versus strong machine consciousness*

The traditional and historically outdated distinction between weak AI and strong AI results from two different requirements: on the one hand, it follows from researchers focusing on different goals (more 'practical' vs. more 'principal'). On the other hand, a comprehensive philosophical debate on the nature of intelligence is driving the debate, including its exclusiveness or non-exclusiveness for humans (or other biological systems), ethical aspects, and the general possibility of reconstructing real intelligence, just to mention a few important aspects.

Weak AI addresses the position of artificial intelligence in philosophy that machines can demonstrate human-like intelligence, but do not necessarily have a mind, mental states or consciousness. Contrarily, strong AI supposes that some forms of artificial intelligence can reason and solve problems[8] as opposed to just making the humans feel that the machines are intelligent. In short: a weak AI-capable agent seems to be intelligent whereas a strong AI-capable agent is intelligent.

Obviously, the philosophical question behind this distinction is strongly related to the problem of consciousness. From that, it is not surprising that some authors suggested the possibility to distinguish between weak and strong artificial consciousness [3] [5]. In analogy to the weak vs. strong AI debate, weak artificial consciousness aims to deal with agents that behave as if they were conscious, at least in some respects, whereas strong artificial consciousness tries to address the design and construction of 'truly' conscious machines. Thus, the distinction between weak and strong artificial consciousness mirrors the dichotomy between true conscious agents and 'as if' conscious agents.

Although the distinction between weak and strong artificial consciousness sets a temporary working ground [5], it suggests a misleading view in so far as it suggests that a concept for a 'weak artificial consciousness' will help to gain a 'first understanding' on what consciousness might be and how it could be realized. Since it misses indispensables for the understanding of cognition – namely experience, i.e. phenomenal consciousness – the concept will not be adequate to overcome 'the riddle': Skipping the 'hard problem' is not a viable option in the business of making conscious machines [42].

Another argument may be raised against the temptation of 'weak artificial consciousness via the easy way': in nature, the development of consciousness goes along with increased intelligence. Most animals are exhibiting behavioural signs at least of phenomenological consciousness, human beings have a phenomenological consciousness and 'above'. 'Evolutionary optimization' is the most powerful optimization known so far (even if it takes its time). Thus, it seems to be highly unlikely that natural selection took such a long way to provide us with consciousness if there was a way to get all the advantages of a conscious being without actually producing it. Of course, this does not mean 'proof' – but the authors cannot help but to sincerely doubt it.

IV. CONCEPTS FOR BUILDING A CONSCIOUS MACHINE

Now, can a machine gain consciousness – that is, strong artificial consciousness as described in the previous section? Why – or why not? Is consciousness not a property of natural system that may thus be, at least in principle, realized by another physical system? And if so, how can that be done? Armed with the absence of a theoretical reason to reject the practical possibility, this paper addresses this issue.

---

[8] Sometimes, the term 'artificial general intelligence' ('AGI') is used to address strong AI, in particular by science fiction writers and within the community of futurists.

In the following, the authors are not outlining a strong theoretical formulation. Also, they are not capable – at this point – to give 'a full proof' (in a strict sense). Rather, it is the intention to show the inherent potential in the given interpretation of consciousness (Chapter II): As long as consciousness is interpreted as an 'internalistic' concept, there would be no change in modelling it: It remains to be something like 'internal interpreter, e.g. transforming 10.000 x 700 nm into "red"'. Nobody knows why and how, except from that it happens. The internalistic interpretation may be true but this would not help oneself to come closer to any understanding of the concept behind it. However, if consciousness is interpreted in the sense as the authors proposed (halfway between 'internalistic' and 'externalistic'), then it could be realized (at least as a toy model) as will be explained in the following. Thus, one can start to understand it and try to run tests on it, and so on. So by proposing this possible solution, the authors will sketch a 'lab scenario'. Here, promising off-the-shelf technologies are considered that may fill the bill if deployed in the proper way.

A. Preparations for a tentative architecture for a conscious agent

Currently, many robotic setups and architectures are the result of careful programming since designers aim to solve specific sensorimotor, relational, or logic issues. A classic example is offered by robotic feats like Robocup[9] where teams of robots exploit algorithms devised by their designers to compete together in a soccer match. Although their behaviours may be very clever it is not the result of real adaptation on a high-cognition level. Of course, there are some robots capable of learning new skills and to adapt to novel situation, at least to a certain degree. However, explicit attempts of integrating consciousness into a robots' intelligence are rare, and so far no model has been exceedingly convincing.

Compared to current robotic agents, biological agents like mammals and humans show a totally different kind of adaptability to novel stimuli. Mammals are capable of dealing with totally unexpected environmental challenges for which they could not possibly have any kind of inborn solution. Furthermore, it is a fair bet to assume that the complexity of their neural structure largely exceeds their genetic blueprint. Most mammals are capable not only of learning how to achieve goals but also of learning what goals have to be pursued [43] [44] – which is an important issue in respect to consciousness. As it has been observed [45] [46], the cortex shows an almost universal capability of autonomously adapting to novel kind of stimuli: 'The fact that humans can learn and adapt to problems that did not exist when the initial model (the neocortex) was created is proof of the generic nature of the mechanisms used by the human brain.' [47]. Thus, it makes sense to look for very general approaches capable, albeit with possible shortcomings, to model a unified and common approach to all aspects of cognition.

Empirical evidence shows that mammals exhibit a very high degree of neural plasticity and cognitive autonomy [48] [49] [50] to the extent that it is fair to suppose that any part of

the cortex might develop almost any cognitive skill. If this supposition were true, it would mean that the neocortex, and possibly the thalamocortical system, exploit some kind of rather general architectural principle, mainly independent of the kind of incoming data.

There have been various attempts in the past to devise a general cognitive architecture [47] [45] [46]. This paper makes yet another attempt and takes advantage of a rather simple idea: true autonomy entails teleological openness. By being teleologically open the authors mean that the system is capable of developing new goals autonomously on the basis of environmental conditions and stimuli [44].

B. Objectives and motivations of the architecture

What are the ideal features that a cognitive architecture should have in order to adapt to a partially unknown body and environment? On the basis of the available literature and the empirical evidence a series of key features and their justification may be listed:

- The architecture must be based on a very limited number of kinds for basic building blocks – each kind exploiting the same common structure. Thus, the description length of the architecture must be kept to a minimum.

- This basic module might be freely replicated in order to cope with multiple sensor modalities and demanding incoming stimuli. This should ensure scalability.

- The basic module has to be able to develop its own goals and to use them both for its own development and for interacting with other modules. This should allow developing intentionality and a tight environment-architecture coupling.

- In principle, adding further modules (constrained only by the system resources) should lead to an increase in performances. Once again, this is important for scalability.

An architecture with the above features should be able to adapt to unknown situations and with a minimum of predesign. Rather than specifying all the algorithms and their mutual relationships, the above approach suggests a recipe to build a cognitive architecture given a body and an environment. Such a recipe is a lot less demanding in terms of description and a priori knowledge than a detailed plan. Furthermore, a recipe of such a universal scope offers many more advantages in terms of adaptability and flexibility.

Thus, the architecture the authors are willing to implement must satisfy the following requirements:

- Structure:
    o it must be scalable
    o it must be adaptable
    o it must take advantage of memory more than speed
    o it must be hierarchical
- Capabilities:

---

[9] http://www.robocup.org/

○ it must take into account the whole history of the system

○ it must develop fine grained new goals

○ it must develop overarching goals emerging out of the finer structure

- Additional Do's and Don'ts:

○ it must not rely on explicit algorithms

○ it might have a limited number of more specialized versions of the same elementary block (for fine tuning, better performance, and optimization)

○ it must be coherent to what one knows about the biological structure of a mammalian brain

C. Combining multi agent systems with genetic algorithms

A tentative approach might be to realize a robot's brain as a multi-agent system (MAS) once such an endeavour may find support by some additional key hypothesis as to the physical foundations of consciousness. MAS have been discussed already as a possible model to realize artificial brains, or as a model to explain the function of a brain (e.g. [51]). They have also been discussed as a possible extension of cognitive architectures as e.g. within the hybrid design oft CLARION (e.g. [52]). In computer sciences, MAS have become a very popular instrument during the last years when modelling complex heterogeneous distributed systems, which are organized 'bottom up'. Their strength lies in predicting appearance of complex phenomena. The single agents have a certain degree of autonomy, they represent local views (in general, no agent has a full global view of the system, due to the complexity and the number of dynamically changing external dependencies), and they work decentralized ('no master brain'). Topics where multi-agent systems are used include in particular the modelling of social and/or cooperative structures. Multi-agent system may be one of the key architectural principles necessary for a conscious mind.

Taking the new approach to consciousness as described in Chapter II as basis, in such a MAS each software agent would represent one 'conscious-lock' to a certain key, an external phenomenon. Thus, the resulting robotical brain would be conscious of the external events it has the appropriate locks for it, and the mechanism of building this consciousness would be exactly the same as for the human brain. So, the tentative idea is that MAS could offer the necessary architectural backbone for a conscious mind and that, once tuned to satisfy to some specific requirement, may be indeed the workable tool to begin designing a new kind of cognition.

At least three questions pose themselves immediately:

- **Complexity:** One may argue that by this approach, only a small number of locks can be realized due to the enormous programming effort needed otherwise.

- **Specification**: An even harder objection might be that in this way, the programmer may tend to mainly 'imitate' the human consciousness but does not develop one which is appropriate for the given robot with a certain form, function and so on.

- **Proof**: A third difficult point is the answer to the question as how one would like to *prove* that a certain robot really has a consciousness in a strong sense (compare Chapter IV.B).

To tackle all three problems with one approach, optimization algorithms have to be integrated, allowing to improve the multi agent system during runtime. Here, due to their 'closeness' to the underlying problem (a developing brain), genetic algorithms might form a natural choice: The 'consciousness-locks' have to be specialized to species, their mode of living, and the challenges presented to them[10]. Their special characteristics are probably not the result of some kind of 'biological master plan' for all living beings, but the result of a species-exclusive evolutionary process, which over millions of years has favoured individuals which are better adapted to their environment than others. In this understanding, the consciousness-lock (realized through multi agents) would be subject to the same evolutionary process, which has driven the whole design of a certain species, including the body shapes, motor skills, brain structure and the like. Genetic algorithms are precisely reproducing this kind of development.

The idea of using genetic algorithms to build a conscious brain is also one of the central design principles behind the cognitive architecture [53]. Genetic algorithms are a part of evolutionary computing, which is a rapidly growing area of artificial intelligence. They are inspired by Darwin's theory about evolution. The idea was introduced in the 1960s by Ingo Rechenberg in his work 'Evolution strategies'. His idea has been extended by many other researchers over the last decades. Today, they play an important role in many complex optimization problems and form an important concept for machine learning approaches. Genetic algorithms use mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. Over several generations, systems are optimized: Pairs of first generation solutions are taken and recombined. The 'fittest' solutions of this match are selected for the next generation. Mutations are used to enhance the genetic variety and thus, the overall solution space. The optimization goal – in nature given through environment and the corresponding challenges – is realized through a so-called fitness function which determines the quality of the solutions. Lately, combining multi agent systems with genetic algorithms has become popular in certain field as e.g. automated testing scenarios.

Assuming that consciousness is a capability of higher development of life forms, the following digest gives a first

---

[10] Consider the following example: Literature states that cats are somewhat colour-blind concerning the colour red, they see it as a shade of grey (whereas they have a perfect colour vision concerning e.g. green and blue). Well, the first finding is that one cannot be really sure about that, since one can only predict that from their eye anatomy – but what kind of 'consciousness' cats really have concerning the colour red is a totally different topic because at this point the design of key-lock-structure is unknown. It may be totally different to the human one. The second – and much more important – insight is that it might be less important for a cat to be capable of seeing red than for example for a bear: cats – being carnivores – do not have to differentiate between ripe and unripe apples since they would not eat them anyway. For a bear on the other hand – being omnivores consuming a large portion of fruits daily – the situation may show itself quite differently.

impression of the number of genetical iterations which are necessary to produce this kind of complex structures: About 3.5 billion years ago, the first life forms developed, monads with a very limited range of functionalities. Based on the development of genetical heredity through DNA molecules, advancements and progresses could be passed over to the next generation leading to first plants and simple animals which arose about 700 million years ago. 200 million years ago, mammals started to populate the earth. Humanoid life forms developed 70 million years ago and the homo sapiens species is only 500.000 years old. Even if it is difficult to tell from which stage in evolution consciousness has first entered the scene, referring to the current state of the art its development is part of a growing and more and more complex brain (ibidem). From this analysis, the reader might understand why the others consider genetic algorithms for the optimization job!

D. The practical side

How to proceed 'practically', meaning: how exactly are the genetic algorithms used to re-build the evolution of a robotical brain?

- Regarding the development of consciousness, one would start with a couple of given perceptions, each of them realized through a single agent, say regarding colours, temperature and the like which seem to play an important role for all living beings – a 'basic set' of conscious perceptions, so to say. This is 'easy' – and would form the 'first generation solution'.

- Now, to make the system learn new conscious elements, the second preparative step is to place the robot in a certain challenging environment – meaning that certain tasks have to be given to him – in order to challenge his 'consciousness enhancement'.

- Next, genetic algorithms come into play in order to produce variants of the robot's 'brain structure': the single agents will be multiplied and altered through the means of the genetic algorithms. They will become multiplied, more complex and more varying. Some of the 'new' solutions will not survive as they do not particularly contribute to the tasks the robot is given. Others will survive as they enhance the robot's capabilities to deal with its tasks. This, the resulting brain structure, will turn out having consciousness-locks which are complex and adapted to the individual needs of the specific kind of robot and its environment.

From that, there are two possibilities to infer that the robotic brain is really developed something like a consciousness by using genetic algorithms:

- First, the direct inspection would address the source code itself. Starting from a 'basis set' of agents in a MAS, the resulting system would consist of old and new software agents, the latter representing new conscious capabilities. The new code can be investigated, varied and different tests cases could be designed and analysed.

- A more 'indirect' inspection would be: A test scenario could be designed where consciousness for a certain perception area would definitely be necessary to solve a certain task. By design, this particular conscious perception would not be part of the basic conscious skills the system is starting with[11]. Now, if – after of a couple of some (more) 'genetic rounds' – the robotic brain would come up with new solutions for the given task, which definitely requires the enhancement of its consciousness, this would be a strong signal that it has developed a new perception in a certain area.

So, the combination of multi agent systems with genetic algorithms allows overcoming the upper mentioned problems:

- **Complexity**: Starting from a small number of locks, their expansion is realized by genetic algorithms which enhance the number of locks in order to optimize the system's behaviour

- **Specification**: Since the optimization takes places in relation to a certain environment including specific challenges and particular tasks, the robotical brain develops a consciousness which is adapted to its own needs.

- '**Proof**': The proof of whether a consciousness has been developed is not complete. However, on the more direct side, the investigation of the auto-generated source-code will deliver new insights. From the perspective of an indirect proof, it would address the development of a new conscious aspect rather than its existence. If a robot can adapt to a certain situation IF and only IF it develops a conscious perception for something that will be a strong hint that consciousness has been developed.

V. SUMMARY & OUTLOOK

Putting it all together: According to [54], there are three motivations to pursue artificial consciousness [55] [56] [57]:

1) implementing and designing machines resembling human beings (cognitive robotics);
2) understanding the nature of consciousness (cognitive science);
3) implementing and designing more efficient control systems.

Based on the presented new approach to consciousness lying between internalism and externalism, a possible technological design for a conscious machine has been

---

[11] Consider the following example: assume having a robot with colour consciousness as one of the basic components. This robot is part of a cooperative structure with other robots and humans, working together in a production line. Due to long geographical distances within factory, it would be absolutely necessary to be capable to 'visualize temperature', meaning to have a visual perception for extended areas within wavelengths between 700 nm and 1 mm (infrared). Here, humans would not be able to 'see' the wavelengths, since they have no consciousness for this wavelength area. However, the robot (using genetic algorithms on the multi agent system which is forming the 'conscious part' of the robotical brain) could develop a perception for this wavelength. By that, the robot might be capable to solve the task – finding a heat leak in an extended machinery – opposed to the human. Due to the fact that the brain is built in the upper described key-lock system that would mean that some kind of 'new' consciousness has been developed.

sketched addressing the upper mentioned goals. The approach is taking advantage of an architecture exploiting self-development of new goals, intrinsic motivation, and situated cognition. From a technological point of view, multi agent systems are used to model independent conscious perceptions. Genetic algorithms – as a subgroup of evolutionary algorithms – come into play to mimic the biological evolution of the brain's structure, thus allowing in general for adaptivity and scalability, and assuring some coherence to what humans know about the biological structure of brains of higher developed animals as e.g. mammals.

The architecture does not pretend to be either conclusive or experimentally satisfying. In the future, this rather sketchy outline of a cognitive architecture will be enhanced to a satisfying and more comprehensive architectural model. At this point, the authors will also integrate components of a cognitive architecture that has been partially implemented in previous setups [58] [59] [60]. The goal of the full architecture model is the implementation of the kind of development and environmental coupling through consciousness which was described in the previous sections.

On the other hand, up-to-date examples of highly distributed systems will be analysed in respect to their decision making processes (e.g., IBM's Watson which is operating on very distributed resources originally). These Systems show a new quality of artificial intelligence from which can be learned from: If high-developed intelligence includes consciousness, and if these big data oriented approached do produce results with a certain intelligence, than the interesting question arises whether these systems MUST have developed a certain consciousness, as part of their intelligence. If there is any merit in that, one could observe the emergence and the 'building' of consciousness in artificial system. Just by watching and interpreting, one could avoid arguing on the basis of biases and presumptions, bringing the whole debate back into the laboratories of natural sciences.

REFERENCES

[1] C. Adami, "What Do Robots Dream Of?," Science, vol. 314, no. 5802, pp. 1093–1094, Nov. 2006.

[2] G. Buttazzo, "Can a machine ever become self-aware?," in Artificial Humans, R. Aurich, W. Jacobsen, and G. Jatho, Eds. Los Angeles: Goethe Institut, 2000, pp. 45–49.

[3] A. Chella and R. Manzotti, Eds., Artificial Consciousness. Imprint Academic, 2007.

[4] D. Gamez, "Progress in machine consciousness," Consciousness and Cognition, vol. 17, no. 3, pp. 887–910, Sep. 2008.

[5] O. Holland, Ed., Machine Consciousness. Imprint Academic, 2003.

[6] D. J. Chalmers, in The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, U.S.A., 1996, pp. xvii, 414.

[7] J. Kim, Mind in a Physical World. Cambridge, Mass: MIT Press, 1998.

[8] T. Nagel, "What is It Like to Be a Bat?," Philosophical Review, vol. 83, no. October, pp. 435–50, 1974.

[9] S. E. Palmer, Vision science: photons to phenomenology. Cambridge, Mass.: MIT Press, 1999.

[10] F. Crick, Astonishing Hypothesis: The Scientific Search for the Soul. New York: Touchstone, 1994.

[11] C. Koch and G. Tononi, "Can Machines Be Conscious?," IEEE Spectrum, vol. 45, no. 6, pp. 47–51, Oct. 2008.

[12] A. Revonsuo, Inner Presence: Consciousness as a Biological Phenomenon. Cambridge, Mass: MIT Press, 2006.

[13] Editors of Scientific American Magazine, The Scientific American Book of the Brain. Lyons Press, 1999.

[14] A. Damasio, Feeling of What Happens: Body and Emotion in the Making of Consciousness. New York: Harcourt Brace & Company, 1999.

[15] G. M. Edelman and G. Tononi, A Universe Of Consciousness: How Matter Becomes Imagination. London: Allen Lane, 2000.

[16] W. James, The Principles of Psychology. New York: Henry Holt and Company, 1890.

[17] B. Kuipers, "Drinking from the Firehose of Experience," Journal of Artificial Intelligence in Medicine, vol. 44, no. 2, pp. 155–170, 2008.

[18] R. Kurzweil, How to Create a Mind: The Secret of Human Thought Revealed. New York: Viking Adult, 2012.

[19] J. R. Searle, Minds, Brains, and Science. Harvard University Press, 1984.

[20] S. Yablo, "Advertisement for a Sketch of an Outline of a Prototheory of Causation," in Causation and Counterfactuals, J. Collins, N. Hall, and L. A. Paul, Eds. Cambridge, Mass: MIT Press, 2004, pp. 119–137.

[21] D. C. Dennett and M. Kinsbourne, "Time and the observer: The where and when of consciousness in the brain," Behavioral and Brain Sciences, vol. 15, no. 02, pp. 183–201, 1992.

[22] R. Flach and P. Haggard, "The cutaneous rabbit revisited," Journal of Experimental Psychology: Human Perception and Performance, vol. 32, no. 3, pp. 717–732, 2006.

[23] M. Cook, "Descartes and the Dustbin of the Mind," History of Philosophy Quarterly, vol. 13, no. 1, pp. 17–33, 1996.

[24] S. Shoemaker, "Qualities and Qualia: What's in the Mind?," Philosophy and Phenomenological Research, vol. 50, pp. 109–131, 1990.

[25] M. R. Bennett and P. M. S. Hacker, Philosophical Foundations of Neuroscience. Malden, Mass: Blackwell, 2003.

