TABLE I
TIMING FOR MAP RESTORATION

Matrix Size	Time per MAP Iteration
128	8.1 s
256	12.8 s
512	88.5 s

speeded up by the use of the FFT, but it is still the limiting computation factor. Other computations, as multiplications for the weights and S_b^{-1}, and computations of S_b^{-1}, also add to this basic time. There is also significant time required for input/output of the several large arrays which must be stored on disk.

The time of an iteration is almost singly a function of the size of the required convolution. These convolutions were done in a size to avoid as much as possible the periodic effects of the FFT. If the image is $M \times M$ and the point spread function is $N \times N$, the FFT multiplication is carried out on matrices $2^k \times 2^k$ where $2^k \leq M + N - 1 > 2^{k-1}$. The timing values given in Table I are functions of the matrix size required for the convolution. The times represent computation done on a CDC-7600 computer.

REFERENCES

[1] B. R. Hunt, "Bayesian methods in digital image restoration," *IEEE Trans. Comput.*, vol. C-26, no. 3, pp. 219–229, Mar. 1977.
[2] T. S. Huang, "Some note on film grain noise," in *Restoration of Atmospherically Degraded Images*, vol. 2, Woods Hole Summer Study, pp. 105–109, July 1966.
[3] E. F. Haugh, "A structural theory for the Selwyn granularity coefficient," *J. Photograph. Sci.*, vol. 2, pp. 65–68, 1963.
[4] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," *IEEE Trans. Syst., Man, Cybern.*, to be published.
[5] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*. New York: Wiley, 1966.
[6] H. J. Trussell, "Improved methods of maximum *a posteriori* image restoration," Ph.D. dissertation, University of New Mexico, New Mexico, NM, 1976.
[7] M. Cannon, "Blind deconvolution of spacially invariant image blurs with phase," *IEEE Trans. Acoust., Speech, Signal. Proc.*, vol. ASSP-24, pp. 58–63, Feb. 1976.

Maryanski's Grammatical Inferencer

B. R. GAINES

***Abstract*—A recent paper in this TRANSACTIONS[1] on an automatic inference procedure for probabilistic grammars contained major anomalies between the technique proposed and the computational examples given. These would mislead others attempting to reproduce the results and also undermine the apparent validity of what is in fact a sound technique. This letter details and corrects the anomalies.**

***Index Terms*—Grammatical inference, probabilistic grammars.**

I. BACKGROUND

Recently Maryanski and Booth[1] published an exposition of a probabilistic inference technique originally proposed in Maryanski's thesis [1]. Unfortunately, the published account is anomalous because an erroneous formula used in the thesis is corrected but the computational examples given are unchanged. This letter details and rectifies the anomalies in the paper, not in any critical spirit, but because: 1) there is a healthy tendency for published data on inference problems to be used by other workers to make comparative tests of techniques; and 2) Maryanski's use of the chi-square statistic as a criterion is a significant innovation in work on grammatical inference that results in more accurate modeling than his published (erroneous) results would suggest.

Manuscript received November 2, 1977; revised April 24, 1978.

The author is with the Department of Electrical Engineering Science, University of Essex, Colchester, England.

[1] F. J. Maryanski and T. L. Booth, "Inference of finite-state probabilistic languages," *IEEE Trans. Comput.*, vol. C-26, pp. 521–536, June 1977.

II. THE ANOMALIES

The actual structure of the anomalies is complex because they are also obscured by some typographical errors. They are as follows.

1) Maryanski, in his thesis, used the chi-square statistic inappropriately, defining it as

$$\chi^2 = \sum_{i=1}^{u} (f_i - Mp(x_i))^2/f_i \tag{1}$$

rather than

$$\chi^2 = \sum_{i=1}^{u} (f_i - Mp(x_i))^2/(Mp(x_i)) \tag{2}$$

using the terminology of his later joint paper with Booth.[1] Because of this, the experimental results in his thesis are incorrect although, since the two definitions are related in value, they do not appear meaningless. The overall effect of using (1) is to depress the estimates of chi-square so that he accepts grammars that are not very close to those generating the data.

