Humans in Loops – a Neocybernetic View at Complex Processes

Heikki Hyötyniemi

Helsinki University of Technology, Department of Automation and Systems Technology
P.O.Box 5500, FIN-02150 TKK, Finland
Tel. +358-50-3841626, heikki.hyotyniemi@tkk.fi

KEY WORDS. Expert knowledge, prototype theory, Case-Based Reasoning, Delphi method; control of complex processes, second-order (neo)cybernetics

ABSTRACT

Neocybernetics offers a general framework for networked agent systems. Neocybernetic structures are locally controlled, but the special adaptation principles result in emergent global-level structures. It turns out that the neocybernetic structures *optimally* exploit the information in the environment, so that they have *evolutionary advantage* over all competing strategies. Thus, it can be claimed that neocybernetic models are universal, representing surviving long-term strategies in different kinds of environments, whatever is the nature of the agents carrying out the evolutionary enhancements. – In technical systems these evolutionary agents are humans. This means that there are some general principles governing the behaviors in complex “human-embedded” systems, regardless of the “free will” of the individual process operators and plant designers. Such vision is elaborated on in this paper, and a general framework for processes containing humans in feedback loops is proposed.

1 INTRODUCTION

When facing truly large-scale systems, one is finally facing always the same challenges: Somewhere in the feedback loops there are *humans* integrated in the systems. Typically, the operators implement the highest-level controls in processes, reacting to disturbances (on-line), and above them, there are the process engineers adjusting the tuning parameters (off-line). Also, when exploiting the “extended product” scheme, where the feedback loops extend beyond the factory walls, when formalizing the whole life-span of the products with service and maintenance, one needs to employ wider views.

To fully master the higher-level feedback loops, one needs to understand *human behavior*. In the field of artificial intelligence (AI) research, different approaches to modeling human expertise have been proposed. For example, embedded *expert systems* were studied a lot back in 1980’s; however, rule revision and maintenance turned out to be quite a challenge in such systems. To attack the deficiencies of strictly symbolic knowledge representations, *fuzzy systems* were studied in 1990’s – but now, the representation seemed to be *too* structureless. And so on – there are architectures proposed for all possible needs, but it seems that explicitly addressing one aspect at a time results in complex architectures where there is no *emergence* – all functionalities are hard-coded, missing cognitive plausibility.

And it is not only the mental models that need to be captured – to be truly useful, the higher-level models should also match the properties of the real world, or the plant being managed, in such a way that not all needs to be explicitly taught to the system. Self-learning capability is necessary, so that the “rules” can adapt according to observations. To reach such adaptation, some kind of compromise between the symbolic and numeric approaches is needed. Does this kind of model framework exist?

It now seems that a higher-level perspective truly is available: Just as the feedback controls define standard-like information flow structures within processes, the new approaches offer standard-like ”knowhowflow” structures above the information flows. At this level, the actions of the human operator or designer can be captured, together with the process models, in a unified framework of *neocybernetics* /8/.

2 PROPERTIES OF EVOLUTIONARY SYSTEMS

Neocybernetics offers a general framework for networked agent systems. Neocybernetic structures are locally controlled, but the special adaptation rules result in emergent global-level structures. It turns out that the neocybernetic structures *optimally* exploit the information in the environment, so that they have *evolutionary advantage* over all competing strategies. Thus, it can be claimed that neocybernetic models are universal, representing surviving long-term strategies in different kinds of environments, whatever is the nature of the agents carrying out the evolutionary enhancements. In technical systems these agents are humans, either process designers (off-line) or process operators (on-line).

What this means when studying complex automation systems? – Indeed, industrial plants are like “artificial cells”, where humans are the agents carrying out the “artificial evolution”. The “metabolism” consists of the processes of raw materials being transformed into products. As the local actors (process designers and operators) gradually adjust the process structures and parameters, robustness against external disturbances gets enhanced. The overall system becomes better and better balanced, variations about the operating point getting minimized. Whereas the evolutionary developments are always stochastic, the *pressures* driving the changes can still be modeled in a consistent way; in the long run, largest enhancements take place in the directions of largest pressures. What determines these pressures, then?