[26] R. Manzotti and Moderato, "Is Neuroscience Adequate As The Forthcoming 'Mindscience'?," Behavior and Philosophy, vol. 38, pp. 1–29, 2010.

[27] J. Dewey, Experience And Nature. Chicago: Open Court, 1925.

[28] W. T. Rockwell, Neither Brain nor Ghost. Cambridge, Mass: MIT Press, 2005.

[29] M. Johnston, "Appearance and Reality," in The Manifest, Princeton, NJ: Princeton University Press, 2002.

[30] C. S. Hurovitz, S. Dunn, G. W. Domhoff, and H. Fiss, "The Dreams of Blind Men and Women: A Replication and Extension of Previous Findings," Dreaming, vol. 9, no. 2–3, pp. 183–193, 1999.

[31] N. H. Kerr and G. W. Domhoff, "Do the Blind Literally 'See' in Their Dreams? A Critique of a Recent Claim That They Do," Dreaming, vol. 14, no. 4, pp. 230–233, 2004.

[32] A. Revonsuo and C. Salmivalli, "A content analysis of bizarre elements in dreams," Dreaming, vol. 5, no. 3, pp. 169–187, 1995.

[33] E. Schwitzgebel, C. Huang, and Y. Zhou, "Do we dream in color? Cultural variations and skepticism," Dreaming, vol. 16, no. 1, pp. 36–42, 2006.

[34] P. S. Churchland and T. J. Sejnowski, "Neural Representation and Neural Computation," Philosophical Perspectives, vol. 4, pp. 343–382, 1990.

[35] F. Tong and M. S. Pratte, "Decoding patterns of human brain activity," Annu Rev Psychol, vol. 63, pp. 483–509, 2012.

[36] B. J. Baars, "In the Theatre of Consciousness: Global Workspace Theory, a Rigorous Scientific Theory of Consciousness," Journal of Consciousness Studies, vol. 4, no. 4, pp. 292–309, 1997.

[37] M. Shanahan, Embodiment and the inner life: Cognition and Consciousness in the Space of Possible Minds. Oxford University Press, USA, 2010.

[38] G. Tononi, "An information integration theory of consciousness," BMC Neuroscience, vol. 5, no. 42, pp. 1–22, 2004.

[39] D. J. Chalmers, "Facing Up to the Problem of Consciousness," Journal of Consciousness Studies, vol. 2, no. 3, pp. 200–219, 1995.

[40] R. Pfeifer and C. Scheier, Understanding Intelligence. Cambridge, MA, USA: MIT Press, 1999.

[41] R. Jackendoff, Consciousness and the Computational Mind. Cambridge, Mass: MIT Press, 1987.

[42] R. Manzotti, "Is Consciousness Just Conscious Behavior?," Int. J. Mach. Conscious., vol. 3, no. 2, pp. 353–360, Dezember 2011.

[43] R. Manzotti, F. Mutti, G. Gini, and S.-Y. Lee, "Cognitive Integration through Goal-Generation in a Robotic Setup," in Biologically Inspired Cognitive Architectures 2012, A. Chella, R. Pirrone, R. Sorbello, and K. R. Jóhannsdóttir, Eds. Springer, 2013, pp. 225–231.

[44] R. Manzotti and V. Tagliasco, "From behaviour-based robots to motivation-based robots," Robotics and Autonomous Systems, vol. 51, no. 2–3, pp. 175–190, Mai 2005.

[45] D. George and J. Hawkins, "Towards a Mathematical Theory of Cortical Micro-circuits," PLoS Comput Biol, vol. 5, no. 10, p. e1000532, Oktober 2009.

[46] J. Hawkins and S. Blakeslee, On Intelligence. New York: Times Books, 2004.

[47] D. George, "How the Brain Might Work: A Hierarchical and Temporal Model for Learning and Recognition," Stanford University, Stanford, CA, USA, 2008.

[48] J. Sharma, A. Angelucci, and M. Sur, "Induction of visual orientation modules in auditory cortex," Nature, vol. 404, no. 6780, pp. 841–847, Apr. 2000.

[49] J. Sharma, V. Dragoi, J. B. Tenenbaum, E. K. Miller, and M. Sur, "V1 Neurons Signal Acquisition of an Internal Representation of Stimulus Location," Science, vol. 300, no. 5626, pp. 1758–1763, Jun. 2003.

[50] M. Sur, P. E. Garraghty, and A. W. Roe, "Experimentally induced visual projections into auditory thalamus and cortex," Science, vol. 242, no. 4884, pp. 1437–1441, Dec. 1988.

[51] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, Principles of Neural Science, Fifth Edition, 5 edition. New York: McGraw-Hill Medical, 2012.

[52] R. Sun, "The CLARION Cognitive Architecture: Extending Cognitive Modeling to Social Simulation," in Cognition and Multi-Agent Interaction, R. Sun, Ed. Cambridge University Press, 2005, pp. 79–100.

[53] B. Goertzel and D. Duong, "OpenCog NS: A Deeply-Interactive Hybrid Neural-Symbolic Cognitive Architecture Designed for Global/Local Memory Synergy," in AAAI Fall Symposium: Biologically Inspired Cognitive Architectures'09, 2009, vol. FS-09–01.

[54] R. Sanz and C. Hernández, "Towards Architectural Foundations for Cognitive Self-aware Systems," in Biologically Inspired Cognitive Architectures 2012, A. Chella, R. Pirrone, R. Sorbello, and K. R. Jóhannsdóttir, Eds. Springer, 2013, pp. 53–53.

[55] J. Bongard, V. Zykov, and H. Lipson, "Resilient Machines Through Continuous Self-Modeling," Science, vol. 314, no. 5802, pp. 1118–1121, Nov. 2006.

[56] R. Pfeifer and J. Bongard, How the Body Shapes the Way We Think: A New View of Intelligence. New York: Bradford Books, 2006.

[57] R. Sanz, I. López, M. Rodríguez, and C. Hernández, "Principles for consciousness in integrated cognitive control," Neural Networks, vol. 20, no. 9, pp. 938–946, Nov. 2007.

[58] R. Manzotti, L. Papi, and S.-Y. Lee, "Does radical externalism suggest how to implement machine consciousness?," in Biologically Inspired Cognitive Architectures 2011, A. Samsonovich and K. Jóhannsdóttir, Eds. Amsterdam: IOS Press, pp. 232–240.

[59] R. Manzotti, "A Process-Based Architecture for an Artificial Conscious Being," in Process Theories: Crossdisciplinary Studies in Dynamic Categories, J. Seibt, Ed. Springer Netherlands, 2003, pp. 285–312.

[60] R. Manzotti, "Machine Free Will: Is Free Will a Necessary Ingredient of Machine Consciousness?," in From Brains to Systems, C. Hernández, R. Sanz, J. Gómez-Ramirez, L. S. Smith, A. Hussain, A. Chella, and I. Aleksander, Eds. Springer New York, 2011, pp. 181–191.

Rough Approximations for Incomplete Information∗

Jun-Fang LUO

College of Mathematics
Southwest Jiaotong University
Chengdu, China

Ke-Yun QIN

College of Mathematics
Southwest Jiaotong University
Chengdu, China

Abstract—**Rough set under incomplete information has been extensively studied. Based on valued tolerance relation for incomplete information system, several approaches were presented to dealing with the attribute reductions and rule extraction. We point out some drawbacks in the existing papers for valued tolerance relation based rough approximations and propose a new kind of rough approximation operators which is a generalization of Pawlak approximation operators for complete information system. Some basic properties of the approximation operators are investigated.**

Keywords—Rough set; tolerance relation; valued tolerance relation

I. Introduction

The rough set theory (RST), proposed by Pawlak[7], is an effective tool for data analysis. It can be used in information system to describe the dependencies among attributes and evaluate the significance of attributes and derive decision rules. In an information system, each object in the universe is associated with some information that is characterized by a set of attributes. Objects characterized by the same information are indiscernible with the available information about them. Based on the indiscernibility relation, classical rough set theory has been used successfully in attribute reduction of information and decision systems.

In many practical situations, it may happen that the precise values of some of the attributes in an information system are not known, i.e. are missing or known partially. Such a system is called an incomplete information system. In order to deal with incomplete information systems, classical rough sets have been extended to several general models by using other binary relations or covers on the universe[1,8-10,14,16]. Based on these extended rough set models, the researchers have put forward several meaningful indiscernibility relations in incomplete information system to characterize the similarity of objects. For instance, Slowinski[11] proposed two different approaches to replace unknown value of attribute by specific subsets of values. Grzymala et al[2,3] performed computational studies on the medical data, where unknown values of attributes were replaced using probabilistic techniques. Kryszkiewicz introduced a kind of indiscernibility relation, called tolerance relation, to handle incomplete information tables[5,6]. Stefanowski[12] introduced two generalizations of the rough sets theory to handle the missing value. The first generalization introduces the use of a non symmetric similarity relation in order to formalize the idea of absent value semantics. The second proposal is based on the use of valued tolerance relations. A logical analysis and the computational

experiments show that for the valued tolerance approach it is possible to obtain more informative approximations and decision rules than using the approach based on the simple tolerance relation. The tolerance relation has also been generalized to constrained similarity relation and constrained dissymmetrical similarity relation[4,13,15]. This paper is devoted to the discussion of valued tolerance relation based rough approximation operators. We pointed out that the lower (upper) approximability presented in [12] is not the generalization of Pawlak approximations. A new kind of lower (upper) approximability is proposed. Some basic properties are analyzed.

II. Similarity Relation for Incomplete Information Table

Rough sets have been introduced by Pawlak[7] as an approach for analyzing vague information. Following Pawlak, an information table is a pair $IT = (U, A)$, where U is a set of objects, A is a set of attributes such that $\forall a \in A$, $a : U \to V_a$, V_a is a domain of a and $V = \bigcup_{a \in A} V_a$. Each subset of attributes $B \subseteq A$ defines an indiscernibility relation $IND(B)$ as:

$$IND(B) = \{(x, y) \in U \times U; \forall a \in B(a(x) = a(y))\} \qquad (1)$$

Clearly, $IND(B)$ is an equivalence relation. Let U/B be the family of all the equivalence classes of the equivalence relation $IND(B)$. For each $X \subseteq U$, the lower and upper approximation of X are defined by [7]:

$$\underline{B}(X) = \{x \in U; [x]_R \subseteq X\},$$

$$\overline{B}(X) = \{x \in U; [x]_R \cap X \neq \varnothing\}.$$

The rough set is characterized by its lower and upper approximations.

Let $IT = (U, A)$ be an incomplete information table. Kryszkiewicz[5] introduced the notion of tolerance relation. The key point in this approach is to interpret an unknown value of the attribute as similar to all other possible values for this attribute. Such an interpretation corresponds to the idea that such values are just missing, but they do exist. The tolerance relation T_B with respect to $B \subseteq A$ is defined as [5]:

$$T_B = \{(x, y) \in U \times U; \forall a \in B(a(x) = a(y) \vee a(x) = * \vee a(y) = *)\} \quad (2)$$

Clearly T_B is a reflexive and symmetric relation, but not necessarily transitive. We denote by $T_B(x)$ the tolerance class

of x, that is $T_B(x) = \{y \in U; (x, y) \in T_B\}$. The lower and upper approximations of $X \subseteq U$ are

$$\underline{T_B}(X) = \{x \in U; T_B(x) \subseteq X\},$$

$$\overline{T_B}(X) = \{x \in U; T_B(x) \cap X \neq \varnothing\}.$$

Stefanowski[12] introduced the absent values semantics for incomplete information tables. In this approach it is assumed that objects may be partially described not only because of our imperfect knowledge, but also because it is definitely impossible to describe them on all the attributes. The unknown values are not allowed to compare. Based on this point, the similarity relation S_B is defined as:

$$S_B = \{(x, y) \in U \times U; \forall a \in B(a(x) = a(y) \vee a(x) = *)\} \quad (3)$$

S_B is a reflexive and transitive relation, but not necessarily symmetric. Based on S_B, the lower and upper approximations of $X \subseteq U$ are defined as:

$$\underline{S_B}(X) = \{x \in U; S_B(x) \subseteq X\},$$

$$\overline{S_B}(X) = \{x \in U; S_B(x) \cap X \neq \varnothing\}.$$

where $S_B(x) = \{y \in U; (x, y) \in S_B\}$.

In order to characterize incomplete information more precisely, Stefanowski[12] introduced the notion of valued tolerance relation. Let $a \in A$ be an attribute and V_a the set of its known values. Given an object $x \in U$ with $a(x) = *$, the probability that $a(x) = e$ for any $e \in V_a$ is equal to $|V_a|^{-1}$. Moreover, if both values are unknown, then the probability that x is similar to y on the attribute is $|V_a|^{-2}$. Thus, the probability $R_a(x, y)$ for x is similar to y is defined by:

$$R_a(x, y) = \begin{cases} 1; a(x) = a(y) \in V_a \\ 0; a(x) \in V_a \wedge a(y) \in V_a \wedge a(x) \neq a(y) \\ |V_a|^{-1}; (a(x) = * \wedge a(y) \in V_a) \vee (a(x) \in V_a \wedge a(y) = *) \\ |V_a|^{-2}; a(x) = * \wedge a(y) = * \end{cases} \quad (4)$$

Let $R_B(x, y) = \prod_{a \in B} R_a(x, y)$. Based on $R_B(x, y)$, the $B-$ lower and the $B-$ upper approximability of X by set Z are defined as:

$$\mu X_B(Z) = T_{z \in Z} T_{x \in \Theta_B(z)} I(R_B(z, x), x) \quad (5)$$

$$\mu X^B(Z) = T_{z \in Z} S_{x \in \Theta_B(z)} T(R_B(z, x), x) \quad (6)$$

where $\Theta_B(z)$ is the tolerance class of element z, x is the membership degree of element x in the set X ($x \in \{0,1\}$), T, S and I are t-norm, t-conorm and fuzzy implication respectively.

In this model, each subset of U may be a lower or upper approximation of X, but to a different degree which is denote as lower (upper) approximability.

Theorem 1 Let $IT = (U, A)$ be an incomplete information table.

(1) If $Z_1 \subseteq Z_2$, then $\mu X_B(Z_1) \geq \mu X_B(Z_2)$, $\mu X^B(Z_1) \geq \mu X^B(Z_2)$.

(2) If $X_1 \subseteq X_2$, then $\mu X_{1B}(Z) \leq \mu X_{2B}(Z)$, $\mu X_1^B(Z) \leq \mu X_2^B(Z)$.

(3) If $B_1 \subseteq B_2$, then $\mu X_{B_1}(Z) \leq \mu X_{B_2}(Z)$, $\mu X^{B_1}(Z) \geq \mu X^{B_2}(Z)$.

Proof: (1) By $T_{x \in \Theta_B(z)} I(R_B(z, x), x) \in [0, 1]$ and $Z_1 \subseteq Z_2$, we have

$$\mu X_B(Z_2) = T_{z \in Z_2} T_{x \in \Theta_B(z)} I(R_B(z, x), x)$$

$$= T_{z \in Z_1} T_{x \in \Theta_B(z)} I(R_B(z, x), x) \cdot T_{z \in Z_2 - Z_1} T_{x \in \Theta_B(z)} I(R_B(z, x), x)$$

$$\leq T_{z \in Z_1} T_{x \in \Theta_B(z)} I(R_B(z, x), x) = \mu X_B(Z_1)$$

So we have $\mu X_B(Z_1) \geq \mu X_B(Z_2)$. $\mu X^B(Z_1) \geq \mu X^B(Z_2)$ can be proved similarly.

(2) Let x_1 be the membership degree of element in the set X_1 and x_2 be the membership degree of element in the set X_2 ($x_1, x_2 \in \{0,1\}$). So $\mu X_{1B}(Z) = T_{z \in Z} T_{x \in \Theta_B(z)} I(R_B(z, x), x_1)$ and $\mu X_{2B}(Z) = T_{z \in Z} T_{x \in \Theta_B(z)} I(R_B(z, x), x_2)$. By $X_1 \subseteq X_2$, it follows that $I(R_B(z, x), x_1) \leq I(R_B(z, x), x_2)$. Consequently we have $\mu X_{1B}(Z) \leq \mu X_{2B}(Z)$. $\mu X_1^B(Z) \leq \mu X_2^B(Z)$ can be proved similarly.

(3) By $B_1 \subseteq B_2$ we have $\Theta_{B_1}(z) \supseteq \Theta_{B2}(z)$ and $R_{B_1}(z, x) \geq R_{B_2}(z, x)$. Thus

$$\mu X_{B_1}(Z) = T_{z \in Z} T_{x \in \Theta_{B_1}(z)} I(R_{B_1}(z, x), x)$$

$$= T_{z \in Z} T(T_{x \in \Theta_{B_2}(z)} I(R_{B_1}(z, x), x), T_{x \in \Theta_{B_1}(z) - \Theta_{B_2}(z)} I(R_{B_1}(z, x), x))$$

$$\leq T_{z \in Z} T_{x \in \Theta_{B_2}(z)} I(R_{B_1}(z, x), x)$$

$$\leq T_{z \in Z} T_{x \in \Theta_{B_2}(z)} I(R_{B_2}(z, x), x) = \mu X_{B_2}(Z).$$

$$\mu X^{B_2}(Z) = T_{z \in Z} S_{x \in \Theta_{B_2}(z)} T(R_{B_2}(z, x), x)$$

$$\leq T_{z \in Z} S_{x \in \Theta_{B_1}(z)} T(R_{B_2}(z, x), x)$$

$$\leq T_{z \in Z} S_{x \in \Theta_{B_1}(z)} T(R_{B_1}(z, x), x) = \mu X^{B_2}(Z).$$

III. ROUGH APPROXIMATIONS BASED ON VALUED TOLERANCE RELATION

Let $IT = (U, A)$ be an incomplete information table. Intuitively, $R_B(x, y)$ is the similarity degree of x and y with respect to attribute set B. Clearly, $R_B(x, y) \in [0, 1]$. If $DT = (U, A)$ is complete, then $R_B(x, y)$ will degenerate to indiscernibility relation $IND(B)$. We note that the lower (upper) approximability will decrease with the increase of elements in Z. This does not coincide with the basic idea of Pawlak's rough set. In Pawlak rough set model, whether a set is lower

(upper) approximation is definite. It does not happen that, the smaller the set, the more possible it is lower (upper) approximation. Actually, (5) and (6) are based on the observation that, in classical rough set,

$$Z = X_B \Rightarrow \forall z \in Z (\Theta_B(z) \subseteq X) ,$$

$$Z = X^B \Rightarrow \forall z \in Z (\Theta_B(z) \cap X \neq \varnothing) ,$$

where X_B and X^B are lower and upper approximations of X respectively. It is worth noticing that this is a necessary condition but not sufficient. Actually, we have

Theorem 2 Let $IT = (U,A)$ be an incomplete information table, $B \subseteq A$.

(1) $Z = X_B$ if and only if $\forall z \in Z (\Theta_B(z) \subseteq X) \wedge \forall z \in \sim Z (\Theta_B(z) \cap \sim X \neq \varnothing)$.

(2) $Z = X^B$ if and only if $\forall z \in Z (\Theta_B(z) \cap X \neq \varnothing) \wedge \forall z \in \sim Z (\Theta_B(z) \subseteq \sim X)$.

Proof: (1) Let $Z = X_B$. For each $z \in Z$, we have $\Theta_B(z) \subseteq X$ by $z \in X_B$. Furthermore, for each $z \in \sim Z$, we have $\Theta_B(z) \not\subset X$ by $z \notin X_B$. Thus $\Theta_B(z) \cap \sim X \neq \varnothing$. Conversely, for each $z \in Z$, by $\Theta_B(z) \subseteq X$ we have $z \in X_B$ and hence $Z \subseteq X_B$. Furthermore, for each $z \in X_B$ we have $\Theta_B(z) \subseteq X$ and hence $\Theta_B(z) \cap \sim X = \varnothing$. It follows that $z \notin \sim Z$ and thus $z \in Z$. Consequently, $X_B \subseteq Z$.

(2) can be proved similarly.

Based on this theorem, we propose the following definition.

Definition 1 Let $IT = (U,A)$ be an incomplete information table, $B \subseteq A$. The $B -$ lower approximability $\gamma X_B(Z)$ and the $B -$ upper approximability $\gamma X^B(Z)$ of X by set Z are defined as:

$$\gamma X_B(Z) = T(\mu X_B(Z), \delta X_B(Z)) ,$$

$$\gamma X^B(Z) = T(\mu X^B(Z), \delta X^B(Z)) ,$$

Where

$$\delta X_B(Z) = T_{z \in U-Z} S_{x \in \Theta_B(z)} T(R_B(z,x), 1-x) ,$$

$$\delta X^B(Z) = T_{z \in U-Z} T_{x \in \Theta_B(z)} I(R_B(z,x), 1-x) .$$

Theorem 3 Let $IT = (U,A)$ be an incomplete information table, and $B \subseteq A$, $X, Z \subseteq U$.

(1) $\gamma X_B(Z) = \gamma(\sim X)^B(\sim Z)$.

(2) $\gamma X^B(Z) = \gamma(\sim X)_B(\sim Z)$.

Theorem 4 Let $IT = (U,A)$ be a complete information table and $B \subseteq A$, $X, Z \subseteq U$.

