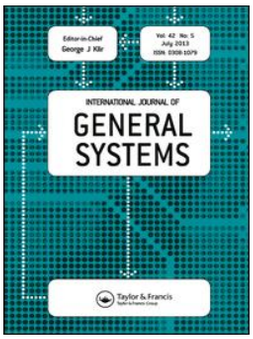

International Journal Of General System

ISSN: 0308-1079 (Print) 1563-5104 (Online) Journal homepage: www.tandfonline.com/journals/ggen20

Letter to the Editor

To cite this article: (1980) Letter to the Editor, International Journal Of General System, 6:2, 101-104, DOI: 10.1080/03081078008934785

To link to this article: https://doi.org/10.1080/03081078008934785

Published online: 19 Oct 2007.

Submit your article to this journal

Article views: 11

View related articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ggen20

Int. J. General Systems, 1980, Vol. 6, pp. 101–104
0308-1079/80/0602-0101 $04.50/0

© Gordon and Breach Science Publishers Inc.,
Printed in Great Britain

Letter to the Editor

IDENTIFICATION OF STOCHASTIC SYSTEMS—reply to A. M. Andrew

Sir—The critical comments made by Andrew on the sequential modelling systems described by Gaines[1] and Witten[2] are based upon a misunderstanding of the way in which the former algorithm operates and the objective of the latter algorithm. It is a significant feature of Gaines' system that it does *not* have the problem Andrew describes, and it is worthwhile to discuss why and to show this in terms of his example. Witten's procedure aims to model *non-deterministic* automata, where only the occurrence or non-occurrence of a transition is considered significant, rather than *probabilistic* automata where the transition frequencies are significant too. We conclude our discussion with a comparison of the two procedures which highlights the important differences between them.

Our studies of stochastic modelling are triggered by the work of Andreae[3,4] who has remarked over several years on the difficulty of performing gedanken experiments on learning systems. Such systems often exhibit surprising behaviour and aslo often solve problems which for some reason one thinks they cannot. In view of this we decided to answer Andrew as a two-stage process: firstly by giving a gedanken experiment view of how Gaines' system should cope with his example, and secondly by running test data from his example through the programs to see how they actually coped.

The essence of Andrew's example is that the sequences generated by the automata of his Figure 1 have the form ((A|B) (C|D))*, which may be generated by the two-state automaton of his Figure 2. However, Figure 2 does not, as he remarks, adequately represent the statistics of the sequences generated as opposed to their structure.

In fact Gaines [Ref. 1, p. 162] gives a similar example in which his program ATOM infers a 3-state model and a 5-state model for the same sequence which do not differ in the structure of the sequences generated, (BAA|BA)*, but only in that the 5-state model better approximates the statistics. Thus he already notes that ATOM can cope with just the problem that Andrew raises.

The reasons why ATOM should cope with such situations are also detailed in the paper. The use of a logarithmic error function is precisely to lead to probabilistic matching [Ref. 1, pp. 153–154]. Arguments are also advanced [Ref. 1, section 3.3.2] to show that if the behaviour is recurrent then the reduced form of that part of the automaton producing it will be precisely matched. Thus ATOM should have no problems with Andrew's example. Probably the easiest way of seeing this is in terms of the "surprise" patterns for Andrew's example. After an "A", ATOM will have a low surprise at a "D" ($-\log 0.9$) and a high surprise at a "C" ($-\log 0.1$), whereas this surprise pattern is reversed after a "B". Since ATOM minimizes its total surprise it will find the two-state model of Figure 2 inadequate since it involves no discrimination between A and B.

Andrew's argument fails for ATOM because it searches *all* possible structures [Ref. 1 p. 160] and outputs as admissible models those whose surprise is not greater than others of the same complexity in terms of numbers of states. Witten's length-k modelling procedure, on the other hand, does not search through all possible structures, nor does it perform multiple passes on the behaviour string. It simply scans through the input once, incrementally updating the canonical model either by augmentation or reduction—the latter occurring when evidence of non-determinism has been found which leads to a simplification of the model. The only criterion which is used to judge possible models is the occurrence or non-occurrence to length-k strings in the in-put, not their frequency [Ref. 2, p. 2]. To obtain probabilistic models one must either perform multiple scans (or, equivalently, postpone any reduction until all of the input has been seen) in order that frequency counts can be made; or use sequential probability estimation techniques. In the latter case, since the structure of the model will affect the statistics that can be gathered, there will be a dilemma between identification (wait until more of the behaviour has been seen before making statistical judgements) and reduction (reduce the model so that the statistics

which are gathered from now on are based on the most likely model). This is similar to the "two-armed bandit" dilemma between identification and control—a rather difficult problem.[5] Since the incremental length-k technique[2] is designed to operate in a single pass over the input, without storing it, the problem is ducked and the statistics of the behaviour sequence are ignored.