2.1 Role of the environment

Whereas AI is defined in terms of mimicking humans, *universal intelligence* can be defined as ability to manage in one’s own environment, whatever are the properties of that environment. In this sense, the problem of artificial intelligence can be extended to studies of *artificial life*. Relevance of behaviors is determined in terms of *feedbacks through the environment*. Thus, the goal of artificial evolution is to become better adapted in one’s own world.

As assumed in *embodied embedded cognition* (EEC), real intelligence emerges out of the interplay between brain, body, and world.

The problem of intelligence has to be attacked from the bottom, not from the top – the simplest form of intelligence are *reflexes*, where the reasonable behavior is reached simply through direct feedback through environment. Such opposite view of intelligence has been exploited in Behavior Based Artificial Intelligence (BBAI) that was made famous by Rodney Brooks /2/. His *subsumption architecture* has become popular especially in robotics, where real-time reaction to surroundings is vital. There are connections to neocybernetics, where orientation towards the environment is still more emphasized:

> *All relevant information is assumed to be available in the high-dimensional observation data; information is determined in terms of statistical quantities, and pragmatic semantics is defined through relevance of functionalities; structures within a framework are dictated directly by the data properties, upper levels becoming “mirror images” of the lower-level data.*

The Brooksian subsumption architecture assumes hierarchies and modularity of functionalities. Now, this kind of predetermined structural assumptions are avoided. Structured information representations *are* relevant, but the models become very different.

2.2 Complex systems in interaction

What the subsumption architecture fails to take into account, is the fact that one should not only be constructing models for a system in isolation: indeed, in complex automation systems one has two complex systems, the automation system domain and the mental system domain, side by side. The environment of the process is largely determined by the humans controlling it, and the environment determining the cognitive actions is the process being monitored. This is where cybernetics can contribute.

The idea of *second-order cybernetics*, as originally coined by Heinz von Foerster, is “cybernetics of observing systems” /5/. The key point is that as the observer affects the system being observed, the overall system becomes a complicated hierarchy. What is more, such view easily leads to *relativism*, where everything is dependent of the observer. In neocybernetics, however, there is new hope: The models of the observer and the observed can have similar structures, and, rather than becoming a hierarchy, coupling of the systems results in yet another system that is not qualitatively more complicated than the component systems:

> *Cognitive and physical systems can be modeled using the same structure; the two complex systems have symmetrical roles, so that ontology (what there is in the world) and epistemology (how one sees it) become the same.*

2.3 Reduction of complexity

When employing the neocybernetic approach, more can be said about the model structures – and there are efficient mathematical tools for analysis of those model structures:

> *The resulting multivariate models are essentially linear; the information (in the form of covariations) is concentrated along the degrees of freedom (DOF); these subspaces can be analyzed in terms of principal components in data.*

Here, *principal component analysis* (PCA) is a traditional data compression method /1/, where the covariation in data becomes maximally captured. In the neocybernetic framework, however, PCA seems to be more than an analysis method: Following the optimal local adjustments, the system starts representing the covariation-induced structure. The principal subspace reveals the *degrees of freedom* for the system. The concentrated covariation (information) means importance and interestingness – there is “most to do” in those most significant directions. As seen from outside, a neocybernetic system looks *elastic* or *resilient*, that is, external disturbances make the system yield until the pressure is released. This yielding takes place in the directions of degrees of freedom, the elasticity in those directions being determined by the *coupling* between the system and its environment.