(1) $Z = X_B$ if and only if $\gamma X_B(Z) = 1$.

(2) $Z = X^B$ if and only if $\gamma X^B(Z) = 1$.

Proof: (1) For complete information table, $R_B(z,x) \in \{0,1\}$. Let $Z = X_B$. For each $z \in Z$, we have $\Theta_B(z) \subseteq X$, where $\Theta_B(z)$ is the equivalence class containing z. For any $x \in \Theta_B(z)$, it follows that $x \in X$, and hence $x = 1$. Thus $I(R_B(z,x), x) = 1$ and hence $\mu X_B(Z) = T_{z \in Z} T_{x \in \Theta_B(z)} I(R_B(z,x), x) = 1$.

For each $z \in U - Z$, it follows that $z \notin X_B$ and thus $\Theta_B(z) \not\subset X$. There exists $x \in \Theta_B(z)$ such that $x \notin X$. Thus $R_B(z,x) = 1 - x = 1$ and $T(R_B(z,x), 1-x)) = 1$. Consequently we have $\delta X_B(Z) = T_{z \in U-Z} S_{x \in \Theta_B(z)} T(R_B(z,x), 1-x) = 1$. So, $\gamma X_B(Z) = 1$ as required.

Conversely, assume that $\gamma X_B(Z) = 1$. It follows that $\mu X_B(Z) = \delta X_B(Z) = 1$. For each $z \in Z$ and $x \in \Theta_B(z)$, we have $1 = I(R_B(z,x), x) = I(1, x)$ and hence $x = 1$. That is $x \in X$ and $\Theta_B(z) \subseteq X$. Thus $Z \subseteq X_B$. On the other hand, for each $z \in U - Z$, by $S_{x \in \Theta_B(z)} T(R_B(z,x), 1-x) = 1$ it follows that there exists $x \in \Theta_B(z)$ such that $T(R_B(z,x), 1-x) = 1$. Thus $x = 0$ and $x \notin X$. So we have $\Theta_B(z) \not\subset X$ and $z \notin X_B$. Consequently $X_B \subseteq Z$ and $Z = X_B$ as required.

(2) Let $Z = X^B$. For each $z \in Z$, we have $\Theta_B(z) \cap X \neq \varnothing$, where $\Theta_B(z)$ is the equivalence class containing z. Hence there exists $x \in \Theta_B(z)$ such that $x \in X$. Thus $T(R_B(z,x), 1) = 1$ and $S_{x \in \Theta_B(z)} T(R_B(z,x), x) = 1$. So $\mu X^B(Z) = T_{z \in Z} S_{x \in \Theta_B(z)} T(R_B(z,x), x) = 1$. For each $z \in U - Z$, it follows that $z \notin X^B$ and thus $\Theta_B(z) \cap X = \varnothing$, i.e. $\Theta_B(z) \subseteq \sim X$. So, for any $x \in \Theta_B(z)$, we have $x \in \sim X$. Thus $I(R_B(z,x), 1-x) = 1$ and hence $\delta X^B(Z) = T_{z \in U-Z} T_{x \in \Theta_B(z)} I(R_B(z,x), 1-x) = 1$.

So, $\gamma X^B(Z) = 1$ as required.

Conversely, assume that $\gamma X^B(Z) = 1$. It follows that $\mu X^B(Z) = T_{z \in Z} S_{x \in \Theta_B(z)} T(R_B(z,x), x) = 1$ and $\delta X^B(Z) = T_{z \in U-Z} T_{x \in \Theta_B(z)} I(R_B(z,x), 1-x) = 1$. For each $z \in Z$, there exists $x \in \Theta_B(z)$ such that $T(R_B(z,x), x) = 1$. Thus $x = 1$ and $x \in X$. So we have $\Theta_B(z) \cap X \neq \varnothing$. Thus $z \in X^B$ and $Z \subseteq X^B$. On the other hand, for each $z \notin Z$ and $x \in \Theta_B(z)$, we have $I(R_B(z,x), 1-x) = 1$. It follows that $x \notin X$ and $\Theta_B(z) \cap X = \varnothing$. So $z \notin X^B$. Thus $X^B \subseteq Z$. Consequently $X^B = Z$.

This theorem shows that Definition 1 is a generalization of Pawlak approximation operators. In what follows, we take $T(x,y) = xy$, $S(x,y) = x + y - xy$ and $I(x,y) = 1 - x + xy$.

Theorem 5 Let $IT = (U,A)$ be an incomplete information table, and $B \subseteq A$, $X, Z \subseteq U$.

(1) $\mu X_B(Z) = \prod_{z \in Z} \prod_{x \in \Theta_B(z) \cap (\sim X)} (1 - R_B(z,x))$.

(2) $\mu X^B(Z) = \prod_{z \in Z} (1 - \prod_{x \in \Theta_B(z) \cap X} (1 - R_B(z,x)))$.

(3) $\delta X_B(Z) = \prod_{z \in U - Z} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X)} (1 - R_B(z,x)))$.

(4) $\delta X^B(Z) = \prod_{z \in U - Z} \prod_{x \in \Theta_B(z) \cap X} (1 - R_B(z,x))$.

Proof: (2) For each $z \in Z$, and $x \in \Theta_B(z)$, $x \in X$ implies $T(R_B(z,x),x) = R_B(z,x)$ and $x \in U - X$ implies $T(R_B(z,x),x) = 0$. Thus

$S_{x \in \Theta_B(z)} T(R_B(z,x),x) = S_{x \in \Theta_B(z) \cap X} R_B(z,x) = 1 - \prod_{x \in \Theta_B(z) \cap X} (1 - R_B(z,x))$.

Consequently, we have

$\mu X^B(Z) = \prod_{z \in Z} (1 - \prod_{x \in \Theta_B(z) \cap X} (1 - R_B(z,x)))$.

(1), (3) and (4) can be proved similarly.

Corollary 1 Let $IT = (U, A)$ be an incomplete information table, and $B \subseteq A$, $X, Z \subseteq U$.

(1) $\mu X_B(Z) = \delta(\sim X)^B(\sim Z)$.

(2) $\mu X^B(Z) = \delta(\sim X)_B(\sim Z)$.

Theorem 6 Let $IT = (U, A)$ be an incomplete information table.

(1)If $Z_1 \subseteq Z_2$, then $\delta X_B(Z_1) \leq \delta X_B(Z_2)$, $\delta X^B(Z_1) \leq \delta X^B(Z_2)$.

(2)If $X_1 \subseteq X_2$, then $\delta X_{1B}(Z) \geq \delta X_{2B}(Z)$, $\delta X_1^B(Z) \geq \delta X_2^B(Z)$.

(3)If $B_1 \subseteq B_2$, then $\delta X_{B_1}(Z) \geq \delta X_{B_2}(Z)$, $\delta X^{B_1}(Z) \leq \delta X^{B_2}(Z)$.

Proof: (1) By Theorem 1 and Corollary 1,

$\delta X_B(Z_1) = \mu(\sim X)^B(\sim Z_1) \leq \mu(\sim X)^B(\sim Z_2) = \delta X_B(Z_2)$

$\delta X^B(Z_1) = \mu(\sim X)_B(\sim Z_1) \leq \mu(\sim X)_B(\sim Z_2) = \delta X^B(Z_2)$

(2),(3)can be proved similarly.

Theorem 7 Let $IT = (U, A)$ be an incomplete information table, and $B \subseteq A$, $X \subseteq U$.

(1) $\gamma X_B(X) = \prod_{z \in X} \prod_{x \in \Theta_B(z) \cap (\sim X)} (1 - R_B(z,x))$.

(2) $\gamma X^B(X) = \prod_{z \in \sim X} \prod_{x \in \Theta_B(z) \cap X} (1 - R_B(z,x))$.

Proof: (1) For each $z \in U - X$, we have $z \in \Theta_B(z) \cap (\sim X)$ and $1 - R_B(z,z) = 0$. Thus $\prod_{x \in \Theta_B(z) \cap (\sim X)} (1 - R_B(z,x)) = 0$ and hence

$\delta X_B(X) = \prod_{z \in U - X} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X)} (1 - R_B(z,x))) = 1$.

Consequently we have

$\gamma X_B(X) = \mu X_B(X) = \prod_{z \in X} \prod_{x \in \Theta_B(z) \cap (\sim X)} (1 - R_B(z,x))$

(2) can be proved similarly.

Theorem 8 Let $IT = (U, A)$ be an incomplete information table, and $B \subseteq A$, $X_1, X_2, Z_1, Z_2 \subseteq U$.

(1) $\mu(X_1 \cap X_2)_B(Z_1 \cap Z_2) \geq \mu X_{1B}(Z_1) \cdot \mu X_{2B}(Z_2)$.

(2) $\delta(X_1 \cap X_2)_B(Z_1 \cap Z_2) \geq \delta X_{1B}(Z_1) \cdot \delta X_{2B}(Z_2)$.

(3) $\gamma(X_1 \cap X_2)_B(Z_1 \cap Z_2) \geq \gamma X_{1B}(Z_1) \cdot \gamma X_{2B}(Z_2)$.

Proof: (1) Let $z \in Z_1 \cap Z_2$, by $\Theta_B(z) \cap (\sim X_1 \cup \sim X_2) = (\Theta_B(z) \cap \sim X_1) \cup (\Theta_B(z) \cap \sim X_2)$,

$\prod_{x \in \Theta_B(z) \cap \sim X_1} (1 - R_B(z,x)) \cdot \prod_{x \in \Theta_B(z) \cap \sim X_2} (1 - R_B(z,x))$

$= \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x)) \cdot \prod_{x \in \Theta_B(z) \cap \sim X_1 \cap X_2} (1 - R_B(z,x))$

$\leq \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x))$,

It follows that

$\mu X_{1B}(Z_1) \cdot \mu X_{2B}(Z_2)$

$= \prod_{z \in Z_1} \prod_{x \in \Theta_B(z) \cap (\sim X_1)} (1 - R_B(z,x)) \cdot \prod_{z \in Z_2} \prod_{x \in \Theta_B(z) \cap (\sim X_2)} (1 - R_B(z,x))$

$\leq \prod_{z \in Z_1 \cap Z_2} \prod_{x \in \Theta_B(z) \cap (\sim X_1)} (1 - R_B(z,x)) \cdot \prod_{z \in Z_1 \cap Z_2} \prod_{x \in \Theta_B(z) \cap (\sim X_2)} (1 - R_B(z,x))$

$= \prod_{z \in Z_1 \cap Z_2} (\prod_{x \in \Theta_B(z) \cap (\sim X_1)} (1 - R_B(z,x)) \cdot \prod_{x \in \Theta_B(z) \cap (\sim X_2)} (1 - R_B(z,x)))$

$\leq \prod_{z \in Z_1 \cap Z_2} \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x)) = \mu(X_1 \cap X_2)_B(Z_1 \cap Z_2)$

(2) For each $z \in Z_1 \cap Z_2$, by

$\delta X_{1B}(Z_1) \cdot \delta X_{2B}(Z_2)$

$= \prod_{z \in \sim Z_1} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_1)} (1 - R_B(z,x))) \cdot \prod_{z \in \sim Z_2} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_2)} (1 - R_B(z,x)))$

$\leq \prod_{z \in \sim Z_1} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x))) \cdot \prod_{z \in \sim Z_2} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x)))$

$= \prod_{z \in \sim Z_1 \cup Z_2} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x))) \cdot \prod_{z \in \sim Z_1 \cap Z_2} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x)))$

$\leq \prod_{z \in \sim Z_1 \cup Z_2} (1 - \prod_{x \in \Theta_B(z) \cap (\sim X_1 \cup X_2)} (1 - R_B(z,x))) = \delta(X_1 \cap X_2)_B(Z_1 \cap Z_2)$

(3) is straightforward from (1) and (2).

Corollary 2 Let $IT = (U, A)$ be an incomplete information table, and $B \subseteq A$, $X_1, X_2, Z_1, Z_2 \subseteq U$.

(1) $\mu(X_1 \cup X_2)^B(Z_1 \cup Z_2) \geq \mu X_1^B(Z_1) \cdot \mu X_2^B(Z_2)$.

(2) $\delta(X_1 \cup X_2)^B(Z_1 \cup Z_2) \geq \delta X_1^B(Z_1) \cdot \delta X_2^B(Z_2)$.

(3) $\gamma(X_1 \cup X_2)^B(Z_1 \cup Z_2) \geq \gamma X_1^B(Z_1) \cdot \gamma X_2^B(Z_2)$.

Proof: (1) By Theorem 8 and Corollary 1,

$\mu(X_1 \cup X_2)^B(Z_1 \cup Z_2) = \delta(\sim X_1 \cap \sim X_2)_B(\sim Z_1 \cap \sim Z_2)$

$$\geq \delta(\sim X_1)_B(\sim Z_1) \cdot \delta(\sim X_2)_B(\sim Z_2) = \mu X_1{}^B(Z_1) \cdot \mu X_2{}^B(Z_2)$$

(2) and (3) can be proved similarly.

IV. CONCLUSIONS

Rough set under incomplete information has been extensively studied. For incomplete information system, researchers have put forward several similarity relations, such as tolerance relation, non-symmetric relation, valued tolerance relation etc. Based on valued tolerance relation, we proposed a new kind of rough approximation operators which is a generalization of Pawlak approximation operators for complete information system. Some basic properties of the approximation operators are investigated. Based on this work, we can further probe the rough set model under incomplete information and its application in knowledge discovery.

REFERENCES

[1] Z.Bonikowski, E.Bryniarski, U.Wybraniec, "Extensions and intentions in the rough set theory," Information Sciences, vol. 107, pp. 149-167, 1998.

[2] J.W.Grzymala-Busse, W.Grzymala-Busse, L.K.Goodwin, A closest fit approach to missing attribute values, Proc. 7th Workshop New Direction in Rough Sets, Data Mining, and Granular-Soft Computing Yamaguchi, Japan, Springer Verlag LNAI 1711, 1999, pp. 405-414.

[3] J.W.Grzymala-Busse, M.Hu, A comparison of several approaches to missing attribute values in data mining, Proceedings of the 2nd Int. Conference on Rough Sets and New Trends in Computing, Banff 2000, pp. 340-347.

[4] L.H.Guan, G.Y.Wang, Generalized approximations defined by non-equivalence relations, Information Sciences, vol. 193, pp. 163-179, 2012.

[5] M.Kryszkiewicz, Rough set approach to incomplete information system, Information Sciences, vol. 112, pp. 39-49, 1998.

[6] M.Kryszkiewicz, Properties of incomplete information systems in the framework of rough sets, Rough Sets in Data Mining and Knowledge Discovery, Physica-Verlag, 1998, pp. 422-450.

[7] Z.Pawlak, Rough sets, Int. J. Computer and Information Sci., vol. 11, pp. 341-356, 1982.

[8] Z.Pawlak, A.Skowron, Rough sets: Some extensions, Information Sciences, vol. 177, pp. 28-40, 2007.

[9] K.Qin, Z.Pei, J.Yang, Y.Xu, Approximation operators on complete completely distributive lattices, Information Sciences, vol. 247, pp. 123-130, 2013

[10] A.M.Radzikowska, E.E.Kerre, A comparative study of fuzzy rough sets, Fuzzy Sets and Systems, vol. 126, pp.137-155, 2002

[11] R.Slowinski, J.Stefanowski, Rough classification in incomplete information systems, Math. Computing Modeling, 12 (10/11), 1989, pp. 1347-1357.

[12] J.Stefanowski, A.Tsoukias, Incomplete information tables and rough classification, Computational Intelligence, 17(3)2001, pp. 545-566

[13] G.Y.Wang, Extension of rough set under incomplete information systems, Journal of Computer Research and Development (in Chinese), vol. 39, pp. 1238-1243, 2002.

[14] Y.Y.Yao, Relational interpretation of neighborhood operators and rough set approximation operator, Information Sciences, 111, pp.239-259, 1998.

[15] X.Yin, X.Jia, L.Shang, A new extension model of rough sets under incomplete information, Lecture Notes in Artificial Intelligence, 4062, pp. 141-146, 2006.

[16] X.H.Zhang, B.Zhou, P.Li, A general frame for intuitionistic fuzzy rough sets, Information Sciences, 216, pp.34-49, 2012.

Incremental Granular Modeling for Predicting the Hydrodynamic Performance of Sailing Yachts

Keun-Chang Kwak

Dept. of Control and Instrumentation Engineering
Chosun University, 375 Seosuk-Dong
Gwangju, Korea

Abstract—This paper is concerned with a design method for modeling Incremental Granular Model (IGM) based on Linguistic Model (LM) and Polynomial Regression (PR) from data set obtained by complex yacht hydrodynamics. For this purpose, we develop a systematic approach to generating automatic fuzzy rules based on Context-based Fuzzy C-Means (CFCM) clustering. This clustering algorithm builds information granules in the form of linguistic contexts and estimates the cluster centers by preserving the homogeneity of the clustered data points associated with the input and output space. Furthermore, IGM deals with localized nonlinearities of the complex system so that the modeling discrepancy can be compensated. After performing the design of 2nd order PR as the first global model, we refined it through a series of local fuzzy if-then rules in order to capture the remaining localized characteristics. The experimental results revealed that the presented IGM showed a better performance in comparison to the previous works for predicting the hydrodynamic performance of sailing yachts.

Keywords—granular networks; particle swarm optimization; linguistic model; two-sided Gaussian contexts

I. INTRODUCTION

During the last decade, it is advantageous to combine several computing techniques synergistically, rather than exclusively, resulting in the construction of complementary hybrid intelligent systems in confronting real-world application domains such as modeling, control, and optimization for complex engineering problems [1]. Among these approaches, the neuro-fuzzy and soft computing as a computational approach to learning and machine intelligence is a widely used computing framework based on the concepts of fuzzy model, neural networks, data clustering, and several stochastic optimization methods for computational intelligence. Here, neural networks provide learning abilities and a connectionist structure to fuzzy systems. Fuzzy systems provide neural networks with a structural framework with human knowledge and reasoning. A considerable number of these studies have been conducted to generate automatic fuzzy if-then rules using clustering techniques from given numerical training data sets [2-4]. On the other hand, Linguistic Model (LM) has been researching to present a nonlinear and complex characteristic based on Context-based Fuzzy C-Means (CFCM) clustering [5]. In contrast to the context-free clustering methods such as FCM clustering [6], subtractive clustering, and mountain clustering, the CFCM clustering is to generate clusters preserving homogeneity of the clustered patterns in connection with their similarity in the input and output variables [7]. The studies associated with this clustering approach are LM [5], RBFN(Radial Basis Function Networks)-CFCM [8], LM with learning [9], TSK(Takagi-Sugeno-Kang)-LM [10], and Incremental Model (IM) [11]. In this paper, we develop the Incremental Granular Model (IGM) for predicting complex hydrodynamic performance of sailing yacht. The presented IGM is based on LM and Polynomial Regression (PR) to capture the localized nonlinear characteristics. First, we build a PR which could be treated as a preliminary construct. Next, all modeling discrepancies are compensated by a collection of rules that become attached to the regions of the input space where the error is localized. The experiments are achieved by the data set obtained from the complex yacht hydrodynamics. The prediction of resistance of the ship at the initial design stage is of a great value for evaluating the ship's performance and for estimating the required propulsive power [12]. We compared the effectiveness of the presented IGM with the previous works such as RBFN, LM, RBFN-CFCM, and IM. This paper is organized as follows. In Section 2, we describe the architecture of CFCM clustering and LM as a framework of user-centric system modeling. In the Section 3, we present the proposed IGM based on LM and PR. In Section 4, we present the prediction problem of the hydrodynamic performance of sailing yachts [13] and the experimental results. Finally, concluding comments are given in Section 5.

II. CFCM CLUSTERING AND LINGUISTIC MODEL

The CFCM clustering as an interesting variant of the fuzzy c-means is realized via individual contexts as shown in Fig. 1. Each linguistic context has defined semantics that can be interpreted as a *large* negative error, *medium* negative error, etc in the design of IGM.

