Artificial Intelligence for Engineering Design, Analysis and Manufacturing (2007), *21*, 19–22. Printed in the USA.

DOI: 10.1017/S0890060407070096

REFLECTIONS

From PUFF to integrated concurrent engineering: A personal evolution

JOHN KUNZ
Stanford Center for Integrated Facility Engineering, Stanford University, Stanford, California, USA

1. INTRODUCTION

Artificial intelligence (AI) emerged from the 1956 Dartmouth Conference. Twenty-one years later, my colleagues and I started daily operational use of what we think became the first application of AI to be used in practice: the PUFF pulmonary function system. We later described the design and initial performance of that system (Aikins et al., 1983; Snow et al., 1998). Today, easily recognizable descendants of that first "expert system" run on commercial products found in medical offices around the world (http://www.medgraphics.com/datasheet_pconsult.html), as do many other AI applications. My research now focuses on integrated concurrent engineering (ICE), a computer and AI-enabled multiparticipant engineering design method that is extremely rapid and effective (Garcia et al., 2004). This brief note compares the early PUFF, the current ICE work, and the modern AI view of neurobiological systems. This comparison shows the dramatic and surprising changes in AI methods in the past few decades and suggests research opportunities for the future. The comparison identifies the continuing crucial role of symbolic representation and reasoning and the dramatic generalization of the context in which those classical AI methods work. It suggests surprising parallels between animal neuroprocesses and the multihuman and multicomputer agent collaborative ICE environment. Finally, it identifies some of the findings and lessons of the intervening years, fundamentally the move to model-based multidiscipline, multimethod, multiagent systems in which AI methods are tightly integrated with theoretically founded engineering models and analytical methods implemented as multiagent human and computer systems that include databases, numeric algorithms, graphics, human–computer interaction, and networking.

Reprint requests to: John Kunz, Stanford Center for Integrated Facility Engineering, Building 550, Stanford University, Stanford CA 94305-4020, USA. E-mail: kunz@stanford.edu

2. SYSTEMS OVERVIEW

Medical patients in a pulmonary lab take a test in which they blow into an airflow measurement device called a spirometer. Entering routine clinical practice in 1977, the PUFF system interprets the digitized airflow and volume data to identify presence and degree of three categories of lung disease. Thus, the original PUFF system did diagnostic reasoning, or analysis, to interpret pulmonary function data. Although the input data are numeric, the representation of disease conditions and the reasoning to interpret the data are entirely symbolic (Fig. 1).

In contrast, ICE is a novel organization form in which multiple human designers each do design work, or synthesis, which is intellectually a much more challenging than analysis. The PUFF system used heuristic knowledge, coded as production rules, from a single domain. Each of 5 to about 20 ICE stations uses one or multiple theoretically founded symbolic models that represent and reason about the function or design intent, form, or design choices and predicted or measured behaviors of integrated product, organization, and process models (Garcia et al., 2004). The ICE method thus uses multiple loosely integrated discipline models (Fig. 2).

PUFF used automated production rule interpretation whereas ICE applications support a mixed initiative method that includes manual synthesis, graphic modeling, and symbolic and numeric analysis of the different integrated models. A prospective (144 case) study measured PUFF performance at 89–96% agreement (SD = 3.8–4.7) of the system to independent experts, whereas the experts had 92% (SD = 1.6) mutual agreement. In multiple sessions, ICE reliably achieves a drop in information processing latency in excess of 4 orders of magnitude ($>$2 days, which is high performance in practice, to $\leq$1 min with $>4\sigma$ reliability) and 2 orders of magnitude for design session duration (e.g., $>$1 month to 2 h, 1 year to 4 days) while maintaining or improving perceived design quality.