The neocybernetic approach has been successfully tested in various large-scale test cases /4/. In those examples, the problem of mental domain modeling has been simplified: It is assumed that the human-defined *goals* of control and all other relevant knowledge can be expressed as explicitly quantifiable *quality measures*. Additionally, assuming that one is only studying the vicinity of the operating point, so that the process data is practically unimodal and linear, makes it possible to map the *qualifiers* onto *qualities* applying simple linear multivariate methods. Degrees of freedom (the latent variable basis vectors) are determined now in terms of correlations between qualifiers and qualities. Constructing a feedback from the qualities back to qualifiers makes it possible to adapt the process parameters gradually towards better process quality. The process parameters being optimized can include for example set points, controller parameters, and tuning factors of models when using some model predictive control scheme. The vision of relevant information being packaged into a low number of latent variables makes such iterative optimization towards locally most beneficial direction a feasible task.

The examples above are based on off-line optimization – but the optimum cannot always be predetermined as the environmental conditions vary. Still, this “truncated” model of cognitive decision making can be applied also for helping in the on-line process operation tasks: The degrees of freedom, as determined by the qualifier/quality correlation structures, reveal the directions where the relevant changes in the process assumedly take place. For example, if the qualities are *accuracy*, *robustness*, and *speed*, as defined by some process expert, one can introduce a virtual “ARS controller”, where one can slide from “more speed” to “less speed”. The underlying real process parameters (like PID tuning parameters) can then be adjusted according to the DOF model in a transparent manner. Thus, the “interface” to the process can be brought nearer to the human: Rather than having to tune the individual PID parameters, resulting in non-optimized operation of the overall process, one can simultaneously adjust all of them by just commanding “more speed”. The acceptability of such ARS controller as an “enhanced PID” would probably be better than with other sophisticated controllers as it matches the operators’ PID control intuition – but this time not on the control loop level, but on the whole plant level.

In both of the cases above, in the off-line and on-line application, human is still needed in the loop, even though his/her task of managing the process is made easier. Indeed, the iteration loop is brought to a higher level: It often turns out after optimization that the defined quality measures did not result in satisfactory production after all, and the quality measures themselves have to be adjusted. In this sense, the new approaches may lead to humans becoming still better experts in their fields – gaining better understanding of what is *good behavior* and what is not. – But we are not satisfied with this, as we want to free the human from the feedback loop altogether; after all, such living at the mercy of processes and machines is not a decent task for humans.

3 MODELING OF COGNITION

Finally, however, the real challenge has to be faced: There are two complex systems side by side, the automation system and the mental system, each of them affecting the other. There can be nonlinearities in both systems, as the structures can change either in the process or in the mental model, but still, a global model has to be pursued to reach real benefits. – Regardless of the challenges, in the "second-order neocybernetic" framework these problems need not necessarily be hopeless: As stated above, the human cognitive machinery is based on the same principles of self-regulation and self-organization as the adapting control systems, and *modeling of expertise* becomes possible in a similar framework.

3.1 Structures in mental models

There is no general agreement of what kind of principles cognition is based on. However, there are some specially promising approaches that seem to be compatible with neocybernetics. In *prototype theory*, as formulated by Eleanor Rosch and others, semantic categories (concepts) are determined in terms of *examples*, where some examples are more central than the others /10/. This view was functionalized in the AI field under the name *case-based reasoning* (CBR), where new problems are solved based on the solutions of similar past problems /9/. The problem here is how the "similarity" should be assessed. In neocybernetics, the categories are more structured, and distance measures are well defined:

> *The chunks (symbolic or subsymbolic memory elements) are patterns constructed of sparse-coded features, or "relevant attractors" among data; inference is pattern matching in the data space.*

The process of *shift from novice to expert* can be explained in a numeric framework better than in a symbolic one. In the neocybernetic framework, the "conceptual spaces" /3/ now have more structure, because sparsity makes it possible to implement mentally plausible nonlinearities (for example, see /6/).