Let us consider a certain fixed context W_j described by some membership function. The data point in the output space is associated with the corresponding membership value. Let us introduce a family of the partition matrices induced by the *l*-th context and denote it by $\mathbf{U}(W_l)$

$$\mathbf{U} = \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^{c} u_{ik} = w_{lk} \ \forall k \text{ and } 0 < \sum_{k=1}^{N} u_{ik} < N \right\}$$

(1)

where w_{lk} denotes a membership value of the *k*-th datum implied by the *l*-th context. The underlying objective function is as follows

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \, || \, \mathbf{x}_k - \mathbf{v}_i \, ||^2 \qquad (2)$$

where $\mathbf{v}_i$ denotes the *i*-th cluster center. The J is minimized under the constraints imposed by (1) as follows

Min Q subject to $\mathbf{U}(\mathbf{W}_l)$, *l*=1, 2, …,p (3)

The minimization of J is realized by iteratively updating the values of the partition matrix and the cluster centers. The successive updates of the partition matrix are completed as follows

$$u_{ik} = \frac{w_{1k}}{\sum_{j=1}^{c} \left(\dfrac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{m-1}}} \qquad (4)$$

where u_{ik} is the partition matrix induced by the *l*-th context. The cluster centers are as the following expression

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} \mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^{m}} \qquad (5)$$
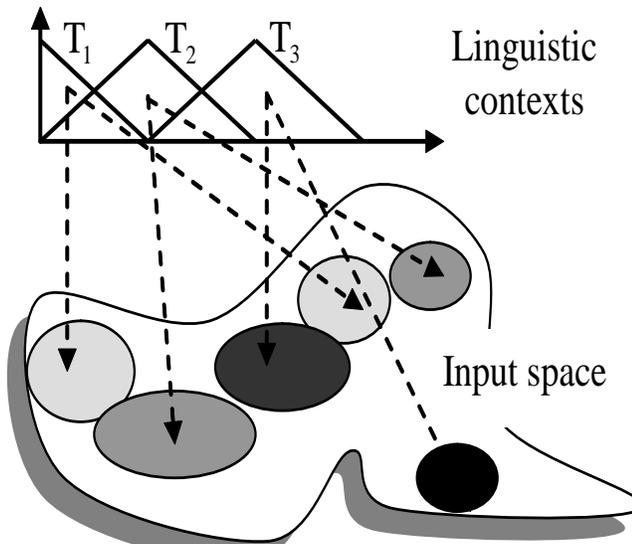
Fig. 1.   Concept of context-based fuzzy clustering

In the design of the LM, we consider the contexts to be described by triangular membership functions being equally distributed in the error space **E** with the 1/2 overlap occurring between two successive fuzzy sets. Each context generates a number of induced clusters whose activation levels are afterwards summed up as shown in Fig. 2.
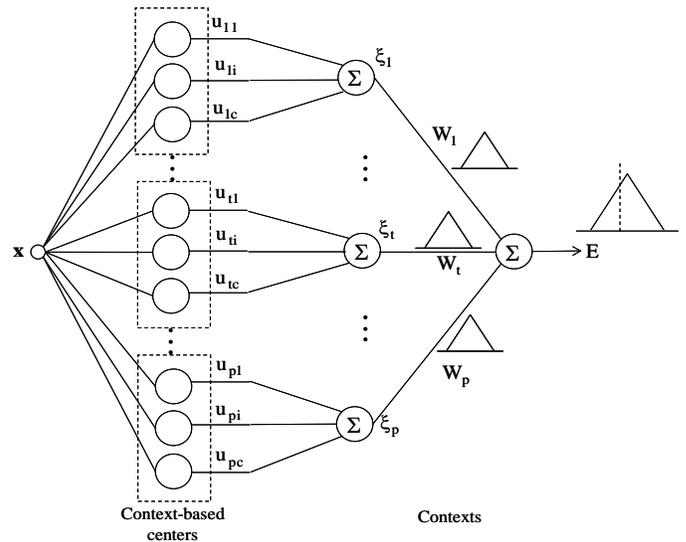
Fig. 2.   Architecture of the LM

Assuming the triangular form of the contexts, the result is a triangular fuzzy number E as follows

$$E = W_1 \otimes \xi_1 \oplus W_2 \otimes \xi_2 \oplus \dots W_n \otimes \xi_n \qquad (6)$$

We denote the algebraic operations by $\otimes, \oplus$ to emphasize that the underlying computing operates on a collection of fuzzy numbers. As such, E is characterized by its three parameters that are a modal value, the lower bound, and upper bound.

III.   INCREMENTAL GRANULAR MODEL (IGM)

The PSO method is one of swarm intelligence methods for solving the optimization problems. The PSO algorithm proposed by Kennedy is performed by social behavior of bird flocking or fish schooling. The character of PSO easily can handle fitness function for solving complex problems. Furthermore, it can control a relationship between global and local search. Here, each particle adjusts information of location with experience of them and their neighborhood. It can form the answer of optimum in short time. As the velocity of particle movement of PSO is only demanded, it is easy to be embodiment and brevity of a theory. The basic element of PSO is simply as follows IGM performs localized nonlinearities of the complex and nonlinear system so that the modeling discrepancy can be compensated. After performing the design of 2nd order Polynomial Regression (PR) as the first global model, we refined it through a series of local fuzzy if-then rules in order to capture the remaining localized characteristics. Fig. 3 shows the main design process of the IGM. Firstly, we decide upon the granularity of information to be used in the development of the model such as the number of contexts and the number of clusters formed for each context. The design procedure of IGM is as follows

[Step 1] Design PR in the input and output space. PR is the extended form of the well-known Linear Regression (LR) in which the relationship between the independent variables $\mathbf{x}_k$ and dependent variable y is modeled as an 2nd order polynomial. PR fits a nonlinear relationship between the value of $\mathbf{x}_k$ and the corresponding conditional mean of y. On the basis of the original data set, a collection of input-error pairs, $(\mathbf{x}_k, e_k)$ is obtained.

[Step 2] Produce linguistic contexts in the error space of the regression model $E_1, E_2, ..., E_p$. The distribution of these fuzzy sets is obtained through the use of statistical distribution or fuzzy equalization while the fuzzy sets are characterized by triangular membership functions with a 0.5 overlap between neighboring fuzzy sets.

[Step 3] Perform CFCM clustering in the input-output space from the linguistic contexts produced in the error space.

[Step 4] Compute the activation levels of the clusters induced by the corresponding contexts and their overall aggregation through weighting by fuzzy sets of the context leading to the triangular fuzzy number of output, $E = F(\mathbf{x}; E_1, E_2, ..., E_p)$.

[Step 5] The output of the IGM is then combined with the output of the linear part. The result is a shifted triangular number $Y = z \oplus E$.
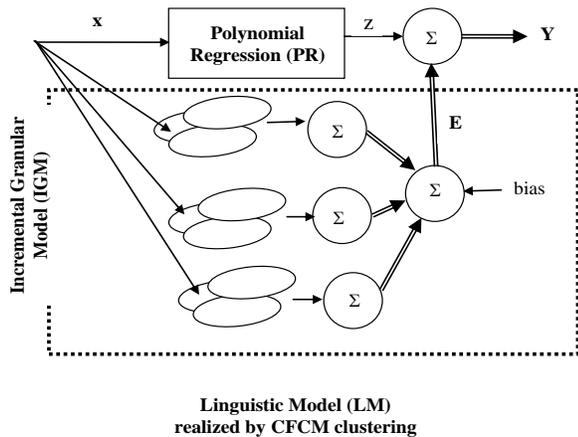

Fig. 3.   Main design process of IGM based on PR and LM

IV.   EXPERIMENTAL RESULTS

In this experiment, we use to predict hydrodynamic performance of sailing yachts from dimensions and velocity. The well-known Holtrop and Mennen method [14] is widely used as the initial design stage of ships for estimating the resistance of the ships. This method provides a prediction of the total resistance's components. The total resistance of a ship has been subdivided as follows

$$R_{total} = R_F(1+k_1) + R_{APP} + R_W + R_B + R_{TR} + R_A \qquad (7)$$

R_F : Frictional resistance according to the ITTC-1957 friction formula

$(1+k_1)$: Form factor describing the viscous resistance of the full form in relation to R_F

R_{APP} : Resistance of appendages

R_W : Wave-making and wave-breaking resistance

R_B : Additional pressure resistance of bulbous bow near the water surface

R_{TR} : Additional pressure resistance of immersed transform stern

R_A : Model-ship correlation resistance.

The form factor of the hull the prediction formula is as the following equation

$$(1+k_1) = c_{13}\{0.93 + c_{12}(B/L_R)^{0.92497}(0.95 - C_p)^{-0.521448} \cdot$$
$$(1-C_p + 0.0225lcb)^{0.6096}\}$$

$$(8)$$

Here, R_B is the prismatic coefficient based on the waterline length L and lcb is the longitudinal position of the centre of buoyancy forward of 0.5L as a percentage of L. L_R is a parameter reflecting the length of the run. In this example, six input variables are composed of longitudinal position of the center of buoyancy, prismatic coefficient, length-displacement ratio, beam-draught ratio, length-beam ratio, and Froude number as shown in Fig. 4. The output variable to be predicted by the six input variables is residuary resistance per unit weight of displacement. The overall data set consists of 308 full experiments, which were performed at the Delft Ship Hydromechanics laboratory. These experiments include 22 different hull forms.
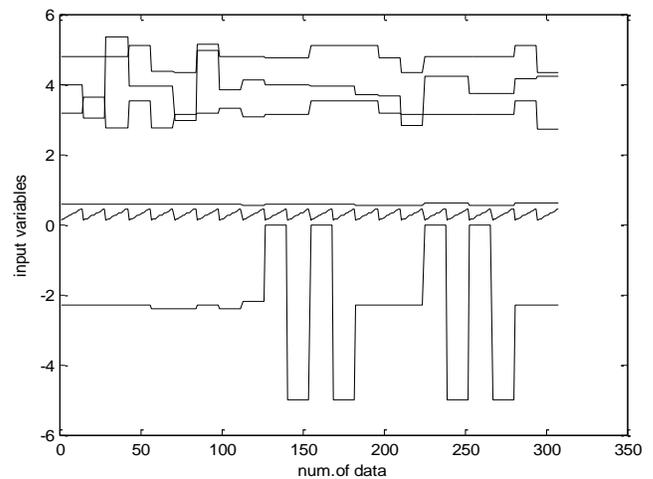
Fig. 4.   Six input variables

We randomly divide the data set into training and test data with 60%-40% ratio in the normalized input space between 0 and 1, respectively. The experiments perform 10 runs. The training data set is used for model construction, while the test set is used for model validation. Thus, the resultant model is not biased toward the training data set and it is likely to have a better generalization performance with respect to new data. Firstly, 2nd order PR is performed in the input and output space.

After that, the regression error is obtained as the histogram shown in Fig.5. Fig. 6 visualizes six linguistic contexts to perform CFCM clustering. These contexts are produced by the use of statistical distribution. Here, we assume that the number of cluster per each context is the same.
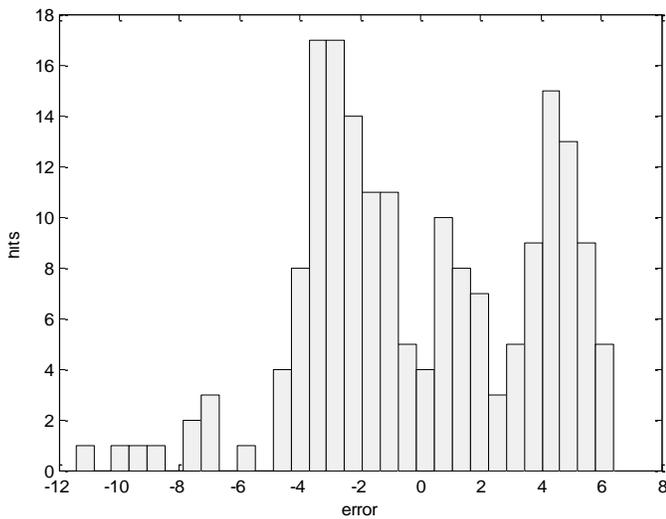
Fig. 5.   Histogram of error obtained by PR
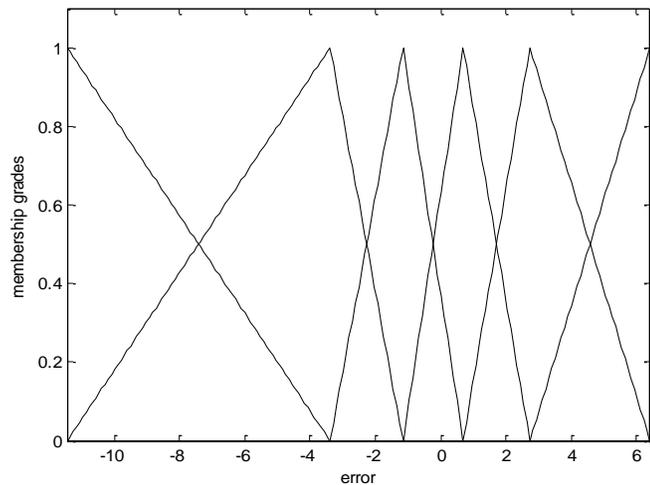
Fig. 6.   Six linguistic contexts obtained in the error space

Fig. 7 shows comparison results between the desired and model output for both the training and test data sets, respectively. As shown in Fig. 7, it is obvious that the proposed IGM has good approximation and generalization performance. Table 1 lists the mean of RMSE and the number of rules for the training and test sets, respectively. In the design of IGM, we obtained the best results in three contexts and two clusters in each context for CFCM clustering as listed in Table 1. Although the conventional LM has a structured knowledge representation in the form of fuzzy if-then rules, it lacks the adaptability to deal with a complex and nonlinear model. Moreover, we obtained each best RMSE result through the construction of RBFN, LM, RBFN-CFCM, and IM in trial and error as listed in Table. 1.
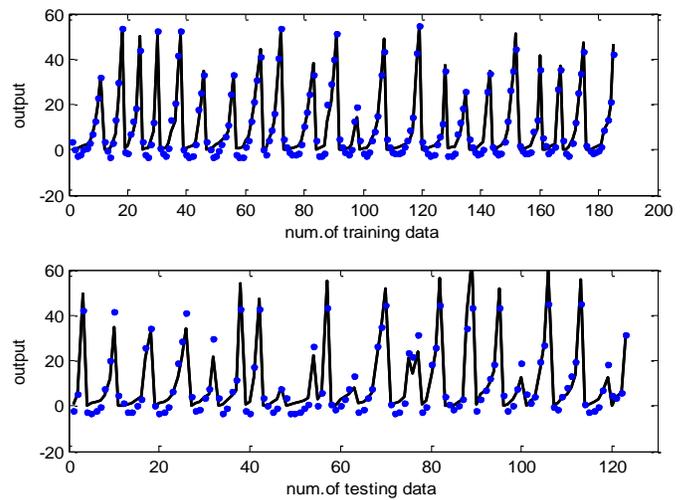
Fig. 7.   Approximation and generalization performance for training and test data set (solid line: actual output, dotted line: model output)

The augmented granular modification of the model was realized by experimenting with the two essential parameters controlling the granularity of the construct in the input and output space, that is "p" and "c". The corresponding results are summarized in Table 2 and 3. Fig. 8 visualizes the uncertain prediction values from upper and lower output obtained from the proposed IGM. As a consequence, the experimental results revealed that the presented IGM yielded a better performance in comparison to RBFN, LM, RBFN-CFCM, and IM for predicting the hydrodynamic performance of sailing yachts as shown in Table 1 and Fig. 8.

TABLE I.   COMPARISON RESULTS OF THE RMSE AND THE NUMBER OF RULES

Methods	no. rule	RMSE (training)	RMSE (test)
Linear regression	-	8.6690	9.3006
RBFN	6*	13.6393	14.9378
LM [5]	6(p=3,c=2)	10.3391	11.7420
RBFN-CFCM [8]	10*(p=5,c=2)	8.1735	9.7952
IM [11]	12(p=6,c=2)	6.4082	7.8182
The proposed IGM	6(p=3,c=2)	3.9333	4.4344
	8(p=4, c=2)	4.0367	4.5022
	10(p=5,c=2)	4.0030	4.5290
	12(p=6,c=2)	3.9298	4.4811

TABLE II.   RMSE (MEAN) - TRAINING DATA

| | | No. of contexts (p) | | | |
		3	4	5	6
No. of clusters per context (c)	2	3.9333	4.0367	4.0030	3.9298
	3	3.9599	3.9589	3.9546	3.8690
	4	3.9428	3.8716	3.8399	3.7762
	5	3.8243	3.8109	3.7325	3.7190
	6	3.8064	2.7643	3.6775	3.6542

TABLE III.   RMSE (MEAN) – TESTING DATA

| | | No. of contexts (p) | | | |
		3	4	5	6
No. of clusters per context (c)	2	4.4344	4.5022	4.5290	4.4811
	3	4.5054	4.5114	4.5705	4.5190
	4	4.5435	4.5418	4.5965	4.5835
	5	4.5182	4.5858	4.5875	4.6137
	6	4.5814	4.6392	4.5997	4.5943

V.   CONCLUSIONS

We developed the incremental granular model with uncertainty output based on polynomial regression and linguistic model realized by the context-based fuzzy c-means clustering. Furthermore, we dealt with an incremental model to deal with localized nonlinearities of the system so that all modeling discrepancies can be compensated. This incremental model is quite different from one frequently used in conjunction with fuzzy modeling with the predominant concept of a rule-based architecture. The experimental results on complex yacht hydrodynamics revealed that the presented method showed a good approximation and generalization ability in comparison to conventional other methods. Thus, these results lead to the conclusion that incremental granular model can be represented as the prototypes that exhibits certain characteristics of the complex system to be modeled.

REFERENCES

[1] J. S. R. Jang, C. T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall, 1997.

[2] K. C. Kwak and D. H. Kim "Adaptive Neuro-Fuzzy Networks with the Aid of Fuzzy Granulation", *IEICE Trans. on Information and Systems*, Vol. E88-D, No. 9, pp.2189-2196, 2005.

[3] J. Abonyi, R. Babuska, and F. Szeifert, "Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models", *IEEE Trans. on Systems, Man, and Cybernetics-Part B*, Vol.32, No.5, pp. 612-621, 2002.

[4] K. C. Kwak, M. G. Chun, J. W. Ryu, and T. H. Han, "FCM-based adaptive fuzzy inference system for coagulant dosing process in a water purification plant", *International Journal of Knowledge-Based Intelligent Engineering Systems*, Vol. 4, No. 4, pp.230-236, 2000.

[5] W. Pedrycz and A. V. Vasilakos, "Linguistic models and linguistic modeling", *IEEE Trans. on Systems, Man, and Cybernetics-Part C*, Vol. 29, No. 6, pp.745-757, 1999.

[6] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
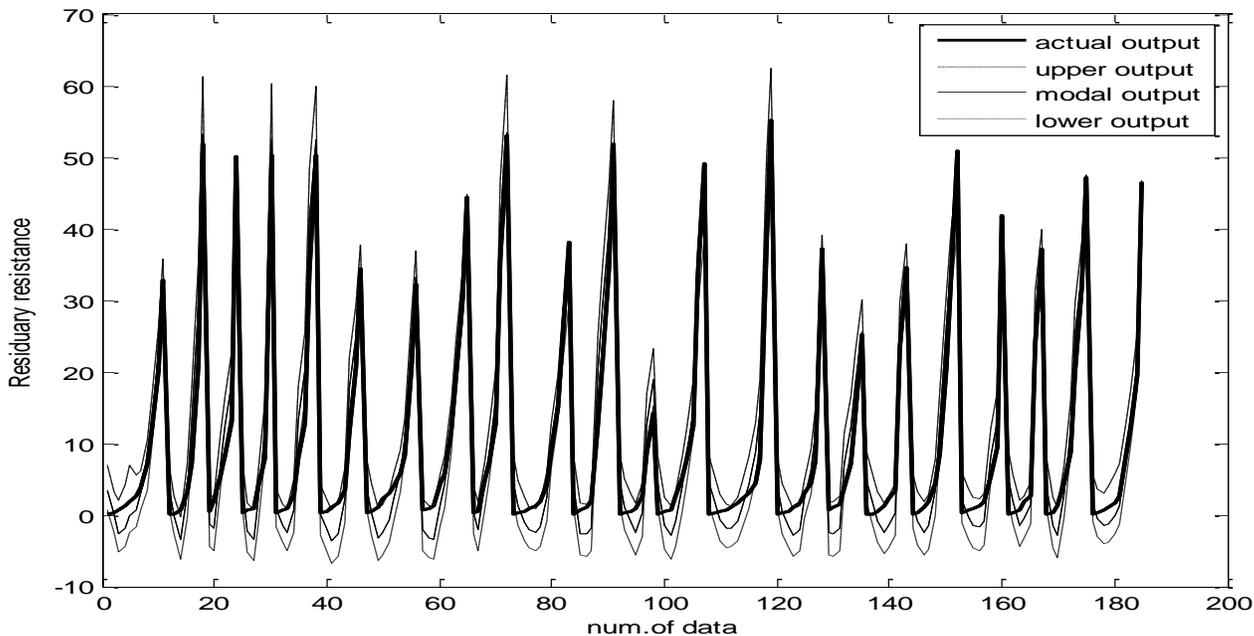
Fig. 8.   Prediction performance with uncertain output

[7]   W. Pedrycz, "Conditional fuzzy C-Means", *Pattern Recognition Letters*, Vol. 17, pp. 625-632, 1996.

[8]   W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks", *IEEE Trans. on Neural Networks*, Vol. 9, No. 4, pp.745-757, 1999.

[9]   W. Pedrycz and K. C. Kwak, "Linguistic models as a framework of user-centric system modeling", *IEEE Trans. on Systems, Man, and Cybernetics-Part A,* Vol. 36, No. 4, pp.727-745, 2006.

[10] K. C. Kwak and D. H. Kim, "TSK-based linguistic fuzzy model with uncertain model output", *IEICE Trans. on Information and Systems*, Vol. E89-D, No. 12, pp.2919-2923, 2006.

[11] W. Pedrycz and K. C. Kwak, "The development of incremental models", *IEEE Trans. on Fuzzy Systems*, Vol. 15, No. 3, pp. 507-518, 2007.

[12] I. Ortigosa, R. Lopez and J. Garcia. "A neural networks approach to residuary resistance of sailing yachts prediction", In Proceedings of the International Conference on Marine Engineering MARINE 2007, 2007.

