The Worrying Statistics of Connectionist Representation

Chris Thornton
Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QN
Email: Chris.Thornton@cogs.susx.ac.uk
Tel: (44)27 678856

December 14, 1994

CSRP 362

Abstract

The paper looks at how the hidden-vector cluster analyses associated with Elman and others seemed to provide a potentially important link between the symbolically-oriented level of analysis and the connectionist level of analysis — a link that might one day help to explain how higher mental processes are grounded in neural architectures. The paper goes on to reconsider the implications of these analyses in light of some recent work by Finch and Chater which shows that linguistically meaningful categories (of the type derived from hidden-vector analyses) are directly evidenced in the N-gram statistics of natural language. The implication of this work seems to be that hidden-vector analyses do not primarily address the link between the symbolic and connectionist levels of explanation but rather tell us something about the statistical properties of the training environments used. The consequences of this result for cognitive science are lightly sketched in.

1 Introduction

There has always been the hope that work in Artificial Intelligence (AI) would help to elucidate and extend the philosophical study of the mind. But, paradoxically, the interface between AI and philosophy appears to have become harder to negotiate as the years have gone by. In the early days links between AI and philosophical studies were readily apparent. AI researchers tended to construct programs that reflected their introspections about mental processes or, in some cases, verbal protocols given by human problem solvers [cf. 1]. This quite naturally produced systems populated with familiar landmarks.

But if the early research was relatively accessible to philosophical minds, developments in the field soon seemed to be carrying AI off into an 'outer space' remote from both introspective experience and philosophical conceptualisation. The effect was, perhaps, particularly noticeable in the area of vision research. Vision researchers of the 1960s, e.g., Roberts [2], were largely concerned with systems which sorted out neat, perspective drawings using rules whose good sense could easily be comprehended. cf. Waltz filtering [3]. But by the late 1970s researchers had begun to work with systems which dealt primarily in abstruse mathematical constructs having nothing obvious to do with the mind or the real world, cf. [4].

esnothingtodowithsubstrates.Beingafunctionalistimpliesthatallthatisrequiredforstatesis
Andwhenitcame,itcameinarushwiththemost.JohnSearledemonstratedtheofAIsystemshedidamore

to the case where we attempt to build computer simulations of mental processes. In simulating mental processes we are simply trying to construct abstractions of the original phenomenon.[2] No matter how accurate our abstraction, some properties of the original phenomenon will necessarily be lost. This is, after all, the essence of abstraction.

Since a computer simulation is just another form of abstraction, and since abstraction necessarily wastes properties, computer simulations of mental processes potentially lose some of the properties of real mental processes. The implication is that Searle was essentially correct: architecture may *not* be irrelevant. The properties we are actually interested in (understanding, belief etc.) may be to do with characteristics of the substrate.

2 Connectionism and the need for good grounding

Searle's attack on strong AI seems to have reflected

showed how the network had constructed an internal hierarchy which flagged linguistically important distinctions (e.g., consonant versus vowel.) The newsworthiness of this work was founded on the fact that these distinctions were not given to the network a priori. Rather they were learned directly from the data.

Sejnowski and Rosenberg's hidden-vector analysis method soon became part of the standard toolkit of the connectionist researcher. Recently, it has been used to particularly good effect by Elman who showed how a copy-back network trained to do word-prediction (given only a diet of raw English sentences), formed an internal hierarchy that captured lexical and semantic categories [12].

For anyone dreaming wistfully of a bottom-up, 'reverse-cascade', this new work by Elman and others looked very promising. The notion that the behaviours of connectionist systems embodied tacit rules was fairly well accepted especially in light of Rumelhart and McClelland's work on the learning of past tenses [13]. But with the new hidden-vector analyses one could now say much more precisely what form the terms of these rules might take. In effect, the hidden-vector analyses provided an initial step-up on the reverse cascade. It built a small bridgehead that connected the mushy and remote world of low-level connectionism (a world of 'weights', 'activation values', 'links', 'units', 'energy levels' etc.) with the rather more tractable world of symbols and class definitions.

Andy Clark was quick to see the potential of this new method. In discussing

4.1 A type-1 theory for copy-back networks?

What should we make of this work? Finch and Chater's own view is that statistical analysis provides us with a better understanding of the performance and behaviour of certain sorts of networks (e.g., Elman, copy-back networks). They conclude that their statistical work shows that the 'copy-back scheme is sampling these [N-gram] statistics successfully.' The go on to say that 'these results suggest that the hidden unit patterns that recurrent neural networks develop can be viewed as reflecting quite directly the statistical structure of the sequences learnt.' [17]

By showing that the internal structures formed by copy-back word-prediction networks closely resemble the structures derived from a particular statistical analysis, they have effectively shown that the networks are sampling the relevant statistic. In a sense, they have provided a type-1 theory [18] for the behaviour of these networks. The theory says that the network is performing a particular computation and it characterizes this computation without making any reference to implementation issues.

For those who want to believe that architecture and grounding are important, this is clearly a worrying demonstration since it seems to eliminate the 'ground' altogether. Surely, if all an Elman network is doing is sampling a certain statistic then its 'networkness' cannot be the origin of significant properties. A functionalist stance towards such networks, then, would seem to be perfectly appropriate. On the other hand it might be argued that any retreat into functionalism *must* be premature. The statistical work in question has only looked at one particular domain (natural language) and has produced results which seem to bear directly on only one type of network (the Elman copy-back net). Our assumptions about the importance of grounding and our hopes for the reverse cascade may then turn out — when other systems are analysed more carefully — to be be fully justified.

5 Is it statistics all the way up?

However things go for the 'grounding' issue, one thing is clear: Finch and Chater's work suggests that we should review our attitude to the value of statistical analysis. Classical AI made practically no use whatsoever of it. New approaches such as reactivism and alife-ism have also tended to largely ignore its potential. Connectionism has used it to a certain degree but typically only for the purposes of analysing the behaviour of models. Finch and Chater's work suggests that it can play a much more direct role in our attempt to understand the nature of concepts and classes. Of course, all that has been shown to date is that certain linguistic classes show up directly in the N-gram statistics of 100(couehn.12T999Of)-1700(mor

they concentrated on analysing 5-gram statistics of text. As they note, 'an N-gram is an ordered sequence of N symbols. The frequencies of occurrence of each N-gram in a continuous stream of data constitutes the N-gram statistics of the data set.' [15]. Their aim was to look at the number of times

[3] Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. Winston (Ed.), *The Psychology of Computer Vision* (pp. 19-92). Mcgraw-Hill.

[4] Scott, G. (1988). *Local and Global Interpretation of Moving Images.* Research Notes In Artificial Intelligence, London: Pitman.

[5] Thornton, C. (1985). A response to searle's thesis. *AISB Quarterly*, No. 52 (pp. 32-33).

[6] Sloman, A. (1985). Strong strong and weak strong AI. *AISB Quarterly*, No. 52 (pp. 26-31).

[7] Dennett, D. (1987). *The Intentional Stance.* Cambridge, Mass.: MIT Press.

[8] Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47* (pp. 139-159).

[9] Brooks, R. (1991). Intelligence without reason. *Proceedings of the Twelth International Joint Conference on Artificial Intelligence* (pp. 569-595). San Mateo, California: Morgan Kaufman.

[10] Hinton, G. and Anderson, J. (Eds.) (1981). *Parallel Models of Associative Memory.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

[11] Sejnowski, T. and Rosenberg,