3.2 Exploiting expertise

How can available expertise be employed in an efficient way? It seems that the *Delphi method* offers a nice framework /11/. Delphi method relies on a panel of independent experts who answer questionnaires in two or more rounds; after each round, a facilitator provides an anonymous summary of the experts' opinions, as well as the reasons they provided for their judgments. Thus, participants are encouraged to revise their earlier answers in light of the replies of other members of the group. It often turns out that during this process the range of the answers decreases and the group will converge towards the "correct" answer. – Indeed, this is a very neocybernetic way of processing and combining information, since for the neocybernetic model there holds:

> *The effects are "pancausal", all variables affecting the others; the model is a balance among internal and external tensions; as the dynamic transients converge, the static pattern emerges.*

In the case of expertise about an industrial process, the problem of representing this expertise is better manageable: The set of variables is easier quantified, and the results can be assessed in terms of money. As compared to traditional Delphi-style expertise, the "world" is now narrower, but decisions can be continuous-valued. For example, when the process operators' expertise is being modeled applying the neocybernetic approach, different operators' actions are used as input data for the model; discrepancies among actions are the "expert tensions", and their balances are captured in the model; this promises better negotiability of knowledge.

4 NEOCYBERNETIC HIGHER-LEVEL CONTROL

Putting it all together, one can get from the very abstract ideas of second-order cybernetics to more concrete "second-order control". The universal model captures the complexity of a natural system (expert knowledge) and that of a technical system (automation process) in the same framework. The structures of *ontology* and *epistemology* become the same; or, as Heraclitus would perhaps have said it, "the way up and the way down in the loop are the same". This symmetry can be interpreted so that as the human observes and affects the process, the *process simultaneously observes the user's actions*, reacting to them actively, and simultaneously makes it necessary for the mental models to adapt according to the changed behaviors (see Fig. 1).

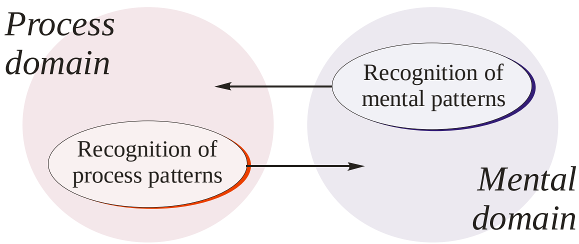

Figure 1. Two complex systems affecting each other

4.1 Capturing expertise in practice

The output information from the process is the input information for the mental system, and *vice versa*. In both cases, that information can be captured in the neocybernetic pattern model. To work in an appropriate manner, the patterns in both ends of the closed loop need to match each other. And, according to the neocybernetic theory, both of the pattern models are based on sparse coded features, and the mapping between them goes through a linear subspace.

How to select variables so that all relevant information is captured? First, as compared to the time scales of the actual process variables, the time scales at the higher level are considerably longer, being manageable to humans (from minutes to days). Typically, these variables are *emergent* in neocybernetic models:

> *"Weak emergence" is reached when, instead of the signals themselves, one uses their statistical and static expectations (variances, spectra, etc.); individual signal instances are abstracted away.*

There typically are very high numbers of process data available for analysis. Even though new sensors may be introduced sparingly, different kinds of *soft sensors* can be used to deliver and manipulate information. For example, when *cameras* are installed, dozens of frames (containing thousands of pixels each) are obtained each second. Whereas the human capability of visual pattern recognition is magnificent, soft sensors are needed to extract (possibly) relevant features out from the raw image data. An example of extracted features and resulting patterns is presented in /7/. – As yet another example of the new challenges, the relevance of pattern recognition is visualized when doing *alarm handling*: Many signals may go off simultaneously, even though there is typically just one fault causing the avalanche.

4.2 Implementation details

The neocybernetic framework offers also some practical rules of thumb, making signal preprocessing simpler. Because the user actions are observed in closed loop, controls assumedly having their effect, there holds:

> *Signals in tightly coupled cybernetic systems become equalized; this means that (traditional) normalization of variables is appropriate.*

In Fig. 2, the general structure of the higher-level control scheme is depicted. Whereas the "easy", traditional measurement and control takes place on the lower level of information processing, the upper part represents *knowledge processing*. The key point is to collect all (possibly) relevant quantifiable information from the process and from its control and signals, extracting appropriate features from them, and find the *patterns* in that data. This pattern recognition takes place separately for the observation data and for the action data, the former being collected by observing the process but the latter being collected by observing the *humans*. Between these two sparse-coded feature models, there is a mapping through a linear subspace. If implemented in control systems, integrated expert models look more or less like "multivariate fuzzy controllers".