[13] UCI   Machine   Learning   Repository, http://archive.ics.uci.edu/ml/index.html

[14] J. Holtrop and G.G.J. Mennen, "An approximate power prediction method", International Shipbuilding Progress, pp. 166-170, 1982.

AUTHOR PROFILE

Keun-Chang Kwak received the B.Sc., M.Sc., and Ph.D. degrees from Chungbuk National University, Cheongju, Korea, in 1996, 1998, and 2002, respectively. During 2003–2005, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. From 2005 to 2007, he was a Senior Researcher with the Human–Robot Interaction Team, Intelligent Robot Division, Electronics and Telecommunications Research Institute, Daejeon, Korea. He is currently the Assistant Professor with the Department of Control, Instrumentation, and Robot Engineering, Chosun University, Gwangju, Korea. His research interests include human–robot interaction, computational intelligence, biometrics, and pattern recognition. Dr. Kwak is a member of IEEE, IEICE, KFIS, KRS, ICROS, KIPS, and IEEK..

What is the Right Illumination Normalization for Face Recognition?

Aishat Mahmoud Dan-ali

Department of Computer Science and Engineering
The American University in Cairo
AUC Avenue, P.O. Box 74, New Cairo 11835, Egypt

Mohamed Moustafa

Department of Computer Science and Engineering
The American University in Cairo
AUC Avenue, P.O. Box 74, New Cairo 11835, Egypt

Abstract—In this paper, we investigate the effect of some illumination normalization techniques on a simple linear subspace face recognition model using two distance metrics on three challenging, yet interesting databases. The research takes the form of experimentation and analysis in which five illumination normalization techniques were compared and analyzed using two different distance metrics. The performances and execution times of the various techniques were recorded and measured for accuracy and efficiency. The illumination normalization techniques were Gamma Intensity Correction (GIC), discrete Cosine Transform (DCT), Histogram Remapping using Normal distribution (HRN), Histogram Remapping using Log-normal distribution (HRL), and Anisotropic Smoothing technique (AS). Results showed that improved recognition rate was obtained when the right preprocessing method is applied to the appropriate database using the right classifier.

Keywords—face recognition; preprocessing; illumination

I. INTRODUCTION

Illumination and pose challenges have been the serious bottlenecks in face recognition algorithms. Many attempts were made to overcome the effects of these challenges; however, the perfect face recognition system invariant to all challenges is still elusive. What researchers concentrate on is trying to overcome a couple of challenges at a time.

This paper is mainly focused on studying how a face recognition system is being affected by illumination variation. Attention is concentrated on the preprocessing part of the system. Different image preprocessing techniques for face recognition were proposed and experimented with. The sequence of execution of the proposed method includes the preprocessing step, PCA/LDA subspace [12], and cosine/Euclidean classifiers.

To facilitate a comprehensive study and analysis, five different preprocessing techniques were implemented on the PCA/LDA model using two different classifiers. These preprocessing techniques were carefully chosen based on their popularity and recorded success.

This yielded ten (10) set of experimentations on each of the three databases used. The preprocessing techniques are Gamma Intensity Correction (GIC), Discrete Cosine Transform (DCT), Histogram remapping with normal distribution (HRN), Histogram remapping with log-normal distribution (HRL) and Anisotropic smoothing technique (AS).

II. PREPROCESSING METHODS FOR FACE RECOGNITION

Preprocessing plays a vital role in face recognition systems, because it always tries to bring the test images and those in the database into a normalized canonical form. The use of preprocessing in face recognition is generally used to overcome the effect of lighting, enhancing image contrast and normalizing the image in terms of rotation and scale. Fig. 1 shows an example of how different illumination conditions of the same person make face recognition a difficult task, even for humans.

Fig. 1. Example of images from the Yale B database with different illumination conditions

We have seen recently a comparison of some preprocessing steps in combination with some face matching methods reported in [13]. In this paper, we are focusing on other set of preprocessing methods combined with a simple matcher based on cosine or Euclidean distances.

The preprocessing methods used in this work are highlighted below

A. Gamma Intensity Correction (GIC)

Gamma intensity correction is used to control the overall brightness of an image by changing the gamma parameter and it can be used to correct the lighting variations in the face image [7]. The gamma correction is the process of taking the exponential of the input image. The output image would be darker or brighter depending on the value of gamma γ. In this work a value of gamma = 0.2 has been used. Gamma correction has been used in [2] and [7] for illumination normalization.

B. Discrete Cosine Transform (DCT)

The Discrete Cosine Transform is a novel approach for illumination normalization under varying lighting conditions used in face recognition algorithms that keeps facial features intact while removing excess lighting variations [1], [6]. It is a popular technique for image compression in which low frequency DCT coefficients that are correlated with illumination variations are curtailed thereby significantly

reducing the effect of illumination variation. Example of application of DCT is in JPEG image compression.

C. Histogram Remapping techniques

This is a variation of the histogram equalization (HE) method in which other distributions are used instead of the default uniform distribution that is being used in the HE method. These distributions include normal, log-normal and exponential distribution. To investigate the possibilities of these distributions experiments were conducted using the Normal distribution and Log-normal distribution in the histogram remapping algorithm as suggested by this paper [9].

D. Anisotropic Smoothing

This technique is based on the reflection perception model. This work was pioneered by Gross and Brajovic [3] in which they found an estimate of the luminance L(x, y) such that reflectance R(x, y) is produced by dividing the input image I(x, y) by L(x, y). These ensure that the local contrast is suitably improved. They obtained this by enforcing a smoothing constraint on the algorithm.

III. EXPERIMENTS

In this section, experiments were carried out that illustrate the effectiveness of the proposed method using three publicly available face databases with considerable illumination variations, the databases are: CAS PEAL database [8]; Extended Yale Face Database B ('Extended Yale-B') [10]; AT&T Database [4]. The sequence of execution includes preprocessing in conjunction with the linear subspace model, and classification using either of the classifiers. The standard protocol used in evaluating the result is also given.

A. Experimental Setup

In these experiments only frontal face views were used in the experiment, but lighting, expression and identity may all vary. All of the images in the CAS PEAL-R1 lighting subset, and all the images were geometrically normalized before preprocessing. While for the AT&T database, the images were 8 bit gray-scale images containing hair and ear regions. All the images in the databases were resized to 100 x 100 pixels.

For the testing, analysis and evaluation of the proposed method, a toolbox containing MATLAB scripts named "The PhD face recognition toolbox" was partially used. The tool was made publicly available free of charge by V. Struct [5]. Another collection of MATLAB face recognition files called "FaceRecEvaluator" was also utilized. The FaceRecEvaluator was developed by Brian and Enrique [11], and is available freely for use in academic and research domains.

B. Results

Here, analysis and discussion of the various performances i.e. recognition accuracy and execution time of the techniques employed are highlighted. Five (5) preprocessing techniques and two (2) distance metrics were experimented with using these databases; the preprocessing techniques are Gamma intensity correction (GIC), Discrete Cosine Transform (DCT), Histogram remapping using Normal distribution (HRN), Histogram remapping using Log-normal distribution (HRL), and Anisotropic Smoothing technique (AS). The distance

metrics are Euclidean (EUC) and cosine (COS) distance metrics. In total we have, ten (10) set of experiments using these databases namely: (GIC+EUC), (GIC+COS), (DCT+EUC), (DCT+COS), (HRN+EUC), (HRN+COS), (HRL+EUC), (HRL+COS), (AS+EUC), (AS+COS).

The analysis and discussion is structured according to each of the three databases CAS PEAL, Extended Yale B, and ATT used and the five illumination normalization techniques with the two different distance metrics. This produces 60 different set of experiments.

C. CAS PEAL Database

The CAS PEAL contains six (6) different subsets in the frontal category. Ten (10) images of 77 subjects making a total of 770 images were used from the Lighting subset. Ten (10) experiments were conducted using this database and the result is as follows:

As shown in Fig. 2, Gamma Intensity correction together with Euclidean distance yielded the second best performance with 64.9% recognition rate in 3.2ms/img, while Gamma Intensity correction together with Cosine distance gave 62.7% recognition rate in 16.5ms/img. However, Discrete Cosine Transform in either case i.e., with Euclidean or Cosine distance metric gave the lowest performance for this database. DCT with Euclidean distance produces 10.4%, while DCT with Cosine distance produces 17.2% in 3.1ms/img and 15.1ms/img respectively. Histogram remapping using normal distribution with Euclidean distance gave 63% recognition accuracy in 3.3 ms/img while Histogram remapping using normal distribution with cosine distance yielded the best performance for this database with 65.6% recognition rate in 15.8 ms/img. However, the only minor setback of this composition (HRN + Cos) is the execution time, which is slower than the (HRN + Euc) combination.
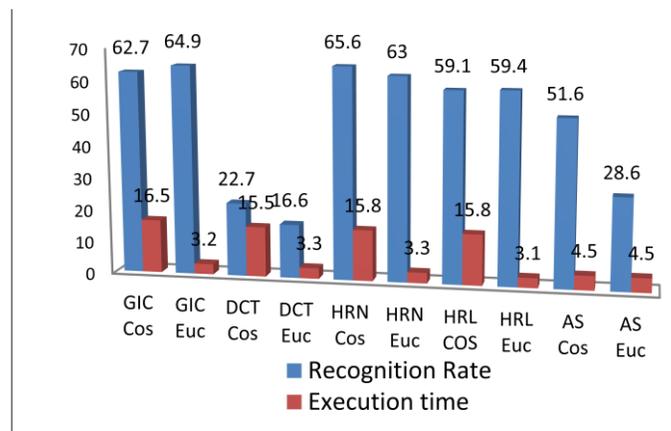
Fig. 2. Performances of different preprocessing techniques on the CAS PEAL database using two distance metrics

On the other hand, histogram remapping using Log-normal distribution with cosine distance generates 59.1% recognition accuracy in 15.3ms/img, while this method together with the Euclidean distance produces 59.4 in 3.1ms/img. Is it important to note that using this technique the Euclidean distance gives better performance in terms of accuracy and execution time.

The last experiment conducted with this database is the Anisotropic Smoothing (AS) method. This technique was also applied using the Euclidean and Cosine distance metrics. The first experiment with Euclidean distance measure yielded an accuracy of 51.6% in 4.5 ms/img, while the second experiment produced 28.6% in 4.5 ms/img with the cosine distance measure.

D. Extended Yale B database

As noted earlier the Extended Yale database is divided into subsets as initially suggested by the authors of the database. These experiments were conducted on subset 2 to subset 5 of the database and the results were analyzed separately.

1) Yale B subset 2 result: This subset is not very challenging, therefore; all of the techniques except DCT yielded 100% recognition accuracy with varying execution times, while DCT produce 96.7% in both cases.

2) Yale B subset 3 result: This subset contains images that have more illumination variation than the previous subset. The GIC plus Euclidean and the GIC plus the cosine distance generates 100% recognition rates each in 1ms/img and 2.1ms/img respectively. A recognition rate of 83.6% in 2ms/img was obtained from the DCT method using the cosine distance, while 78.9% was obtained from this subset using DCT + Euc combination in 1ms/img. In the histogram remapping technique using normal distribution, a recognition rate of 100% was also obtained in both cases using the two distance metrics. The execution times differ with 0.9ms/img, with Euclidean distance having 1.1ms/img and cosine distance having 2ms/img. Histogram remapping using log-normal distribution yielded 100% recognition rate in 1.9ms/img using cosine distance and same method using Euclidean distance gave same result but in 1ms/img. The Anisotropic smoothing technique is the final experiment carried-out using this subset. The result obtained showed a recognition rate of 99.3% in 1.5ms/img and 98% in 1.1ms/img using the cosine and Euclidean distance metrics.

3) Yale B subset 4 result: The fourth subset of this database contains images with second highest degree of illumination variation. As shown in Fig. 3, GIC + Cos yielded a recognition rate of 92.1% in 2.1ms/img, while GIC + Euc 94.7% in 1.1ms/img. It can be realized here also that the Euclidean distance gave the best performance. In DCT experiments, DCT + Cos produces 64.5% in 2 ms/img while, DCT + Euc generates 51.3 % in 0.9 ms/img. Analyzing this and the previous result under this technique closely, one can see that the DCT method is best matched with the cosine distance measure. On the other hand, experiments with histogram remapping using the normal distribution (HRN + Euc) on this subset produce 80.3% in 0.9 ms/img using Euclidean distance and (HRN + Cos) produces 78.3 % in 2 ms/img using cosine distance. Similar results were obtained using this remapping technique with lognormal distribution. Experiments with cosine distance produce 78.3 % in 2 ms/img, while that of Euclidean distance produce 78.9% in 1 ms/img. Between these two methods the HRN + Euc gives the

best result. The last experiment on this subset is the anisotropic smoothing method. The AS + cos technique gives 82.9 % in 1.5 ms/img while the AS + Euc gives 86.5 % in 1.3 ms/img. It can be noted that the anisotropic technique (AS) perform better than the three methods on this subset; that is the DCT, HRN and HLN. However the gamma intensity correction (GIC) out perform all the techniques on this subset.

Fig. 3. Performances of different preprocessing techniques on subset 4 of the Yale B database using two distance metrics

4) Yale B subset 5 result: This subset is the last and most complex of the extended Yale B database. As depicted in Fig. 4, GIC with cosine distance produces 95.4% recognition rate in 2 ms/img whereas GIC with Euclidean distance produces the same result in 1 ms/img. The next experiment is the DCT with the cosine distance measure which gave a recognition accuracy of 59.2% in 2 ms/img while the DCT + euc drops performance by generating only 50% recognition rate in 1ms/img. It is relevant to point out that the DCT always perform better when combined with cosine distance metric. In the next set of experiments histogram remapping with normal distribution (HRN) was applied to this subset, the result indicated HRN perform better than HRL in both cases using the two distance metrics. HRN+euc gave 80.3% in 1 ms/img whereas HRN+cos gave 77.6% in 2ms/img. Similarly, HRL with cosine distance produced 77.6% in 2ms/img and HRL+euc produced 73.7% in 0.9ms/img. In the anisotropic smoothing experiment, the AS + cos combination yields an amazing 92.1% recognition accuracy in 1ms/img, and the AS + euc method yields 90.8% recognition accuracy in just 1ms/img. It can be gathered from the experiments conducted on this complex subset that Gamma Intensity correction performed better that all the techniques used on the subset..

E. AT&T database

This database was used as a means of assessing the performance of the techniques on a database devoid of any illumination challenge. However, the database contains slight variations due to pose, accessories (glasses), and expression. As shown in Fig. 5, the gamma Intensity correction with the cosine distance yielded 95.6% in 1.4ms/img, while the same technique with Euclidean distance yielded 94.4 % in 0.9

ms/img. DCT technique with the cosine distance resulted in 28.8% recognition rate in 1.5ms/img whereas DCT +euc indicated a recognition accuracy of 28.8% in 0.7ms/img. In histogram remapping techniques the following results were obtained: HRN+cos produced 95% in 1.5ms/img recognition accuracy whereas HRN + euc produced 95% accuracy in 0.8ms/mg. In the same way, HRL +cos resulted in 95% recognition rate in 1.4ms/img while HRL+euc resulted in 95% accuracy in 0.7ms/img. These show that the remapping techniques produce somewhat similar results with little difference in execution time. Anisotropic smoothing technique result indicated a drop in recognition rate from the previous results; AS+cos produced 87.5% in 1.5ms/img but AS + euc produced 86.3% in 0.7ms/img.

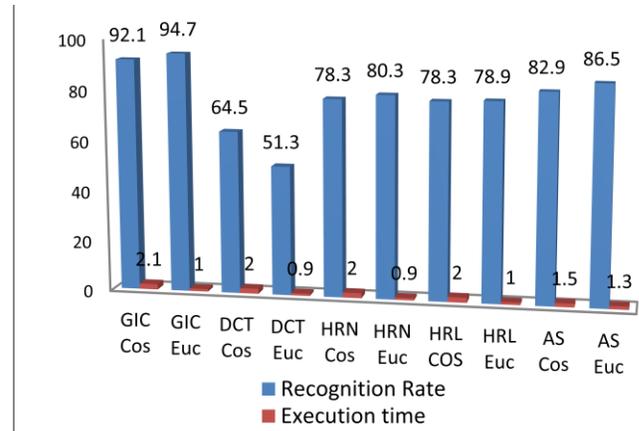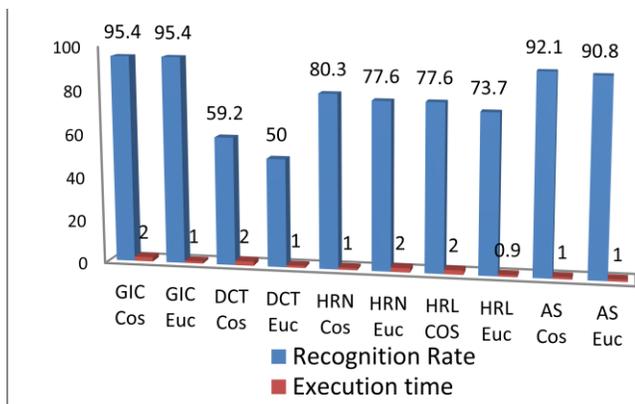Fig. 4. Performances of different preprocessing techniques on subset 5 of the Yale B database using two distance metrics
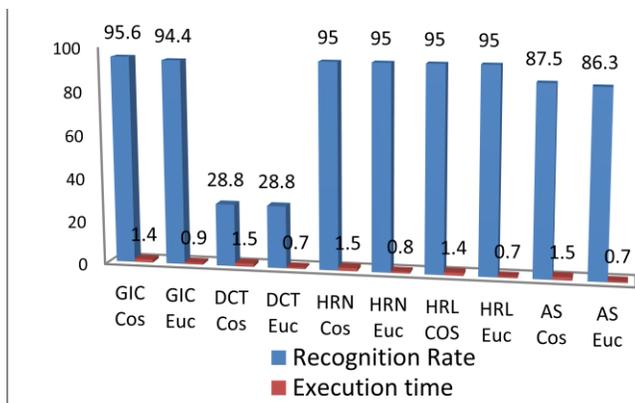

Fig. 5. Performances of different preprocessing techniques on the AT&T database using two distance metrics

IV. CONCLUSIONS

In this paper, different preprocessing techniques for face recognition systems have been proposed and implemented using hybrid approach and linear subspace modeling for feature extraction, and dimensionality reduction and cosine or Euclidean distance metric for classification. The proposed preprocessing techniques are Histogram Remapping using Normal distribution (HRN) and Histogram Remapping using Log-normal distribution (HRL). Other preprocessing techniques experimented with are Gamma Intensity Correction (GIC), Discrete Cosine transform (DCT), and Anisotropic Smoothing (AS). The above mentioned techniques resulted in ten types of face recognition methods: (GIC+EUC), (GIC+COS), (DCT+EUC), (DCT+COS), (HRN+EUC), (HRN+COS), (HRL+EUC), (HRL+COS), (AS+EUC), (AS+COS).

The performances of these ten methods have been evaluated in terms of percentage of recognition accuracy, and for the total execution time to monitor efficiency. The following conclusions are made based on the results and analyses of the above mentioned face recognition techniques:

- The Gamma Intensity correction provides good performance on all the databases particularly those with extreme illumination condition like the subset 4 and subset 5 of the Extended Yale B database. However, the technique was outperformed by histogram remapping using normal distribution on the CASPEAL lighting subset. Generally, the GIC method performed at its best when combined with the Euclidean distance metric, i.e. the (GIC + EUC) arrangement.

- The Histogram remapping technique play a vital role in the CASPEAL complex database in which the lighting and illumination variation is at the extreme with some images very much over-exposed while others are very much under-exposed. In this category of illumination variation technique, the HRN performed better than the HRL in almost all the cases. HRN provides better results when combined with the cosine distance metric.

- The Discrete cosine transform method (DCT) generates the worst recognition accuracy in all the databases used. It can be concluded from these set of experiments that this method is not the best in terms of illumination normalization for face recognition purposes. However, despite this low performance, the DCT method perform better when merged with the cosine distance metric even though it takes some time to finish execution.

- The Anisotropic Smoothing technique (AS) provides second best performance on the Yale B database subsets, however, not so good result was obtained when applied to the CASPEAL lighting subset. This method work well when combined with the Cosine distance metric.

- Between the two distances metrics studied, the cosine distance produce superior performance in almost all the experiments carried out with some minor exceptions like in the GIC technique. The only drawback of this method is the execution time with is slower than the Euclidean distance measure.

REFERENCES

[1] W. Chen, Er. M. J, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," IEEE Transactions on Systems, Man and Cybernetics, vol.36, no.2, 2006, pp. 458-466.

[2] X. Tan, and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," In Proceedings of the IEEE international workshop on analysis and modeling of faces and gestures, 2007. pp. 168-182.

[3] R. Gross, V. Brajovic, "An image preprocessing algorithm for illumination invariant face recognition". In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003 LNCS, vol. 2688, 2003.pp. 10–18. Springer, Heidelberg

[4] S. Ferdinando, H. Andy, "Parameterisation of a Stochastic Model for Human Face Identification." Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, December 1994.