2) In the published paper, this error of definition is corrected but the results given are still taken directly from the thesis so that the values of "chi-square" are those of (1) and *not* of definition (2) actually given in the paper. Because of this, the actual grammars given as selected by the technique described are, as noted, very much simpler than they should be. The two-state model inferenced for a five-state source in example G_1 of the paper (G_3 in the thesis) is misleading to anyone evaluating the results. *It would not be accepted on the basis of the technique described in the paper.*

3) An error of omission in the paper that does not show up because the chi-square values have not been recalculated is that the summation in (2) extends not just over the f_i of the sample but also over the (zero) f_i of words *predicted by the grammar which do not occur in the sample.* It is perhaps simpler to regard these as a "correction term" to a version of (2) in which the summation is over the sample only:

$$\chi^2 = \sum_{i=1}^{u} (f_i - Mp(x_i))^2/(Mp(x_i)) + F \tag{3}$$

where F is the total predicted frequency of occurrence of words not in the sample. This was a clue that Maryanski missed to the error in (1) since zero values of f_i cause (1) to "blow up" but not (2).

4) There was a typographical error in Maryanski's thesis such that the "chi-square" value (incorrect definition 1) was reported as 1.582 when it is actually 15.82. This has been transferred to the paper.

5) There is a typographical error in the paper in reporting the sample set used for G_1. The table at the beginning of Section A. *Test Grammar 1*, on page 532 has one sample missing: word 011 with a count of 3. This table is correct in the thesis and was used to derive the result given in the paper.

6) However, the actual sample set for G_1 is itself anomalous in both thesis and paper. The counts associated with the words deviate so far from the probabilities in the definition of the grammar as to make it a biased sample. It is not clear how the samples were generated and whether this is statistical fluctuation or an error—

0018-9340/79/0100-0062$00.75 © 1979 IEEE

no such anomaly arises in the other samples in the thesis. It would have been immediately apparent that this was an unreasonable sample for inference purposes had Maryanski used the correct definition of chi-square, since the sample set given is significantly different from its own grammar!

The remainder of this letter details the effects of these anomalies, and their correction, on the results given.

III. Reanalysis of Maryanski's Data

A. Test Grammar 1

First define a new derived grammar G_1^A which is that obtained by taking the structure of the grammar G_1 but estimating the rule probabilities from the (full) word sample given. This turns out to be

G_1^A
$0.10: B \to 1$ $0.73: C \to 1D$
$0.40: B \to 0A$ $0.62: D \to 1$
$0.50: B \to 1E$ $0.07: D \to 0C$
$0.15: A \to 0C$ $0.31: D \to 1D$
$0.85: A \to 1D$ $1.00: E \to 1D$
$0.27: C \to 0C$

The chi-square values [from (3)] for the two derived grammars are

$$G_1^I: \chi^2 = 128.1 \quad \text{(0.05 acceptance level 18.3, 10 d.f.)}$$

$$G_1^A: \chi^2 = 55.9 \quad \text{(0.05 acceptance level 15.5, 8 d.f.)}.$$

Thus, not only would the technique described reject the derived two-state grammar reported in the paper but it would also reject the derived five-state grammar with the same structure as the actual generator!

This discrepancy, if noted, would have led to a check on the data which are meant to be a fair sample of the output of the probabilistic grammar G_1. In fact it is a most peculiar sample with the words: 1 in at twice its expected frequency; 011 at one and one-half times expectation; 0011 at one-quarter expectation; and the high probability words 00011, 01011, 01111, and 001011 are missing, when very much lower probability words are included. While clearly, such a sample might arise in a "real" inferencing situation, its statistically significant deviation from the underlying grammar makes it an unsuitable example from which to draw conclusions about the accuracy of an inferencing technique.

A better 50-word sample with frequencies far closer to their expected values is

3:1	1:01011
15:111	1:01111
8:011	1:110011
4:0011	1:001011
4:1111	1:110111
2:0111	1:111011
2:11011	1:000011
1:00011	1:000111
1:00111	1:010011
1:11111	

The derived grammar with the same structure as G_1 for this sample is

G_1^B
$0.06: B \to 1$ $0.74: C \to 1D$
$0.44: B \to 0A$ $0.68: D \to 1$
$0.50: B \to 1E$ $0.12: D \to 0C$
$0.41: A \to 0C$ $0.20: D \to 1D$
$0.59: A \to 1D$ $1.00: E \to 1D$
$0.26: C \to 0C$

which is rather closer than G_1^A to G_1 and has a chi-square value against the new sample of

$$\chi^2 = 18.4 \text{ (0.05 acceptance level 21.0, 12 d.f.)}$$

so that it is accepted.