It needs to be recognized that the information flows between the process domain and the mental domain can be rather abstract flows; they are all-invasive, ubiquitous just as human observation and intervention is – after all, such high-level feedback loops finally *cannot* be fully automated (what a surprise!), but new models can offer tools for humans to easier manage the complex plant.

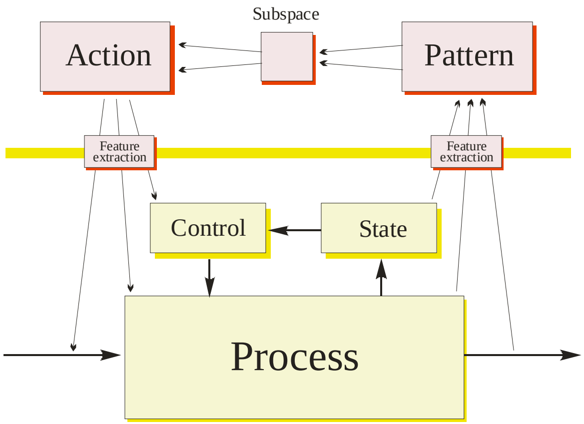

Figure 2. General structure of the "second-order control"

5 CONCLUSION

Neocybernetics has been successfully applied to off-line tuning of complex processes, and the promising results have made it possible to propose yet another level of on-line feedbacks in complex plant models: Including the user model in the feedback loop. This paper presented the general guidelines how this can be done; the practical applicability of the approach remains to be evaluated.

6 REFERENCES

[1] Alexander Basilevsky: *Statistical Factor Analysis and Related Methods: Theory and Applications*. John Wiley & Sons, New York, 1994.

[2] Rodney A. Brooks: *Intelligence Without Representation*. Artificial Intelligence, Vol. 47, 1991, pp. 139–159.

[3] Peter Gärdenfors: *Conceptual Spaces: The Geometry of Thought*. Bradford Books (MIT Press), Cambridge, MA, 2000.

[4] Kalle Halmevaara and Heikki Hyötyniemi: *Process Performance Optimization Using Iterative Regression Tuning*. Helsinki University of Technology, Control Engineering Laboratory, Report 139, April 2004.[1]

[5] Francis Heylighen and Cliff Joslyn: *Cybernetics and Second-Order Cybernetics*. In R.A. Meyers (ed.): *Encyclopedia of Physical Science & Technology* (3rd edition), Academic Press, New York, 2001. Available also in Internet at "`http://pespmc1.vub.ac.be/Papers/Cybernetics-EPST.pdf`"

[6] Heikki Hyötyniemi: *Expertise: an optimized data model?* Proc. of Finnish Artificial Intelligence Days (STeP 2000), August 28–30, 2000, Vol. 2, pp. 95–102.[1]

[7] Heikki Hyötyniemi and Vesa Hasu: *Independent components of expertise*. Proc. of Finnish Artificial Intelligence Days (STeP 2000), August 28–30, 2000, Vol. 3, pp. 37–44.[1]

[8] Heikki Hyötyniemi: *Neocybernetics in Biological Systems*. Helsinki University of Technology, Control Engineering Laboratory, Report 151, August 2006.[1]

[9] Janet Kolodner: *Case-Based Reasoning*. Morgan Kaufmann, San Francisco, CA, 1993.

[10] Eleanor Rosch: *Principles of Categorization*. In E. Rosch and B.B. Lloyd (eds.): *Cognition and Categorization*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1978, pp. 27–48.

[11] Gene Rowe and George Wright: *The Delphi technique as a forecasting tool: issues and analysis*. International Journal of Forecasting, Vol. 15, Issue 4, October 1999.

---

[1] Papers available through "`http://www.control.hut.fi/research/cybernetics/publications.html`"