[5] V. Štruc, N. Pavešic, "Photometric normalization techniques for illumination invariance", in: Y.J. Zhang (Ed), Advances in Face Image Analysis: Techniques and Technologies, IGI Global, 2011. pp. 279-300.

[6] Ruiz-del-Solar, J., Quinteros, J., Comparing Pre-Processing Approaches for Illumination Invariant Face Recognition, Technical Report UCH-DIE-VISION-2006-04, Dept. Elect. Eng., Universidad de Chile. Available (August 2006) in: http://vision.die.uchile.cl/

[7] S. Shan, W., Gao, B., Cao, D., Zhao, "Illumination normalization for robust face recognition against varying lighting conditions", In: Proc. IEEE Workshop on AMFG, 2003, pp157-164.

[8] W. Gao, S. Member, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations," vol. 38, no. 1, 2008, pp. 149–161.

[9] V. Štruc, J. Žibert, and N. Pavešic, "Histogram remapping as a preprocessing step for robust face recognition", WSEAS transactions on information science and applications, vol. 6, no. 3, pp. 520-529, 2009."

[10] A. S. Georghiades, and P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", "IEEE Trans. Pattern Anal. Mach. Intelligence", volume 23, number 6, pages "643-660",2001

[11] B.C., Becker, E.G Ortiz. "Evaluation of Face Recognition Techniques for Application to Facebook," in Proceedings of the 8th IEEE International Automatic Face and Gesture Recognition Conference, 2008.

[12] P.N Belhumeur, J.P Hespanha, and D.J Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Transactions on pattern analysis and machine, vol.19, 1997, pp.711-720.

[13] Hu Han, Shiguang Shan, Xilin Chen, Wen Gao, "A comparative study on illumination preprocessing in face recognition", Pattern Recognition, vol. 46, no. 6, 2013, pp. 1691-1699.

Checking the Size of Circumscribed Formulae

Paolo Liberatore

DIAG

Sapienza University of Rome

Rome, Italy

Abstract—**The circumscription of a propositional formula T may not be representable in polynomial space, unless the polynomial hierarchy collapses. This depends on the specific formula T, as some can be circumscribed in little space and others cannot. The problem considered in this article is whether this happens for a given formula or not. In particular, the complexity of deciding whether $CIRC(T)$ is equivalent to a formula of size bounded by k is studied. This theoretical question is relevant as circumscription has applications in temporal logics, diagnosis, default logic and belief revision.**

Keywords—*Circumscription; computational complexity; belief revision.*

I. INTRODUCTION

The circumscriptive reasoning mechanism requires a set of variables to be minimized [1], [2], that is, set to the logical value false whenever possible. Similarly to the closed world assumption [3], it formalizes the assumption that lack of information on certain conditions can be considered evidence that they do not hold. Applications include temporal domains [4], [5], diagnosis [6], induction [7] and belief revision [8]. Contrary to the basic closed world assumption, circumscription takes into account all possible ways variables can be set to false; for example, $x \vee y$ is consistent with either $\neg x$ and $\neg y$ but not both, leading to the two possible cases $(x \vee y) \wedge \neg x$ and $(x \vee y) \wedge \neg y$. These may be up to 2^n, if the number of variables is n: a trivial representation of the circumscribed formula may be exponential. However, it may be equivalent to a smaller formula.

Expressing propositional circumscription as a formula of size bounded by a polynomial has been proved not possible in general [9], unless the polynomial hierarchy collapses [10], a condition generally deemed unlikely. As a result, the problem of whether propositional circumscription can be represented in space bounded by some number k has not an obvious answer: it is possible in some cases but not in others. The problem considered in this article is whether this is possible; in particular, the complexity of this problem is studied. This is similar to the problem of minimizing propositional formulae: given a formula F, is there an equivalent formula of size bounded by k [11]? For circumscription, the question is whether the circumscription of a formula is equivalent to some formula of size bounded by k. For example, the circumscription of $x \vee y$ accounts for both $(x \vee y) \wedge \neg x$ and $(x \vee y) \wedge \neg y$ to be possible; therefore, the result is the formula $((x \vee y) \wedge \neg x) \vee ((x \vee y) \wedge \neg y)$. However, this formula is equivalent to $(x \wedge \neg y) \vee (\neg x \wedge y)$. By the standard metric of formulae where size is defined as the number of variable occurrences, this formula has size 4. Therefore, the circumscription of $x \vee y$ is equivalent to a

formula of size bounded by $k = 4$, but not for example $k = 1$ as no formula of a single variable is equivalent to $(x \wedge \neg y) \vee (\neg x \wedge y)$. The answer is not this easy when the formula is more complex than $x \vee y$. Indeed, it will be proved that the problem is hard for the complexity class Π_2^p, that is, harder than problems such as propositional satisfiability, vertex cover and Hamiltonian cycle [10].

The question of the size of the representation has an implementation impact. Indeed, verifying which conditions hold under the circumscription assumption amounts to $CIRC(T) \models C$, where T represents the current information and C the condition to check, and this is an hard problem [12], [13], [14]. However, if $CIRC(T)$ can be represented by a formula F of bounded size, the problem can be solved by first finding F and then solving the easier (coNP) problem $F \models C$. Once F is determined, any number of other conditions $C_1, C_2, \ldots$ can then be checked against F at the same cost.

Since circumscription is also used as the target of translation of several belief revision operators, the question concerns the dynamic of logic. Indeed, changing a formula to accommodate for new information is generally expected to produce a result of bounded size.

The article is organized as follows: the next section contains the formal definition of circumscription and the notations used in this article, plus two preliminary lemmas; in the section afterwards, the complexity of the problem of whether the circumscription of a formula can be represented in size bounded by some number is studied; the final section comments the practical implications of this analysis and its open problems.

II. PRELIMINARY RESULTS

Propositional formulae are denoted by the capital letters T and F, and are always assumed to be in Negation Normal Form (NNF). Sets of variables are denoted by X, Y and Z. Notation $X^\neg$ indicates the set $\{\neg x \mid x \in X\}$. The shorthand $x \not\equiv y$ indicates $(x \wedge \neg y) \vee (\neg x \wedge y)$.

Models are denoted by ω_X, where the suffix X indicates the set of variables: ω_X is a truth evaluation of the variables X, ω_Y is a truth evaluation of the variables Y, etc. Models are identified by the sets of variables they assign to true; this allows to write $\omega_X \subseteq \omega'_X$ to mean that ω'_X assigns true to all variables ω_X assigns true, but not necessarily the converse. The model assigning true to all variables X is denoted ω_X^+, the one assigning false to all ω_X^-.

The following notation is used to denote a formula that represents a single model: $Form(\omega_X) = \bigwedge \{x \mid \omega_X \models x\} \cup$

$\{\neg x \mid \omega_X \models \neg x\}$. If F is a formula over variables $X \cup Y$ and ω_X a truth evaluation over X, the notation $F|_{\omega_X}$ indicates the formula obtained by replacing each variable X in F with its truth value according to ω_X.

In this article, circumscription is defined over propositional logic, and restricted to the case where all variables are minimized. This gives rise to the following definition.

Definition 1: Given a formula T over variables X, its circumscription $CIRC(T)$ is defined as follows, where $X^\neg = \{\neg x \mid x \in X\}$.

$$CIRC(T) = \bigvee \left\{ T \wedge S \;\middle|\; \begin{array}{l} S \subseteq X^\neg \\ T \wedge S \not\models \bot \\ \forall S' \subseteq X^\neg \\ S \subseteq S' \Rightarrow T \wedge S' \models \bot \end{array} \right\}$$

Some formulae T have small circumscription. For example, $T = \bigwedge X$ has a circumscription equal to itself, since $S = \emptyset$ is the only subset of $X^\neg$ satisfying the definition. Some other formulae have larger circumscription, such as $T = \bigvee X$; indeed, for this formula $S = X^\neg \backslash \{x\}$ satisfies the definition for every $x \in X$. Some formulae do not even have polynomial-size equivalent representations of their circumscription [9].

Circumscription is simple to compute on formulae that imply either x, $\neg x$, or $x \not\equiv x'$ for some variables x and x':

Property 1: The following equivalences hold:

$$\begin{array}{rcl} CIRC(T \wedge x) & = & x \wedge CIRC(T|_{\omega^+_{\{x\}}}) \\[2mm] CIRC(T \wedge \neg x) & = & \neg x \wedge CIRC(T|_{\omega^-_{\{x\}}}) \\[2mm] CIRC(T \wedge (x \not\equiv x')) & = & \begin{array}{l}(x \not\equiv x') \wedge \\ (CIRC(T|_{\omega^-_{\{x\}}\omega^+_{\{x'\}}}) \vee \\ CIRC(T|_{\omega^+_{\{x\}}\omega^-_{\{x'\}}})) \end{array} \end{array}$$

These are well-known properties. The third equivalence allows evaluating $CIRC(T)$ separately for x true and x false, if T does not contain x'.

The size of formulae is defined by the following metrics.

Definition 2: The size of a formula F, denoted $||F||$, is the number of variable occurrences in F.

For example, the size of $(a \wedge \neg b) \vee c \vee \neg(\neg a)$ is four, since the variable a occurs twice in it and b and c once each. According to this definition, the size of a formula and of its NNF form obtained by applying the De Morgan rules coincide. A bound on the size of a formula derives from its models.

Lemma 1: If a NNF formula F has a model that satisfies a literal l but not the modified model where the value of l is inverted, then F contains l.

Proof: Let F be a formula and ω_X its model satisfying l. Let us assume, on the converse, that F does not mention the literal l. Since F is in NNF, no part of it is turned to false by changing the value of l from true to false. As a result, the model ω'_X obtained by changing the value of l in ω_X satisfies F, contradicting the assumption of the lemma. $\square$

As a consequence, if a formula is satisfied by a model where x is true but not by the same model where x is false,

and vice versa, then any formula equivalent to it contains both x and $\neg x$. Therefore, if a formula contains $x \not\equiv y$, either conjoined with a satisfiable formula not containing x and y or disjoined with a non-valid formula not containing x and y, then it must contain at least two literal occurrences for x and two for y. The following lemma shows a sufficient condition for the presence of a literal in a formula.

Lemma 2: Let F be a formula over $X \cup Y$. For any truth evaluation ω_X, no formula equivalent to F is smaller than the smallest formula equivalent to $F|_{\omega_X}$.

Proof: Let T be a formula equivalent to F. Equivalence is preserved when replacing a variable with a truth value in both formulae. As a result, $F|_{\omega_X} \equiv T|_{\omega_X}$. Furthermore, such a replacement does not increase the number of literal occurrences in T, since it only replace some variables with either true or false. As a result, the size of $T|_{\omega_X}$ is less than or equal to the size of T. Since $T|_{\omega_X}$ is a formula equivalent to $F|_{\omega_X}$, it is at least as large as the smallest formula equivalent to $F|_{\omega_X}$. Since T is larger or has the same size, the claim is proved. $\square$

This lemma is useful when formulae contain parts that are satisfiable only for a specific truth evaluation of some variables X. Such formulae are built to the aim of generating a (relatively) large subformula whenever a condition is met.

III. The Size of Circumscriptive Formulae

In this section, we analyze the problem of deciding whether the circumscription of a formula can be represented by a formula of size bounded by an integer k, in unary notation. The unary notation is used to avoid exponentially-sized formulae to be taken into account. Equivalently, the problem could be reformulated as: is there any formula that is equivalent to $CIRC(T)$ and has size less or equal than another formula G?

Theorem 1: The problem of deciding whether $CIRC(T)$ is equivalent to a formula F with $||F|| \leq k$, where k is a number in unary notation, is in Σ^p_3.

Proof: The problem can be reformulated as follows: check whether there exists a formula F that is equivalent to $CIRC(T)$ and $||F|| \leq k$. The problem $F \models CIRC(T)$ is in coNP, since it amounts to check whether $\omega \not\subset \omega'$ for every $\omega \models T$ and $\omega' \models F$. Since coNP is a subclass of Π^p_2, this problem is also in Π^p_2. The problem $CIRC(T) \models F$ is instead Π^p_2-complete [12], [13], [14]; therefore, it is in Π^p_2. The problem under consideration can be therefore solved by guessing a formula F of size bounded by k and then checking whether $F \models CIRC(T)$ and $CIRC(T) \models F$. Since both problems are in Π^p_2, they can be checked by reversing the result of a Σ^p_2 oracle. The problem can therefore be solved by a first nondeterministic step generating all formulae F with $||F|| \leq k$ and then by calling the oracle. It is therefore in Σ^p_3. $\square$

The problem can be proved hard for the class Π^p_2.

Theorem 2: The problem of deciding whether $CIRC(T)$ is equivalent to a formula T' with $||T'|| \leq k$ is Π^p_2-hard.

Proof: Let F be a formula over variables $X \cup Y$. The proof shows how to build in polynomial time a formula T and

a number k in unary notation such that $\forall X \exists Y.F$ is valid if and only if $CIRC(T)$ is equivalent to a formula of size $\leq k$.

Let us assume, without loss of generality, that $|X| = |Y| = n$. The reduction introduces a set of new variables X' in in one-on-one correspondence with X. It also introduces a set of new variables Y' in correspondence with Y and a set of new variables Z of cardinality $m = 3n + ||F|| + 1$.

In this proof the following notations are used, where X and X' are sets of variables in one-to-one correspondence and each x corresponds to $x' = c(x)$:

$$
\begin{aligned}
X^\neg &= \{\neg x \mid x \in X\} \\
X \equiv X' &= \bigwedge\{x \equiv x' \mid x \in X,\ x' = c(x)\} \\
X \not\equiv X' &= \bigwedge\{x \not\equiv x' \mid x \in X,\ x' = c(x)\}
\end{aligned}
$$

Formula T and number k are as follows.

$$
\begin{aligned}
T &= (X \not\equiv X') \wedge \\
&\quad \Big(((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y') \vee \\
&\quad (F \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg)\Big) \\
k &= 14n + 3||F|| + 2
\end{aligned}
$$

The reduction works as follows: $X \not\equiv X'$ allows expressing $CIRC(T)$ in terms of the disjunction of $CIRC(T|_{\omega_X})$ for all possible ω_X; if $\forall X \exists Y.F$ is true, all these formulae $CIRC(T|_{\omega_X})$ can be expressed in the same way, so that a single formula equivalent to $CIRC(T)$ exists with size bounded by k; otherwise, for the evaluation ω_X that makes F false $CIRC(T|_{\omega_X})$ alone has size greater than k.

The first step employs the third equivalence of Property 1, when applied to every $x \in X$ and its respective $x' \in X'$, since T contains $X \not\equiv X'$:

$$
CIRC(T) \equiv \bigvee_{\omega_X} Form(\omega_X) CIRC(T|_{\omega_X})
$$

The second step of the proof is to analyze $CIRC(T|_{\omega_X})$ for an evaluation ω_X. Formula $T|_{\omega_X}$ can be rewritten as follows.

$$
\begin{aligned}
T|_{\omega_X} &\equiv \Big((X \not\equiv X') \wedge \\
&\quad (((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y') \vee \\
&\quad (F \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg))\Big)|_{\omega_X} \\
&\equiv (X \not\equiv X')|_{\omega_X} \wedge \\
&\quad (((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y')|_{\omega_X} \vee \\
&\quad (F \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg)|_{\omega_X}) \\
&\equiv Form(\omega_{X'}) \wedge \\
&\quad (((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y') \vee \\
&\quad (F|_{\omega_X} \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg)))
\end{aligned}
$$

In this last formula, $\omega_{X'}$ is the evaluation of X' setting each variable in X' to the opposite value of the corresponding

variable in X. This formula does not contain any variable in X. Therefore, $CIRC(T|_{\omega_X})$ is defined by taking into account only the other variables: X', Y, Y' and Z. Since X' has a fixed value, it holds:

$$
\begin{aligned}
CIRC(T|_{\omega_X}) &\equiv \\
&Form(\omega_{X'}) \wedge \\
&CIRC(((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y') \vee \\
&(F|_{\omega_X} \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg))
\end{aligned}
$$

The first subformula of circumscription $((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y')$ has only models $\omega_Y^+ \cup \omega_{Y'}^+, \cup \omega_Z \cup \omega_{Z'}$ in which ω_Y^+ and $\omega_{Y'}^+$ set all variables in Y and Y' to true. This model contains a model of the second subformula if $F|_{\omega_X}$ is satisfiable. Indeed, let ω_Y be the model that satisfies $F|_{\omega_X}$. This model is contained in ω_Y^+. The model $\omega_{Y'}$ that assigns $y' \in Y'$ to true if and only if the corresponding $y \in Y$ is false in ω_Y also satisfies $(F|_{\omega_X} \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg)$, and is contained in $\omega_{Y'}^+$. A model of the second subformula is therefore $\omega_Y \cup \omega_{Y'} \cup \omega_Z^- \cup \omega_{Z'}^-$, where $\omega_Z^- \cup \omega_{Z'}^-$ set all variables to false and are therefore contained in $\omega_Z \cup \omega_{Z'}$.

This proves that every model of the first subformula contains a model of the second, if $F|_{\omega_X}$ is satisfiable. If this is the case, the first subformula is irrelevant to circumscription. Otherwise, the second subformula is unsatisfiable.

$$
\begin{aligned}
CIRC &\left(\begin{array}{l} ((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y') \vee \\ (F|_{\omega_X} \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg) \end{array} \right) \\
&\equiv CIRC(F|_{\omega_X} \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg) \\
&\quad \text{if } F|_{\omega_X} \text{ is satisfiable} \\
&\equiv CIRC((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y') \text{ otherwise}
\end{aligned}
$$

The rest of the proof depends on whether F is satisfiable for every ω_X. If it is, then $CIRC(T|_{\omega_X})$ is equivalent to $\omega_{X'} \wedge CIRC(F|_{\omega_X} \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg)$ for every ω_X. As a result, $CIRC(T)$ is equivalent to $CIRC((X \not\equiv X') \wedge F \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg)$, which is equivalent to $(X \not\equiv X') \wedge F \wedge (Y \not\equiv Y') \wedge \bigwedge Z^\neg \wedge \bigwedge Z'^\neg$ by Property 1. This formula has size $4n + ||F|| + 4n + 2m = 8n + ||F|| + 6n + 2||F|| + 2 = 14n + 3||F|| + 2 = k$.

If F is false for some ω_X, then $CIRC(T|_{\omega_X})$ is equivalent to $Form(\omega_{X'}) \wedge CIRC((Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y')$, which is also equivalent to $(Z \not\equiv Z') \wedge \bigwedge Y \wedge \bigwedge Y'$ by applying the second and third equivalence of Property 1. For every $z \in Z$, this formula has a model that makes z true, but changing only the evaluation of z results in a model not satisfying this formula. The same applies to all variables in Z and Z' and their negation, and to all variables in Y and Y'. By Lemma 1, every formula equivalent to this one has size greater than or equal to $4|Z| + 2|Y| = 4m + 2n = 4(3n + ||F|| + 1) + 2n = 12n + 4||F|| + 4 + 2n = 16n + 4||F|| + 4 > k$. By Lemma 2, every formula equivalent to $CIRC(T)$ has size greater than or equal to this amount. □

IV. Conclusions

The problem of checking whether the circumscription of a formula can be represented by a formula of size bounded by k turned out to be Π_2^p-hard in Σ_3^p. These two classes are at the second and third level of the polynomial hierarchy, respectively. As a result, the problem cannot be solved by a propositional satisfiability solver. It can, however, be translated into a QBF and then passed as input to one of the existing QBF solvers [15].

An open question is how much complexity decreases if the formulae are in Horn form, and in particular if some additional restriction makes the problem tractable. If k is in binary representation rather than unary, the question is whether $CIRC(T)$ can be represented by a formula that may be exponential, but still bounded by k. The necessity of considering such large formulae is likely to make this problem harder than with k in unary notation: polynomial space may not be sufficient to solve it.

Indeed, assuming k in unary notation amounts to requiring the equivalent formula to have size comparable to that of the input data. This is equivalent to ask whether $CIRC(T)$ is equivalent to a formula of the same size of another formula G, for example. Allowing k to be stored in binary form with n bit allows the bound be as large as $2^n - 1$. As a result, even formulae of exponential size are allowed as representations of $CIRC(T)$. What complicates the analysis is that the usual guess-and-check algorithm for finding such a formula does not work in polynomial space, as this may not be enough for even storing the formula. A cycle of the minimal models of T is still feasible, but this may not allow determining the size of a formula satisfied exactly by all of them, unless such a formula is explicitly produced.

Acknowledgements

References

[1] G. Brewka, I. Niemelä, and M. Truszczynski, "Nonmonotonic reasoning," in *Handbook of Knowledge Representation*, V. Lifschitz, B. Porter, and F. van Harmelen, Eds. Elsevier, pp. 239–284, 2007.

[2] K. Frankish, "Non-monotonic inference," in *The Encyclopedia of Language and Linguistics*, A. Barber, Ed. Elsevier, 2005.

[3] K. Sengupta, A. Krisnadhi, and P. Hitzler, "Local closed world semantics: Grounded circumscription for owl," in *The Semantic Web–ISWC 2011*. Springer, pp. 617–632, 2011.

[4] M. Shanahan, "The frame problem," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., 2009.

[5] J. van Benthem, "McCarthy variations in a modal key," *Artificial Intelligence*, vol. 175, no. 1, pp. 428–439, 2011.