Using Maryanski's algorithm on this sample, the actual derived grammar has four states:

G_1^C
$0.06: B \to 1$ $0.68: D \to 1$
$0.44: B \to 0C$ $0.12: D \to 0C$
$0.50: B \to 1E$ $0.20: D \to 1D$
$0.33: C \to 0C$ $1.00: E \to 1$
$0.67: C \to 1D$

with

$$\chi^2 = 19.3 \text{ (0.05 acceptance level 22.4, 13 d.f.)}.$$

This is essentially G_1 with symbols A and C equivalenced, a not unreasonable reduction since both structure and transition probabilities are little affected. A statistical discrimination between G_1 and G_1^C requires a great deal of data, and certainly for up to 1000-word samples, G_1^C is the grammar that an inferencer based on any of the current techniques will infer. It is the same in structure as the second grammar that is derived in the paper from an (unreported) 1000-word sample.

B. Test Grammar 2

The data sample for the second grammar G_2 (in the paper, G_5 in the thesis), is a reasonable one. However, again the calculation of chi-square being wrong leads to the reported derivation of a grammar that would actually be rejected. The value for G_2^I is not 6.00 as reported, but instead

$$G_2^I: \chi^2 = 22.9 \text{ (0.05 acceptance level 9.5, 4 d.f.)}.$$

The actual grammar acceptable by this criterion is a three-state one:

G_2^A
$1.00: B \to 1C$ $0.19: C \to 1C$
$0.70: C \to 1$ $0.65: D \to 1$
$0.11: C \to 0D$ $0.35: D \to 0D$

with

$$\chi^2 = 6.7 \text{ (0.05 acceptance level 7.8, 3 d.f.)}.$$

If we tighten up the criterion slightly to an acceptance level of 0.10, then this grammar also is rejected and the derived grammar becomes

G_2^B
$1.00: B \to 1C$ $0.65: D \to 1$
$0.68: C \to 1$ $0.35: D \to 0D$
$0.13: C \to 0D$ $0.79: E \to 1$
$0.19: C \to 1E$ $0.21: E \to 1E$

with

$$\chi^2 = 2.6 \text{ (0.10 acceptance level 4.6, 2 d.f.)}$$

which has the structure of G_2^D in the paper, the actual four-state unambiguous version of the grammar G_2.

IV. Discussion

The reanalysis of two of Maryanski's examples given here demonstrates that the technique suggested in his thesis and reported in the paper works rather better than the published results suggest. On the first example it leads to a four-state model of a five-state grammar on the basis of a 50-word sample, and this is the best

result that can be expected with such a small sample of what is a statistically ambiguous grammar so far as the distinction between four and five states is concerned. On the second example, the technique leads to a three-state model of a four-state grammar on a fairly weak statistical test, and to the correct four-state model on a slightly stronger one.

It is clearly unreasonable to expect a probabilistic grammatical inference procedure to derive firm conclusions from any but virtually infinite samples, and statistical decisions on small samples will always be open to variation with criteria. I have given a full discussion of this phenomenon in [2]—in terms of the terminology of that reference, for example, both the three- and four-state models of G_2 are *admissible*. What one asks of a probabilistic grammatical inference technique is that the actual grammar generating the data does at least appear in the admissible set inferenced for reasonable sized samples. The reanalysis of Maryanski's data given here and similar reanalysis of all the examples in his thesis demonstrates that his proposed use of a chi-square criterion, if applied correctly, does lead to a viable inferencing technique.

There is a feature of the chi-square criterion that goes beyond this basic requirement. It does actually discriminate in a reasonable way between admissible grammars (those such that all grammars of a smaller size are a worse fit to the sample). As the grammar inferenced gets bigger, so the degrees of freedom decrease and hence the chi-square criterion gets tighter. Hence the criterion also provides a *decision-procedure* for selecting amongst admissible grammars. Horning [3] achieves a similar criterion by postulating an absolute probability of a grammar (derived from his *grammar-grammar*) that declines as the size of the grammar increases. His technique [4], and its variants [5], also give viable grammatical inferencers.