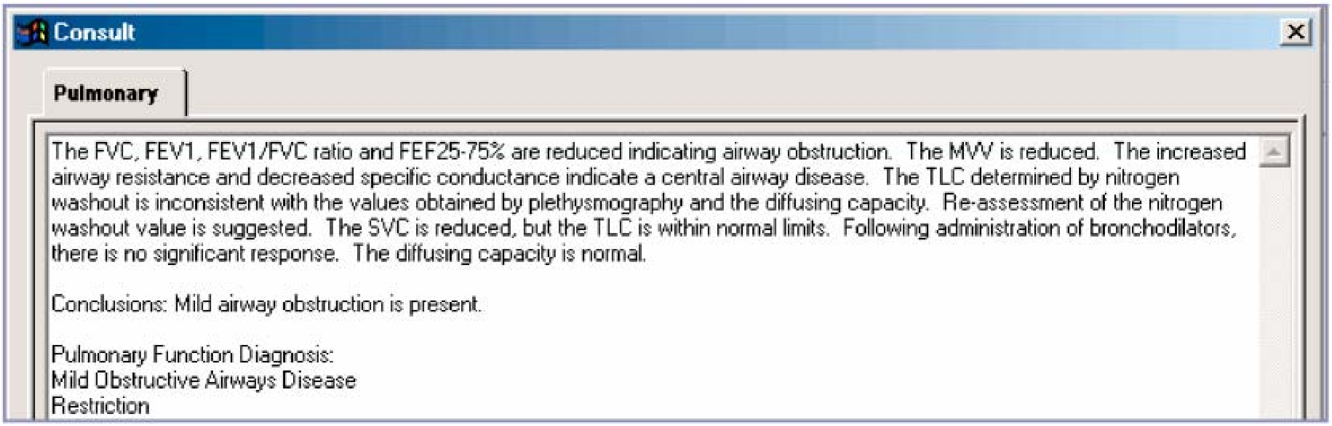

Fig. 1. The early PUFF system created and explained diagnoses of measured data such as this example from the commercial version of the system. Reprinted with permission from http://www.medgraphics.com/datasheet_pconsult.html [A color version of this figure can be viewed online at www.journals.cambridge.org]

PUFF uses text description of its diagnostic reasoning and conclusions. Our implementation of ICE uses graphic, tabular, text, and verbal explanation of descriptive design content, predictions, and their bases; explanations of prediction; and design choice rationale and evaluation of design adequacy given requirements.

3. COMPARING NEUROPROCESSING WITH ICE

Table 1 compares the attributes of ICE with those of a neurobiological system such as current modern AI systems.

4. DISCUSSION

4.1. Good knowledge representation makes reasoning (relatively) easy

PUFF is the unique domain area in my personal experience in which a pure rule-based knowledge representation was simultaneously appropriate for the domain expert, for the developing "knowledge engineers" and as a programming language for the system content. Scores of applications later, some successful and some not, most AI practitioners find that excellent declarative representations are required to make the programming simple enough to both do and to maintain.

Fig. 2. The current multiuser ICE process embeds models and analyses in a social context. The models are symbolic or mathematical and have associated anaysis methods plus specialized interactive graphic visualizations of both the modeled systems and the predicted results. [A color version of this figure can be viewed online at www.journals.cambridge.org]

4.2. Difficulty and cruciality of defining appropriate metrics of performance

In both our early and recent work, we spent almost as much discussing potential performance metrics as we did doing the validation tests that verified the baseline performance of the existing practice and the measured performance of the knowledge system. We find that definition of those performance metrics to be a substantial contribution of the work we do. We now conclude that latency is fundamentally important performance metric of ICE, and underlying ICE design and operating mechanisms must work to reduce both its mean and variance to extremely low levels. Very low latency is also crucial for neurobiological system performance.

4.3. System performance can exceed human performance

The PUFF measured system performance was very high. It was astonishing at the time that the statistical performance of the PUFF system exceeded the performance of expert humans in practice because they had distractions, became weary, and simply missed important features of the data. Our recent ICE work finds the same result: better results in a few days than normally in a month or more.

5. METHOD DIFFERENCES

5.1. Routine dependence on many knowledge sources and diverse computational and social methods

My post-PUFF application efforts always involved creating the new domain knowledge to enable scientifically

https://doi.org/10.1017/S0890060407070096 Published online by Cambridge University Press

Table 1. *A summary of the attributes of ICE methods (2006), neurobiological models of cognitive structure and performance, and PUFF (1977), which has two agents (sensing and interpretation of cognition)*