[6] M. Magnusson, J. Kvarnström, and P. Doherty, "Abductive reasoning with filtered circumscription," *Nonmonotonic Reasoning, Action and Change*, p. 55, 2009.

[7] K. Inoue and H. Saito, "Circumscription policies for induction," in *Inductive Logic Programming*, ser. Lecture Notes in Computer Science, R. Camacho, R. King, and A. Srinivasan, Eds., 2004, vol. 3194, pp. 164–179.

[8] P. Liberatore and M. Schaerf, "Reducing belief revision to circumscription (and viceversa)," *Artificial Intelligence*, vol. 93, no. 1–2, pp. 261–296, 1997.

[9] M. Cadoli, F. M. Donini, M. Schaerf, and R. Silvestri, "On compact representations of propositional circumscription," *Theoretical Computer Science*, vol. 182, pp. 183–202, 1997.

[10] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[11] D. Buchfuhrer and C. Umans, "The complexity of boolean formula minimization," *Journal of Computer and System Sciences*, vol. 77, no. 1, pp. 142–153, 2011.

[12] T. Eiter and G. Gottlob, "Propositional circumscription and extended closed world reasoning are Π_2^p-complete," *Theoretical Computer Science*, vol. 114, pp. 231–245, 1993.

[13] M. Cadoli and M. Lenzerini, "The complexity of propositional closed world reasoning and circumscription," *Journal of Computer and System Sciences*, vol. 48, pp. 255–310, 1994.

[14] M. Thomas, "The complexity of circumscriptive inference in posts lattice," *Theory of Computing Systems*, vol. 50, no. 3, pp. 401–419, 2012.

[15] F. Donini, P. Liberatore, F. Massacci, and M. Schaerf, "Solving QBF with SMV," in *Proc. of the 8th International Conference on Principles of Knowledge Representation and Reasoning*, pp. 578–589, 2002.

A two-level on-line learning algorithm of Artificial Neural Network with forward connections

Stanislaw Placzek
IT Faculty
Vistula University
Warsaw, Poland

Abstract—An Artificial Neural Network with cross-connection is one of the most popular network structures. The structure contains: an input layer, at least one hidden layer and an output layer. Analysing and describing an ANN structure, one usually finds that the first parameter is the number of ANN's layers. A hierarchical structure is a default and accepted way of describing the network. Using this assumption, the network structure can be described from a different point of view. A set of concepts and models can be used to describe the complexity of ANN's structure in addition to using a two-level learning algorithm. Implementing the hierarchical structure to the learning algorithm, an ANN structure is divided into sub-networks. Every sub-network is responsible for finding the optimal value of its weight coefficients using a local target function to minimise the learning error. The second coordination level of the learning algorithm is responsible for coordinating the local solutions and finding the minimum of the global target function. In the article a special emphasis is placed on the coordinator's role in the learning algorithm and its target function. In each iteration the coordinator has to send coordination parameters into the first level of sub-networks. Using the input X and the teaching Z vectors, the local procedures are working and finding their weight coefficients. At the same step the feedback information is calculated and sent to the coordinator. The process is being repeated until the minimum of local target functions is achieved. As an example, a two-level learning algorithm is used to implement an ANN in the underwriting process for classifying the category of health in a life insurance company.

Index Terms—Keyword1; neural network, learning algorithm, hierarchical structure, decomposition, coordination

I. INTRODUCTION

In practice many ANN structures are used but the most popular are the ANNs with forward connections that have a complete or semi-complete set of weight coefficients. The structure of an ANN is depicted in (Fig.1). Neurons in both the hidden and the output layers use sigmoid or tanh activation functions. In the output layer the linear activation function is usually used for approximation tasks. In the most common structures hidden layers include more neurons than input layers, so input information is not compressed in the hidden layers. In this paper two assumptions are accepted:

- To define an ANN structure only the hidden layers and output layer are included. A network described as ANN (10-15-8) includes 10 neurons in the input layer, 15 neurons in one hidden layer and 8 in the output one. It is a two-layer ANN.

- To implement a two-level learning algorithm, an ANN with one hidden layer is used.

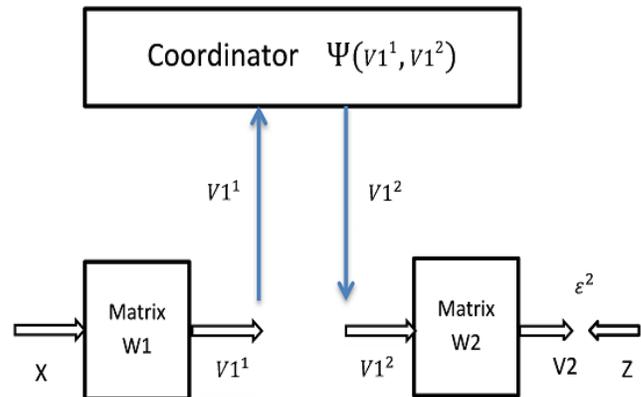The concept layer is used in the primary sense.

Fig. 1: Scheme of the ANN with forward connections

Using (Fig.1) symbols a set of forward and back formulas can be written.
For forward

$$X \rightarrow U1 \rightarrow V1^1 \rightarrow V1^2 \rightarrow U2 \rightarrow V2 \rightarrow \epsilon^2 \leftarrow Z \quad (1)$$

For the first layer (the hidden layer)

$$u1_i = \sum_{j=0}^{n_0} W1_{ij} \cdot x_j \quad (2)$$

$$v1_i^1 = f_1(u1_i) \quad (3)$$

The coordinator is described by the Ψ function

$$\Psi(V1^1, V1^2) \quad (4)$$

For the second layer (the output layer)

$$u2_k = \sum_{i=0}^{n_1} W2_{ki} \cdot v1_i^2 \quad (5)$$

$$v2_k = f_2(u2_k) \quad (6)$$

The target function

$$\epsilon^2 = \frac{1}{2} \cdot \sum_{k=1}^{n_2} (v2_k - z_k)^2 \qquad (7)$$

Where:
$j = 0, 1, ..n_0$ - number of input neurons,
$i = 0, 1, ..n_1$ - number of hidden neurons,
$k = 1, 2, ..n_2$ - number of output neurons.

Using a standard backpropagation notation, derivatives with respect to the weight coefficients are achieved

$$\frac{\partial \epsilon^2}{\partial w2_{ki}} = (v2_k - z_k) \cdot f_2'(u2_k) \cdot v1_2 \qquad (8)$$

$$\frac{\partial \epsilon^2}{\partial w1_{ij}} = \sum_{k=1}^{n^2} (v2_k - z_k) \cdot f_2'(u2_k) \cdot w2_{ki} \cdot f_1'(u1_i) \cdot x_j \quad (9)$$

Equation (4) is known as coordination function.

II. TYPES OF HIERARCHICAL MODELS

Using concepts described in [1], an ANN will be treated as a complex system in an internal hierarchical structure. Three terms are introduced in relation to an ANN:

- The layer of both an ANN and a learning algorithm description or abstraction,
- The layer of algorithm complexity,
- The layer of algorithm structure.

To distinguish between these concepts, the following three terms: a stratum, a level, and an echelon, are used respectively. The term layer is used as a common term referring to any of the aforementioned concepts. For future use of the formal description of different concepts of the hierarchical structures, we describe an ANN as a relation between sets X and Y.

$$ANN : X \to Y \qquad (10)$$

Where:
ANN is a function
X an input set, $X = (x_0, x_1.x_2...x_{n_0})$
Y an output set or response of the ANN, $Y = (y_1, y_2, ...y_{n_2})$.

A. ANN layers of description or abstraction

To treat an ANN as a complex system and to describe it in a complete and detailed way, a different approach should be used. There arises the dilemma of the simplicity of description and the complete understanding of an ANN's behavior [1]. At the first stage, a verbal description is used to help understand how an ANN is built. For a more detailed analysis, mathematical descriptions using algebra and/or differential equations are required. Finally, math formulas have to be implemented in a computer program or an electronic device. Therefore, to achieve a complex description of an ANN, a family of concepts and models from different fields of science and technology have to be used. Every model uses its own set of

variables, laws, principles and terminology by means of which an ANN is described. For such a hierarchical description the functioning on any level should be as independent as possible. To separate this concept of hierarchy from others, a new name is used, a stratified ANN or a stratified description [1]. The layer of abstraction will be referred as stratum (Fig. 2).
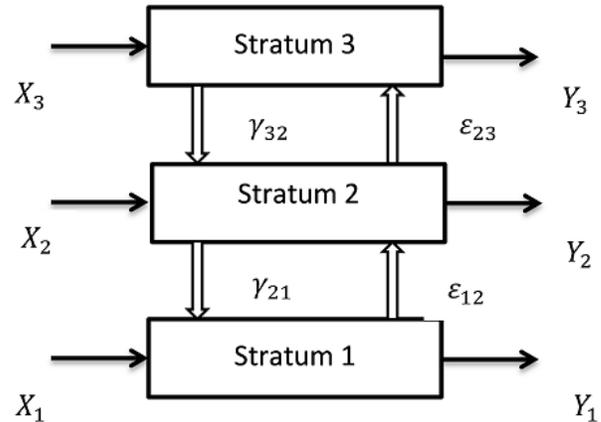

Fig. 2: The ANN stratification description

Using this definition we can state that [1]:
- the selection of strata, in terms of which an ANN is described, depends on the scientist, their target and needs.
- the concepts in which every stratum is described should be as independent as possible.
- one can comprehensively understand how an ANN is working, moving down from the hierarchy of strata.
- a stratified description implies a reduction in information sent up the hierarchy by the reduction of information.

The input set X and the output set Y are both representable as Cartesian products. It is assumed that there are given two families of sets :

$$X = X_1 \times X_2 \times .... \times X_{n_s} \qquad (11)$$

$$Y = Y_1 \times Y_2 \times ... \times Y_{n_s} \qquad (12)$$

Where:
n_s -the number of strata in which one describes an ANN structure.
If concepts in which every stratum is described are fully independent, the ANN stratification can be described as:

$$ANN_i : X_i \to Y_i \qquad (13)$$

Where: $i = 1, 2...n_s$

B. Organisational hierarchy

For a multi-layer ANN a lot of hidden layers and one output layer are sectioned off. The smaller part will be described as a sub-network. Every sub-network has its own output vector that is, at the same time, an input vector of the succeeding one

(IJARAI) International Journal of Advanced Research in Artificial Intelligence,
Vol. 3, No.12, 2014
Extended Paper from Science and Information Conference 2014

$Vi^j, i = 1, 2, ...n - 1, \ j = 2, 3...n$, where n is the number of sub-networks. Because of the specific organisation of an ANN's hierarchy there are a lot of sub-networks on the first level, for each of which local target functions are defined:

$$\Phi = (\phi_1, \phi_2...\phi_n) \tag{14}$$

These sets of local tasks have to be coordinated to achieve the global solution. The coordinator, as an independent task, will have its own target function Ψ. Taking everything into account, this concept is the base on which one may build the new scheme of the ANN learning algorithm structure (Fig. 3). It is the hierarchical organisational structure. To distinguish this concept of hierarchy structure from others, the echelon is used.
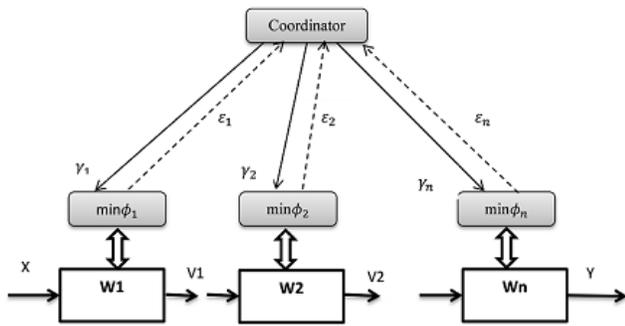

Fig. 3: The hierarchical structure of an ANN learning algorithm

The two-level ANN learning algorithm can be described as a set of procedures. The procedures on the first level are responsible for solving their local tasks and calculating the part of matrix weight coefficients . The second-level procedure has to coordinate all the local procedures (tasks) using its own local target function. There is the vertical interaction between the procedures and two types of information are sent. One is a downward transmission of control signals:

$$\Gamma = (\gamma_1, \gamma_2...\gamma_n) \tag{15}$$

The second is upward from the first level to the second. It is a feedback signal that informs the coordinator about the behaviour of the first-level tasks:

$$E = (\epsilon_1, \epsilon_2...\epsilon_n) \tag{16}$$

Consequently, in all the structures, three different task are defined:

- the global target function

$$\Phi : X \to Y \tag{17}$$

- a set of the first-level tasks (the first level task)

$$\phi_1 : X \times \gamma_1 \to V1 \tag{18}$$

$$\phi_i : Vi \times \gamma_i \to Vi + 1 \tag{19}$$

$$\phi_n : V_n \times \gamma_n \to Y \tag{20}$$

Where:
$i = 1, 2, ...n - 1$
- a coordinator task Ψ

$$\Psi : E \to \Gamma \tag{21}$$

Where: $\Gamma = (\gamma_1, \gamma_2...\gamma_n), \quad E = (\epsilon_1, \epsilon_2...\epsilon_n)$

To build the two-level learning algorithm two assumptions have been made:

- There is no explicit relation between the procedures on the first level for direct communication. The procedures are using only input and output vectors and a coordinator signal.
- There is no direct relation between the global target function Φ and the coordinator task Ψ.

C. Levels of calculation complexity

The standard ANN learning algorithm is a non-linear minimisation task without constraints. To solve this task, iteration procedures are used. Using the most popular back propagation algorithm, one has to choose a lot of control parameters. From a theoretical point of view one can have only general suggestions and recommendations regarding the choice of real parameters, for example learning parameters. The algorithm is time-consuming and its convergence is not fast. Dividing the primary algorithm for all ANNs into the sub-network tasks, the local target functions are simpler and can be used in different procedures. Additionally, a new procedure is needed: the coordination procedure. In practice, however, the coordinator does not have the ability to find all the parameters needed for the first-level procedures. To solve this problem, a multi-level decision hierarchy is proposed [1]. Solving the problems in the iteration algorithm on both the first and the second level, one can observe certain dynamic processes. These processes are non-linear and use a lot of control parameters. During the learning process these parameters are stable and do not change. Practice proves that this solution is not optimal. To control the way learning parameters change, an additional level could be used the adaptation level (Fig. 4).

Thus, one can build three levels at a minimum:

- The local optimisation procedures: the algorithm is defined directly as a minimisation task without constraints.
- The coordination procedure: this algorithm could be defined directly as a minimization of the target function as well. Constraints could exist or not.
- The adaptation procedure: the task or procedure on this level should specify the value of learning parameters not only for the coordinator level, but also on the first level. To solve this task, a procedure should achieve dynamic characteristic of the learning process on all the levels.

As a conclusion one can state that the complexity of the problem increases from the first level to the next one. The coordination and adaptation procedures need more time to solve their own procedure.
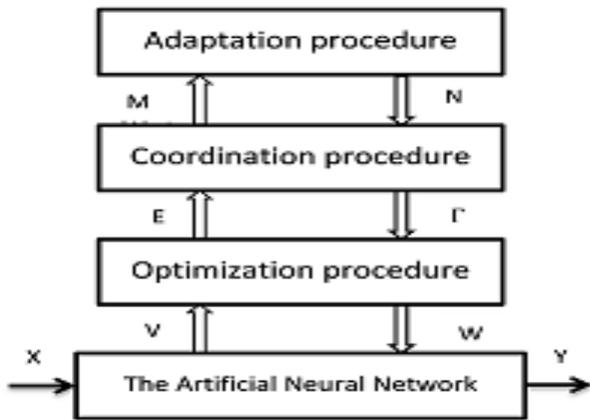
(IJARAI) International Journal of Advanced Research in Artificial Intelligence,
Vol. 3, No.12, 2014
Extended Paper from Science and Information Conference 2014

Fig. 4: Functional multi-level hierarchy for an ANN learning algorithm

III. Two-level algorithm structure

The two-layer ANN with an input layer, one hidden layer and an output layer can be used for further considerations. This simple structure is very popular and by using it one can solve a lot of practice tasks. Since this network is used to solve different classification tasks, sigmoid activation functions are used. To decompose the standard learning algorithm structure into a sub-network task, the coordination target function has to be built. The two-level learning algorithm structure for the ANN with one hidden layer is shown in (Fig. 5).
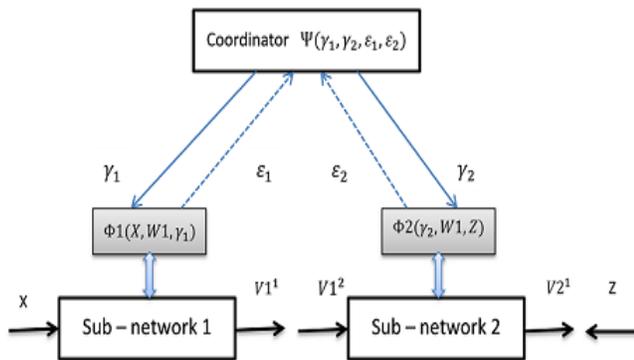

Fig. 5: Scheme of two-level learning algorithm structure

According to [2][8] the following set of formulas can be written.

1) For the first sub-network: The local target function $\Phi 1$ is defined as error- mean- square

$$\Phi 1(W1, X, \gamma 1) = \frac{1}{2} \sum_{i=1}^{n_1} (v1_i - \gamma 1_i)^2 \qquad (22)$$

Other relations

$$u1_i = \sum_{j=0}^{n_1} W1_{ij} \cdot X_j \qquad (23)$$

$$v1_i = f_1(u1_i) \qquad (24)$$

Where:
$\gamma 1 -$ a target value given by the coordinator
Total derivatives with respect to the weight coefficients of matrix $W1$.

$$\frac{\partial \Phi 1}{\partial W1_{ij}} = (v1_i - \gamma 1_i) \cdot f_1'(u1_i) \cdot X_j \qquad (25)$$

A sigmoid derivative function

$$f_1'(u1_i) = a \cdot v1_i \cdot (1 - v1_i) \qquad (26)$$

The new value of weight coefficients

$$
\begin{aligned}
W1_{ij}(n+1) &= W1_{ij}(n) - \alpha_1 \cdot \frac{\partial \Phi 1}{\partial W1_{ij}} \qquad (27) \\
&+ \beta_1 \cdot [W1_{ij}(n) - W1_{ij}(n-1)] \quad (28)
\end{aligned}
$$

Where:
α_1 -the learning coefficient
β_1 - the regularisation coefficient
The feedback information sent by the first sub-network to the coordinator

$$\epsilon 1_i = v1_i \qquad (29)$$

Where:
$i = 1, 2, ... n_1$

2) For the second sub-network: The local target function $\Phi 2$ is also defined as error-mean-square.

$$\Phi 2(W2, \gamma 2, Z) = \frac{1}{2} \sum_{k=1}^{n_2} (v2_k - z_k)^2 \qquad (30)$$

Other relations

$$v2_k = f_2(u2_k) \qquad (31)$$

$$u2_k = \sum_{i=0}^{n_1} W2_{ki} \cdot \gamma 2_i \qquad (32)$$

Where:
$\gamma 2_i$ the input value given by the coordinator.
Total derivatives with respect to the weight coefficients of matrix W2.

$$\frac{\partial \Phi 2}{\partial W2_{ki}} = (v2_k - z_k) \cdot f_2'(u2_k) \cdot \gamma 2_i \qquad (33)$$

And the sigmoid derivative

$$f_2' = a \cdot v2_k \cdot (1 - v2_k) \qquad (34)$$

The new value of the weight coefficients

$$
\begin{aligned}
W2_{ki}(n+1) &= W2_{ki}(n) - \alpha_1 \cdot \frac{\partial \Phi 2}{\partial W2_{ki}} \qquad (35) \\
&+ \beta_2 \cdot [W2_{ki}(n) - W2_{ki}(n-1)] \quad (36)
\end{aligned}
$$

(IJARAI) International Journal of Advanced Research in Artificial Intelligence,
Vol. 3, No.12, 2014
Extended Paper from Science and Information Conference 2014

As we state in (30), the local target $\Phi2$ is the function of $W2$ weight coefficient and $\gamma2$ parameters given by the coordinator. The total derivatives with respect to the coordination parameters $\gamma2$ are:

$$\frac{\partial \Phi2}{\partial \gamma2_i} = \sum_{k=1}^{n_1} (v2_k - z_k) \cdot f_2'(u2_k) \cdot W2_{ki} \qquad (37)$$

New feedback information sent by the second sub-network to the coordinator:

$$\epsilon2_i(n+1) = \gamma2_i(n) - \alpha_3 \cdot \cdot \frac{\partial \Phi2}{\partial \gamma2_i} \qquad (38)$$

3) For the coordinator: In a two-level learning algorithm, the coordinator plays the main role. It is now time to decide what kind of coordination principle will be chosen. This principle specifies various strategies for the coordinator and determines the structure of the coordinator. In [1] three ways were introduced in which the interaction could be performed.

- Interaction Prediction. The coordination input may involve a prediction of the interface input.
- Interaction Decoupling. Each first-level sub-system is introduced into the solution of its own task and can treat the interface input as an additional decision variable to be free. This means that sub-systems are completely decoupled.
- Interaction Estimation. The coordinator specifies the ranges of interface inputs over which they may vary.