Finally, it is worth noting that grammatical inference is a problem well-suited to small machines. The published core occupancies for Maryanski's technique are a function of the compiler, etc., not basic requirements. The ATOM inferencer [2], [6] uses a different search strategy than that of the published paper and will accept a range of criteria for approximation amongst which Maryanski's chi-square had been included (from his thesis but in corrected form). To obtain the results in this paper, ATOM was run as a Fortran program on an 1si11 microprocessor in less than 8K 16-bit words. Most of the program is concerned with nice printouts and the actual algorithm could be machine coded in a few hundred words. Timings are comparable to those given in the paper for a much larger processor: G_1^I (best two-state) 5 s; G_1^C (best four-state) 120 s; G_2^I (best two-state) 0.4 s; G_2^A (best three-state) 2.4 s; G_2^B (best four-state) 10 s.

References

[1] F. J. Maryanski, "Inference of probabilistic grammars," Ph.D. dissertation, Univ. Connecticut, Storrs, CT, July 1974, Univ. Microfilms 75-10, 645.

[2] B. R. Gaines, "System identification, approximation and complexity," *Int. J. Gen. Syst.*, vol. 3, pp. 145–174, 1977.

[3] J. J. Horning, "A study of grammatical inference," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1969, Univ. Microfilms 70-10, 465.

[4] A. Van der Mude and A. Walker, "On the inference of stochastic regular grammars," Dep. Comput. Sci., Rutgers Univ., New Brunswick, NJ, 1977, to be published.

[5] C. M. Cook, A. Rosenfeld, and A. R. Aronson, "Grammatical inference by hill-climbing," *Inform. Sci.*, vol. 10, pp. 59–80, 1976.

[6] B. R. Gaines, "Behavior/structure transformations under uncertainty," *Int. J. Man-Machine Stud.*, vol. 8, pp. 337–365, May 1976.

Authors' Reply[2]

FRED MARYANSKI AND TAYLOR L. BOOTH

The authors would like to thank B. R. Gaines for his careful reading of both the paper[1] and the thesis [1]. Dr. Gaines is correct in pointing out that for application to the grammatical inference problem

$$\chi^2 = \sum_{i=1}^{u} (f_i - MP(X_i))^2/(MP(X_i) + F$$

where

$$F = M^*\left(1 - \sum_{i=1}^{u} p(X_i)\right).$$

We have found, as Gaines has stated, that making this correction to the program does in fact yield more accurate models. We also concur with Dr. Gaines as to the two typographical errors pointed out in test grammar 1.

However, we are not in agreement with Gaines' claim that the 50 word sample set for grammar G_1 is anomalous. The basic tenet of grammatical inference is that the inference program is presented with a random sample from the grammar to be inferred [2]. The 50 word sample set for G_1 was produced by a program which exercised G_1 in a random manner. The fact that the structure of G_1 could not be inferred from the sample implies that the sample was too small; not that it is incorrect in anyway. This fact is discussed in our paper[1] in which the inference procedure using a 1000 word sample is shown to produce an inferred grammar very close to G_1.

In Gaines' letter, a 50 word sample more representative of the structure of G_1 is presented. This sample yields an inferred grammar closer to G_1 than the inferred grammar in our paper.[1] However, the use of Gaines' sample violates the basic rule of grammatical inference that the sample be taken at random from the grammar. The example presented in our paper[1] is intended to demonstrate the limitations of probabilistic grammatical inference when small random samples are supplied.

In summary, we concur with the corrections put forth by Gaines, with the exception noted above. However, we feel that the anomalies, while unfortunate, do not detract from the significance of the grammatical inference technique presented in our paper. We once again thank Dr. Gaines for his interest in and use of our chi-square criterion and for his careful reading of our paper.

References

[1] F. J. Maryanski, "Inference of probabilistic grammars," Ph.D. dissertation, Comput. Sci. Group, Dep. Elec. Eng. and Comput. Sci., Univ. Connecticut, Storrs, CT, July 1974.

[2] K. S. Fu and T. L. Booth, "Grammatical inference: Introduction and survey—Part I," *IEEE Trans. Syst., Man, Cybernetics*, vol. SMC-5, pp. 95–111, Jan. 1975.

[2] Manuscript received April 24, 1978.

F. Maryanski is with the Department of Computer Science, Kansas State University, Manhattan, KS 66506.

T. L. Booth is with the Department of Electrical Engineering and Computer Science, University of Connecticut, Storrs, CT 06268.

0018-9340/79/0100-0064$00.75 © 1978 IEEE