Figure A shows the outputs from ATOM (part (a)) and the incremental length-k modeller (part (b)), for a behaviour sequence from Andrew's Figure 1 automation. A much longer sequence was used for length-k modelling than for ATOM; the former needs more evidence of nondeterminism while the latter runs extremely slowly, even with short input sequences. ATOM produces optimal probabilistic models for 1 state, 2 states, and so on. We chose to display Mealy models, and Andrew's structure of his Figure 1 appears as the best 3-state model; whilst Figure 2 is the best 2-state one. The enormous size of the search space is indicated by the number of models considered. The length-k technique produces only one model for a given value of k: results are shown for k=2, 3, 4, and 5. Since it makes Moore models, it finds the structure of Andrew's Figure 3, and this is indeed an accurate model. However, as he points out, more complicated examples can be found which confound Moore modelling as well as Mealy modelling: the length-k technique will fail on these in the way that he describes. When k grows beyond 4, more complex models than Andrew's are formed because the behaviour sequence does not contain all possible k-tuples that can be generated by the source of Figure 1. The automation model grows initially and then, if the source is indeed nondeterministic, shrinks as the behaviour sequence continues to unfold and reveals that a smaller one will do. In this case, a maximum model size of 11 states was reached before the extent of nondeterminism became apparent, in the length-4 case.

To summarize the difference between the two procedures, ATOM searches the model space, whereas the length-k technique augments or reduces a single model. While ATOM can guarantee an optimal model under any criterion of goodness-of-fit, length-k model only optimizes within the length-k criterion of goodness-of-fit. Naturally, to compensate for this lack in power, the length-k technique is very substantially faster than ATOM, which has a practical limit of 10- or 11-state models with current technology.

BRIAN R. GAINES
IAN H. WITTEN
Man-Machine Systems Laboratory,
Department of Electrical Engineering Science,
University of Essex,
Colchester C04 3SQ, U.K.

```
ATOM:  file: andrew.200 containing 200 tokens;  delim='';  inputs = {}

ATOM: Mealy model with 1 states (4 links): plogp = 1.9981
-> 1: b->1(53) d->1(48) a->1(47) c->1(52)

1 models with 1 states [time: 0 min 10 sec]

ATOM: Mealy model with 2 states (4 links): plogp = 0.9981
-> 1: b->2(53) a->2(47)
   2: d->1(48) c->1(52)

169 models with 2 states [time: 0 min 6 sec]

ATOM: Mealy model with 3 states (6 links): plogp = 0.6815
-> 1: b->2(53) a->3(47)
   2: d->1(4) c->1(49)
   3: d->1(44) c->1(3)

15251 models with 3 states [time: 3 min 37 sec]
```

FIGURE 1(a). Models for a behaviour sequence from Andrew's automaton: from ATOM.

```
ILK: file: andrew.3000 containing 3000 tokens;  delim=''

ILK: length-2 model with 4 states, 8 links [maximum used: 4 states, 8 links]
-> 1: b:   ->2  ->4
>> 2: d:   ->1  ->3
   3: a:   ->2  ->4
   4: c:   ->1  ->3

predictions: none 5(0.2%); unique correct 16(0.5%); unique incorrect 1(0.0%);
             non-unique correct 2978(99.3%); non-unique incorrect 0(0.0%)
[time: 0 min 7 sec]

ILK: length-3 model with 4 states, 8 links [maximum used: 8 states, 14 links]
>> 1: d:   ->2  ->3
-> 2: b:   ->1  ->4
   3: a:   ->1  ->4
   4: c:   ->2  ->3

predictions: none 10(0.3%); unique correct 76(2.5%); unique incorrect 3( 0.%);
             non-unique correct 2911(97.0%); non-unique incorrect 0(0.0%)
[time: 0 min 7 sec]

ILK: length-4 model with 4 states, 8 links [maximum used: 11 states, 19 links]
>> 1: d:   ->4  ->3
   2: c:   ->4  ->3
   3: a:   ->1  ->2
-> 4: b:   ->1  ->2

predictions: none 19(0.6%); unique correct 113(3.8%); unique incorrect 8(0.3%);
             non-unique correct 2860(95.3%); non-unique incorrect 0(0.0%)
[time: 0 min 8 sec]

ILK: length-5 model with 7 states, 13 links [maximum used: 17 states, 29 links]
>> 1: d:   ->5  ->4
-> 2: b:   ->3  ->6
   3: d:   ->4  ->7
   4: b:   ->1  ->6
   5: a:   ->1  ->6
   6: c:   ->5  ->2
   7: a:   ->1

predictions: none 36(1.2%); unique correct 185(6.2%); unique incorrect 16(0.5%);
             non-unique correct 2763(92.1%); non-unique incorrect 0(0.0%)
[time: 0 min 9 sec]
```

FIGURE 1(b) Models for a behaviour sequence from Andrew's automaton: the length-4 model.

REFERENCES

1. B. R. Gaines, "System identification, approximation and complexity." *International Journal of General Systems*, **3**, 1977, pp. 145–174.
2. I. H. Witten, "Approximate, non-deterministic modelling of behaviour sequences." *International Journal of General Systems*, **5**, 1979, pp. 1–12.
3. J. H. Andreae, *Man-Machine Studies Reports UC-DSE/1 through 14*. University of Canterbury, Christchurch, New Zealand, 1972–79.
4. J. H. Andreae, *Thinking with the teachable machine*. Academic Press, London, 1977.
5. I. H. Witten, "The apparent conflict between estimation and control." *Journal of the Franklin Institute*, **301**, 1976, pp. 161–189.