In this article the Interaction Prediction is used. The coordinator predicts the interface between sub-networks. This means that the output of the first sub-network $V1^1$ and the input of the second sub-network $V1^2$ are predicted. The signal $\gamma1$ predicts the output signal of the first sub-network $V1^1$. The first sub-network uses this signal as a teachers value and it is a part of the target function $\Phi1$ of the first sub-network according to formula (22). The coordinator predicts signal $\gamma2$ as well. The second sub-network uses this signal as the input value $V1^2$. Using this assumption, it gives the ability to define the local target function $\Phi2$ of the second sub-network (30). Consequently, two local target functions $\Phi1$ and $\Phi2$ can be defined. As stated above, the coordinator needs the feedback information from the first-level sub-networks checking if the predicted signals $\gamma1$, $\gamma2$ were true. If not, the coordinator using its own target function should find the new value of the coordination signal $\gamma1$, $\gamma2$. The first sub-network using the formula (24) calculates the new value of its output signal which,,at the same time, is the feedback signal $\epsilon1$ to the coordinator (29). The second sub-network is trying to minimise the local target function $\Phi2$ and calculate the new optimal value of input signal $\epsilon2$ (38), which is sending to the coordinator. Therefore, the coordinator has full information and is ready to calculate and predict the new value of the coordination input signal $\gamma1$, $\gamma2$. Taking into account that,

$$\gamma1 = \gamma2 = \gamma \qquad (39)$$

the coordinator target function is defined as:

$$\Psi = \frac{1}{2} \cdot \sum_{i=1}^{n_1} (\epsilon1_i - \gamma_i)^2 + \frac{1}{2} \cdot \sum_{i=1}^{n_1} (\gamma_i - \epsilon2_i)^2 \qquad (40)$$

Using gradient algorithm one can calculate using gradient algorithm one can calculate

$$\frac{\partial \Psi}{\partial \gamma_i} = (\gamma_i - \epsilon1_i) + (\gamma_i - \epsilon2_i) \qquad (41)$$

$$\frac{\partial \Psi}{\partial \gamma_i} = 2 \cdot \gamma_i - (\epsilon1_i + \epsilon2_i) \qquad (42)$$

The new value of the coordinator signal $\gamma1_i$ $\gamma2_i$

$$\gamma1_i(n+1) = \gamma2_i(n+1) = \gamma_i(n) - \lambda_1 \cdot \frac{\partial \Psi}{\partial \gamma_i} \qquad (43)$$

$$\gamma1_i(n+1) = \gamma2_i(n+1) = \gamma_i \cdot (1 - 2 \cdot \lambda_1) + \lambda_1 \cdot (\epsilon1_i + \epsilon2_i) \qquad (44)$$
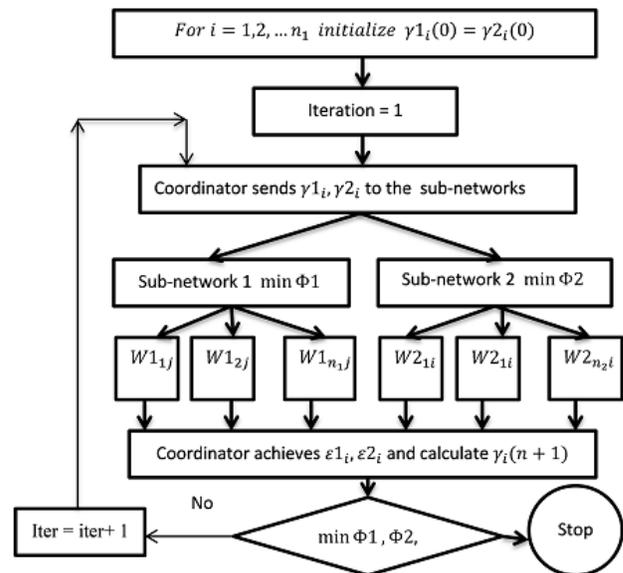
Fig. 6: Functional multi-level hierarchy for an ANN learning algorithm

IV. EXAMPLE

In a life insurance company the underwriting process has been playing the main role in risk control and premium calculation. The ANNs could be used to help the insurance agents to classify the insurance applicant and calculate the first level of premium. Therefore, a special short questionnaire was prepared which only includes 10 main questions. The data were used to teach the ANN work as an insurance specialist, known as the underwriter. All data were divided into three subsets:

- The first set of both the input data X and the output data Z included 250 records of data. This set is known as the learning set. As an example, a small part of the input data is shown in the (Fig.7). The learning epoch includes 250

(IJARAI) International Journal of Advanced Research in Artificial Intelligence,
Vol. 3, No.12, 2014
Extended Paper from Science and Information Conference 2014

vectors that are sending into the ANN input one by one. When this sequence of the process is finished, the next iteration begins.

- The second set is used to verify the quality of the learning process. This set contains 150 records. It is known as verification set.
- Finally, the third set, known as the testing set, contains only 100 records. It helps the decision-making specialist to decide whether the ANN achieves good quality and if it is ready for use.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Nr	Sex	Years old	Weight in kg	Growth in cm	Smoking	Father's long life	Mather's long life	Children number	Living condition	Living place
1	1	48	70	190	0	68	67	0	1	2
2	1	62	71	182	0	82	91	1	0	0
3	0	64	83	180	5	89	100	2	1	2
4	1	40	107	177	10	74	66	2	1	2
5	1	58	120	183	20	90	77	3	1	1
6	1	56	80	174	0	87	75	1	1	2
7	1	60	108	186	0	88	83	3	1	0
8	1	47	91	182	0	78	79	1	1	3
9	1	52	71	186	0	71	81	0	1	3
10	1	62	77	177	0	96	81	2	0	1

Fig. 7: An input data example

To achieve this, the two-level learning algorithm has been used to teach the ANN. The structure of the ANN includes only one hidden layer: 10 input neurons as the dimensionality of the vector X, 15 neurons in the hidden layer and 8 neurons in the output layer. This structure can be shortly described as the ANN (10-15-8). Two sub-networks were introduced in accordance with the algorithm description. The first sub-network includes the hidden layer and its local target function $\Phi1$. The second sub-network includes the output layer with its local target function $\Phi2$. The coordinator has its own local target function Ψ and coordinates local tasks to achieve the minimum of the global target function Φ (the whole ANN). The main goal of this example is to study the dynamic characteristic of the ANN learning process, especially the relations between all the target functions: two local target functions $\Phi1$ and $\Phi1$, the coordinator target function Ψ and the global target function Φ. Al values have to achieve the minimum value according to the relation, as shown in formula (22)(30)(40).

In Fig. 8. the dynamic characteristic of the learning process of the first sub-network is shown. In the beginning phase, the target function $\Phi1$ decreases its value from 1.2 to less than 0.1 during 4,000 iterations. After that, it very slowly decreases its value to the target value 0.001.

To study the dynamic characteristic of the second sub-network (Fig. 9.), one can say that in the beginning phase of the learning process, errors occur less frequently and achieve value 0.1 after only 2,000 iterations. The differences between the sub-networks can be explained by the dimensionality of $W1$ and $W2$ matrices. The matrix $W1$ includes 165 weight coefficients while $W2$ includes only 128 weight coefficients.
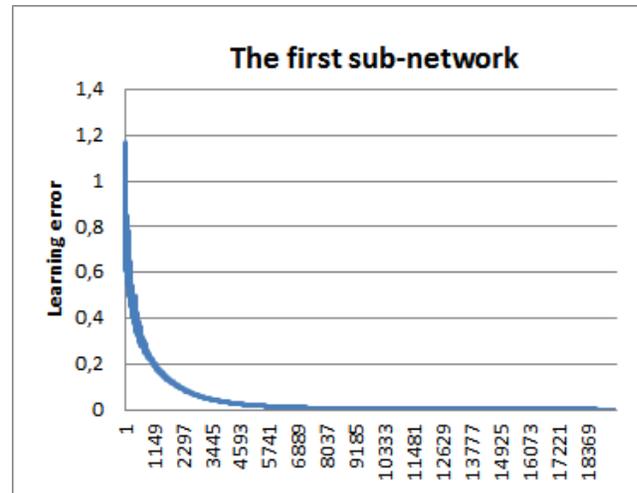

Fig. 8: The target function $\Phi1$ of the first sub-network depending on the iteration number
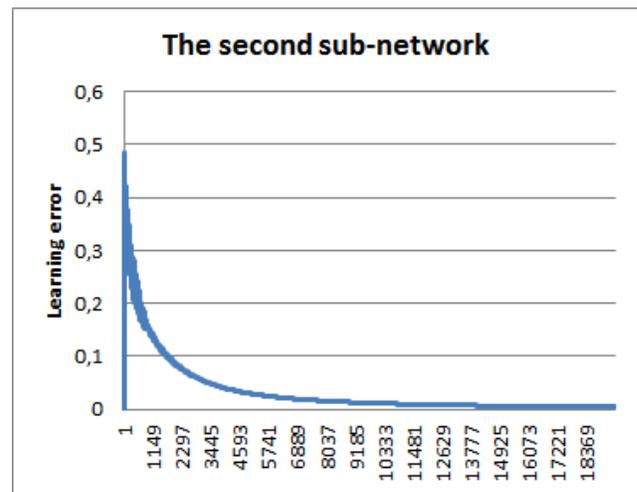

Fig. 9: The target function $\Phi2$ of the second sub-network depending on the iteration numberr

Therefore, in the second sub-network the learning process achieves the minimum value faster.

Finally, the coordinator target function is shown in Fig. 10. The quality of the dynamic processes is the same for both $\Phi1$ and $\Phi2$. The starting error is the greatest and achieves its value 0.1 after 6,000 operations, which can be explained by the relations between the sub-networks. When the learning process started, the sub-networks were not connected (decoupled). This means that every sub-network has to change both the weight coefficient and the input− output vectors using the coordinator gamma signal. The coordinator calculates the optimal γ value using feedback information $\epsilon1$, $\epsilon2$ from both sub-networks. This is the iteration process and it has two stages. During the first stage all errors decrease their value dramatically, rather quickly achieving value less than 0.1. After that, the process stabilises and achieves the final value after a

(IJARAI) International Journal of Advanced Research in Artificial Intelligence,
Vol. 3, No.12, 2014
Extended Paper from Science and Information Conference 2014

considerably long time (number of iterations). This part of the algorithm is not optimal and the coordinator should change the strategy to calculate a new γ value.
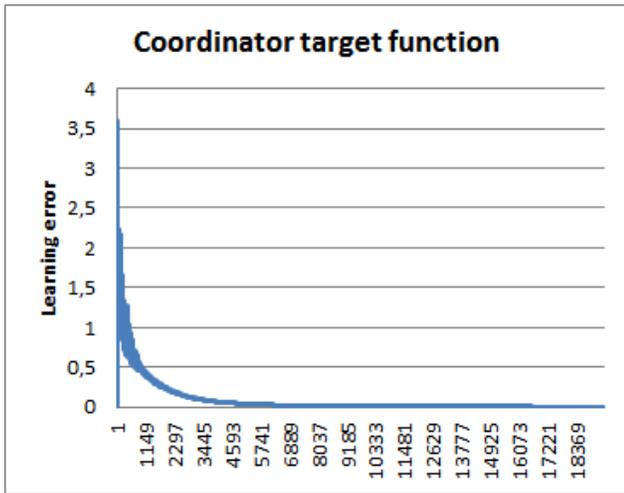

Fig. 10: The coordinator target function value Ψ depending on the iteration numberr

It should be emphasised that in the first stage all the dynamic processes are having vibrations. This can be explained by the fact that when trying to increase the learning process, learning coefficients $\alpha 1$, $\alpha 2$ and $\alpha 3$ have a rather big value, namely 0.8. When learning coefficients are decreased, dynamic processes are more stable, without the vibrations. In Fig.11. a part of the iteration process is shown.
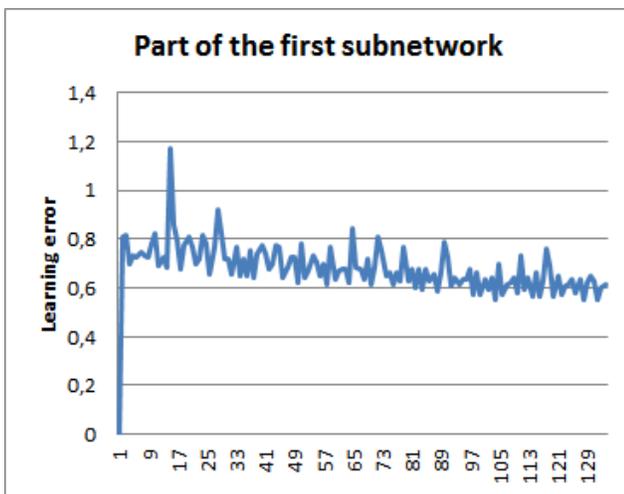

Fig. 11: The part of the learnig process including vibrations

When the ANN achieves the final learning result, the verification sets are used and the differences between the teachers data and the ANN's calculations are collocated (Fig.12). It can be seen that not all the output vectors are the same. In most cases, the ANN calculates less than the teacher (an insurance specialist). The chart shows a number of categories for a couple of insurance candidates. When a category is higher, the insurance premium is greater as well.
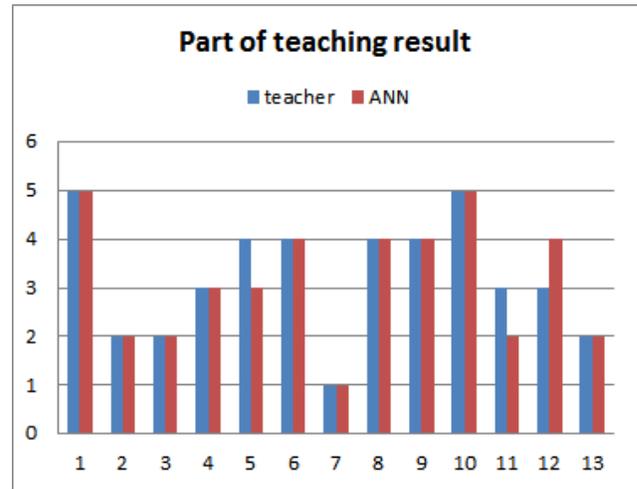

Fig. 12: Result of the ANN's learning

Therefore, one can state that the ANN is more conservative than a life insurance company specialist. For a few candidates it calculates the higher category and they would pay higher premium. Finally, the dynamic characteristic of the global target function is shown (Fig. 13). The characteristic is closely related to all the above. The maximum value is higher and the process needs more time to achieve value less than 0.1, namely after about 8,000 iterations (the sum of the local values).
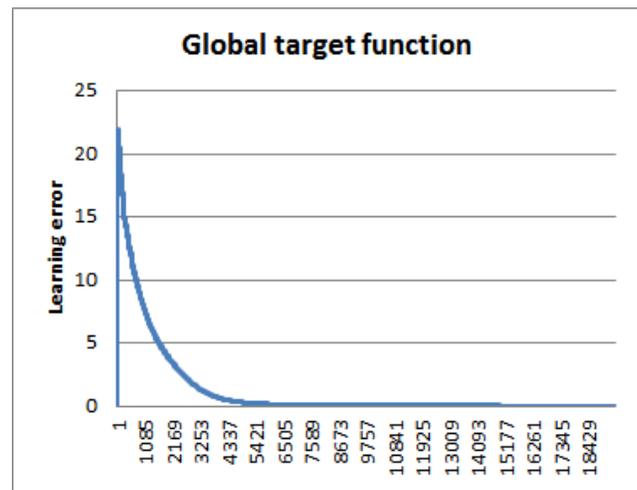

Fig. 13: Learning the global target function Φ for the ANN

Characteristics of the feedback signals $\epsilon 1(n)$ end $\epsilon 2(n)$ are depicted on Fig.14.

V. Conclusion

In [1] for the big systems with hierarchical structure, three coordination principles are defined. For the ANN learning process interaction prediction was used. Each sub-network is
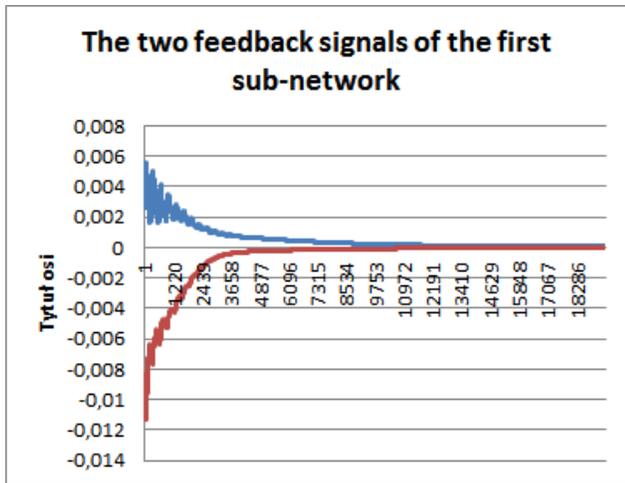
(IJARAI) International Journal of Advanced Research in Artificial Intelligence,
Vol. 3, No.12, 2014
Extended Paper from Science and Information Conference 2014

Fig. 14: Example of dynamic characteristic of $\epsilon 1_1$ and $\epsilon 1_2$

Fig. 15: Example of dynamic characteristic of $\Delta W1$ and $\Delta W2$ matrixes

responsible for finding the minimum value of its own target function to treat the interface inputs as additional variables. The γ signal plays this role. For the first sub-network, $\gamma 1$ is used as the teacher data and the sub-network should change its own weight coefficients in such a way that the final sub-networks output should be as close to the teacher value as possible (in a square error sense, of course). For the second sub-network, gamma works as the input vector. This vector and the teacher vector are used to train the sub-network. The coordinator is responsible for finding the optimal value of the $\gamma 1$ signal, using its own target function Ψ. The underwriting process was used to show that this learning algorithm structure is able of finding the minimum of the global target function Φ for quite a complicated problem and the ANN is ready to work. The weight coefficient for both $W1$ and $W2$ matrices were memorised and the program was sent to insurance agents to use. As it has been emphasised, during the second stage, the learning algorithm works far away from the optimal value. Convergence errors decrease their values very slowly and (Fig.15.) reeffirmed the quality of the dynamic characterics:

$$\Delta W1(n) = W1(n) - W1(n-1) \tag{45}$$

and

$$\Delta W2(n) = W2(n) - W2(n-1) \tag{46}$$

Where:
$n-$ number of iterations

Therefore coordinator should use a more complicated coordination algorithm that has to include not only a PD algorithm structure but a PID algorithm as well. This work will be continued. Analysing the learning result shown in Fig. 12. one can see that the ANN should find a discrete category value (from the set 1 to 8). From time to time, the network solutions are different than the specialist's decisions. Shifting the solution into fuzzy sets could solve this conflict. What follows is the suggestion that the second sub-network should
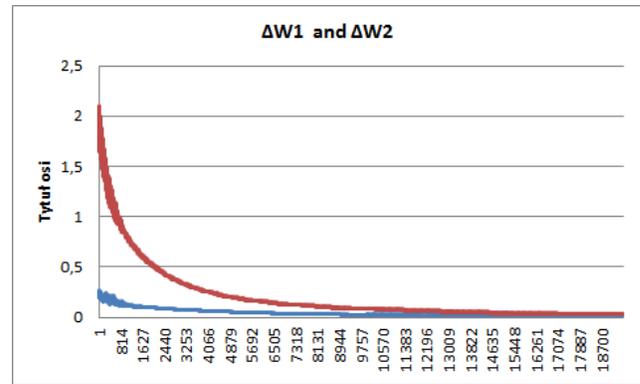
include in its structure the logic module based on the fuzzy solutions.

REFERENCES

[1] M. D. Mesarocic, D. Macko, and Y. Takahara, Theory of hierarchical multilevel systems, Academic Press, New York and London, 1970.
[2] Ch. M. Bishop, Pattern Recognition and Machine Learning, Springer Science + Business Media, LLC 2006.
[3] W. Findeisen, J. Szymanowski, A. Wierzbicki, Teoria i metody obliczeniowe optymalizacji. Pastwowe Wydawnictwo Naukowe, Warsaw 1977.
[4] Zeng-Guang Hou, Madan M. Gupta, Peter N. Nikiforuk, Min Tan, and Long Cheng, "A Recurrent Neural Network for Hierarchical Control of Interconnected Dynamic Systems", IEEE Transactions on Neural Networks, Vol. 18, No. 2, March 2007.
[5] Joarder Kamruzzaman, Rezaul Begg, Artificial Neural Network in Finance and Manufacturing, Idea Group Publishing, Hershley, Pennsylvania 2006.
[6] O. B. Lapunow, On possibility of circuit synthesis of diverse elements, Mathematical Institute of B.A. Steklova 1958.
[7] Toshinori Munakate, Fundamentals of the New Artificial Intelligence, Second Edition, Springer 2008.
[8] S. Osowski, Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006.
[9] S. Placzek, B. Adhikari, "Analysis of Multilayer Neural Network with Direct Connection Cross-forward Connection", CS&P Conference 2013, The University of Warsaw, Warsaw 2013.
[10] L. Rutkowski, Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa 2006.