Attribute	Neurobiological System	PUFF	ICE
Functional elements	Neuron	Sensor, interpretation system	"Station" that is a skilled person with specialized modeling and analysis modeling tools and methods
Integration of functional elements	Tightly integrated cognitive and memory elements in each neuron system	Tight integration feeds data forward, no feedback	Individual ICE "stations" have responsibility to create and analyze models, and they store libraries of detailed models. In addition, ICE has shared memory of information that is common among multiple stations.
Task(s)	Some neuronal complexes process sensory data; others integrate (coarsely coded) detailed sensory data.	Sense (airflow), interpret data	Each functional station models and analyzes details that are never shared. Functional stations share small amounts of (coarsely coded) information with other stations; in addition, they participate in higher level data integration.
States	State free (no locked activation or inhibition)	Same	Same
Delays in sending information from one element to another	Very short (milliseconds to a few seconds), disruption dramatically impedes or stops performance	NA	Very short (seconds to a few minutes), disruption dramatically impedes or stops performance "Latency" is the fundamental process performance metric.
Organization of functional elements	Parallel	Serial	Parallel creation and analysis of multidisciplinary models and analyses and of work flow
Feedback	Large amounts, including local feedback with inhibition	NA	Large amounts, both internally to stations and among stations
Control	Autonomous behavior of elements with emergent making of "sense"	Simple data feedforward	Largely autonomous behavior of elements with emergent making of "sense" that is facilitated by feed-back initiated by a facilitator
Number of external inputs	One or small number of external inputs for each neuron	One	Functional requirements from only a single perspective for each station
Scale	Billions of neurons in the system	One sensor, one interpreter	Ten–twenty stations in a session
Types of functional elements	Neuronal process, inhibitors	Sense, interpret	Modeling and analysis; facilitator, which operates as an inhibitor and focuser of attention
Cognitive complexity of nodes	Very simple	"Knowledge-based" interpretation	Cognitively complex and rich
Information richness of data exchange	Simply coded information shared among many nodes	Simply coded information sent from sensor to interpreter	Simply coded information shared among all nodes
Network topology	Closed (no exogenous nodes)	Same	Same (only people in the room participate in the session)
Prediction	Critical process in cognition	NA	Critical process in design interpretation
Role of visual data	Critical as system input, at least for primates	NA	Critical to describe models and present predictions

founded *model*-based reasoning, where the emphasis was on the model, not the reasoning method. The models need to be good enough so that reasoning can be developed, explained, maintained, and extended. Data and graphics also provide crucial power. Two crucial measures of success are believability of the final results and social engagement of stakeholders, which they need if they are to take the actions that the results imply.

5.2. Users' expectations of high-performance knowledge processing

Users expect high-performance knowledge processing to be social, possibly involving at least one and often many participants, and our systems have evolved to engage multiple stakeholders simultaneously using a variety of models, knowledge and data sources, and reasoning and analysis methods. The change from single- to multiple-user focus is my most recent and probably most significant change from the early days of AI.

6. CONCLUSIONS

AI has progressed a long way in 30 years, judging by this comparison between PUFF, ICE, and the modern AI perspective of neurobiology. Many important future research frontiers seem to lie in different instantiations of the multiagent neurobiology model, with its high cognitive and sensory capability and, in robotics, with its additional actuator capabilities. If ICE is a representative example, early knowledge-based cognitive system models ("good old-fashioned AI") will continue to play an important role in the many kinds of instantiations of multiagent neurobiological models. Symbolic representation and reasoning methods from the early days of AI remain completely relevant today. However, symbolic models and analysis must work effectively with high-performance multiagent organizations, databases, visualization, networking, and human stakeholder participation.

REFERENCES

Aikins, J.A., Kunz, J.C., & Shortliffe, E.H. (1983). PUFF: an expert system for interpretation of pulmonary function data. *Computers and Biomedical Research 16(3)*, 199–208.

Garcia, A., Kunz, J., Ekstrom, M., & Kiviniemi, A. (2004). Building a project ontology with extreme collaboration and virtual design and construction. *Advanced Engineering Informatics 18(2)*, 71–83. Also available on-line at http://cife.stanford.edu/online.publications/TR152.pdf

Snow, M.G., Fallat, R.J., Tyler, W.R., & Hsu, S.P. (1988). Pulmonary consult: concept to application of an expert system. *Journal of Clinical Engineering 13(3)*, 201–205.