Twelve Papers

R. Jones

Contents

Experiments with Asa H

R. Jones

Departments of Physical Science

Emporia State University

Emporia, KS 66801

Abstract

We test the abilities of our Asa H software agent by applying it to problems in classification, robotics, protolanguage learning, and control. Asa is successful in creating/learning a description of each of these application environments.

Introduction

Our recently developed "Asa H" software architecture (hierarchical autonomous software agent) consists of a hierarchical network of clustering modules. (Jones, 2006 and 2007) The first (one or more) layers of the hierarchy comprise an array of classifiers which accepts inputs at each time step and compares them with stored categories. Each higher layer of classifiers accepts the matches output from the layer below as its own input vector and is implemented in the same way creating a self-adaptive hierarchical set of categories. (Fig. 1)

As each category at each time step is evaluated and relayed from the lowest layers of the hierarchy the developing pattern is compared with stored temporal sequences. The best matching sequences output a set of "value measures." The topmost module in the hierarchy combines all of the values input from the sequence classifiers below it and produces an estimated utility. Upper layers in the hierarchy correspond to higher levels of abstraction. This architecture bears some resemblance to Amit's recent work (Amit and Mataric 2002), to Selfridge's "pandemonium" (Selfridge 1958), and to Kohonen's feature maps (Kohonen, 1958). However, many of the algorithms and learning methods are original. (Jones, 2006)

Asa H

The lowest layer(s) in our architecture is a set of classifiers which accepts inputs at each time step, **In**, and compares them with stored categories. As each category at each time step is passed on from the categorizing subsystem the developing temporal sequence is compared with stored sequences (and time dilated, compressed, and shifted sequences (Agrawal, et al. 1995)) in the next higher layer(s) of the network.

Additional layers of sequence learners are stacked one above the other thus learning a hierarchically structured "sequence of sequences," each covering a longer time period then the one before. The idea is to use the hierarchical organization to reduce computational complexity, each layer corresponding to a higher level of abstraction than the one below it. The top most sequence classifying layer outputs a set of predicted "value measures" to the final module in the hierarchy. These "values" are things like system "health", damage, energy supply, "foresight," speed, etc and are detectable in the input signal by hard wired (hand coded) feature detectors. (The top layer of sequence classifiers learns to predict the values that will be attained later in time.) "Foresight," for example, is measured based upon the sequence's ability to predict the next one or more inputs, **In**(t+1), **In**(t+2), etc. (judged by a vector dot product or other similarity measure).

The final element in the hierarchy (Fig. 1) combines all the value measures passed up to it from the sequence classifiers below and produces an estimated utility. Prospective actions can be treated the same as (occasional) predictions and can consist of innate predefined categories (action primitives) and sequences. At each time step that action (or inaction) is chosen which maximizes the predicted ultimate utility given the recent context (input history). Full details of Asa and its learning mechanisms are given in Jones (2006)

Testing Asa H

In the present paper we explore Asa H's capabilities across a broad range of cognitive tasks and applications. While doing this we are also suggesting some practical uses for Asa software. In some cases learning is faster and performance is better than that achieved by humans.

Simulating Neural Networks

We first tested Asa on simple Boolean function learning: AND, OR, XOR, etc. On a single exposure to the XOR training set, for instance, Asa learns the categories:

C (1,1) = .707

C (1,2) = .707

C (2,3) = -.999

C (3,1) = .707

C (3,2) = -.707

C (4,3) = .999

C (5,1) = -.707

C (5,2) = .707

C (6,1) = -.707

C (6,2) = -.707

And the sequences relating those categories:

S (1,1,1) = .707

S (1,2,2) = .707

S (2,3,1) = .707

S (2,4,2) = .707

S (3,5,1) = .707

S (3,4,2) = .707

S (4,6,1) = .707

S (4,2,2) = .707

(Asa H 2.0's representation is more efficient but here we use 1.0 representation for its simplicity.) Upon subsequent exposures to the inputs (or some noisy inputs that are reasonably similar) Asa responds with predictions (output):

(Input):(Output);(1,1):(-1); (1,-1):(1);(-1,1):(1); (-1,-1):(-1)

On exposure to the AND training set Asa learns the categories:

C (1,1) = .707

C (1,2) = .707

C (2,3) = .999

C (3,1) = .707

C (3,2) = -.707

C (4,3) = -.999

C (5,1) = -.707

C (5,2) = .707

C (6,1) = -.707

C (6,2) = -.707

And the sequences relating those categories:

S (1,1,1) = .707

S (1,2,2) = .707

S (2,3,1) = .707

S (2,4,2) = .707

S (3,5,1) = .707

S (3,4,2) = .707

S (4,6,1) = .707

S (4,4,2) = .707

On exposure to the OR training set Asa learns the categories:

C (1,1) = .707

C (1,2) = .707

C (2,3) = .999

C (3,1) = .707

C (3,2) = -.707

C (4,1) = -.707

C (4,2) = .707

C (5,1) = -.707

C (5,2) = -.707

C (6,3) = -.999

And the sequences relating those categories:

S(1,1,1) = .707

S(1,2,2) = .707

S(2,3,1) = .707

S(2,2,2) = .707

S(3,4,1) = .707

S(3,2,2) = .707

S(4,5,1) = .707

S(4,6,2) = .707

Where humans confuse exclusive disjunction and biconditionals and learn them slowly, if at all. (Hunt et al, 1966) Asa learns them readily; biconditional:

C (1,1) = .999

C (2,5) = .999

C (3,1) = .707

C (3,3) = .707

C (4,1) = .577

C (4,3) = .577

C (4,4) = .577

And the sequences relating those categories:

S (1,1,1) = .707

S (1,2,2) = .707

S (2,3,1) = .707

S (2,2,2) = .707

S (3,4,1) = .707

S (3,2,2) = .707

Exclusive disjunction:

C (5,2) = .999

C (6,2) = .707

C (6,4) = .707

C (7,2) = .577

C (7,3) = .577

C (7,4) = .577

And the sequences relating those categories:

S (4,5,1) = .707

S (4,2,2) = .707

S (5,6,1) = .707

S (5,2,2) = .707

S (6,7,1) = .707

S (6,2,2) = .707

Learning Mobile Robot Control

We next had AsaH control a robot in a simulator (Fig. 2). (Heiserman, 1981 and 1982and Jones, 2006) The complexity of the experiment is adjustable by modifications of the environment and by running over multiple generations. The value module is retrained after each run and accepts as inputs "damage", integrated energy input to the robot, and "foresight" and predicts utility u as output. Damage, energy input, foresight, and u are each extracted from raw data inputs by hand coded detectors. "Foresight", for instance, is measured based upon the sequence's ability to predict the next (one or more) inputs **In**(t). "Damage," on the other hand, is signaled when an action is commanded at time t but not detected in **In**(t+1).

Over successive runs AsaH learns to value (prefer) sequences which extend the robot's life expectancy and avoid those which lead to damage. Sequence classifiers that have been reliably learned by AsaH include:

High speed AND contact ⟶ collision

Collision ⟶ damage

Contact energy ⟶ recharge

Sense obstacle left ⟶ turn right

Sense obstacle right ⟶ turn left

Sense obstacle forward AND move forward ⟶ contact

Sense obstacle left AND turn left ⟶ contact

Sense obstacle right AND turn right ⟶ contact

Sense obstacle left AND turn left ⟶ sense obstacle forward

Sense obstacle right AND turn right ⟶ sense obstacle forward

Sense obstacle forward AND turn left ⟶ sense obstacle right

Sense obstacle forward AND turn right ⟶ sense obstacle left

Sense energy forward ⟶ move forward

Sense energy left ⟶ turn left

Sense energy right ⟶ turn right

recharge ⟶ charged

charged ⟶ back away

Where " ⟶ " means, basically, "followed by."

Chained sequences (longer sequences) are also formed such as:

Sense obstacle forward ⟶ turn left ⟶ sense obstacle right ⟶ turn left

And:

Sense energy right → turn right → sense energy forward → move forward → contact energy → recharge → charged → back away

with, of course, possible time warping.

Learning Protolanguage

We have experimented with linguistic input and output using binary alphanumeric coding. While learning mobile robot control the linguistic input "collision" is provided simultaneous with robot-obstacle contact. While moving forward the trainer says "moving forward". Asa learns to associate the action and the words. Similarly, a linguistic input "warning" is supplied concurrent with robot obstacle detection. After such training it is possible for an observer to input a cry "collision" and/or "warning" and evoke preemptive evasive actions from the robot. The chained input pair "warning", "collision" typically has a stronger influence on the robot behavior than either input alone. Asa has learned, for instance:

Hear "collision" → collision

Hear "warning" → damage

Hear "obstacle" → slow down

Hear "clear ahead" → speed up

Hear "turn left" → turn left

Hear "turn right" → turn right

Hear "stop" ⟶ stop

Sense obstacle ⟶ say "obstacle"

Slow down ⟶ say "slow down"

Turn right ⟶ say "turn right"

Turn left ⟶ say "turn left"

Recharge ⟶ say "recharging"

Collision ⟶ say "collision"

Move ⟶ stop ⟶ say "stop"

Stop ⟶ move ⟶ say "moving"

Chaining also produces phrases like "obstacle", "slow down", "turn left". Once a substantial amount of real world experience has been recorded we may try natural language I/O in multiple languages. World knowledge might then aid in translation between different input and output languages.

Vocabulary and Concepts

In an effort to move beyond protolanguage we have begun to teach Asa H standard sign language (Jones, 1968). So far Asa H has learned the vocabulary/concepts:

Collision

Warning

Obstacle

Clear

Turn

Stop

Move

Slow

Fast

Recharge

Damage

Rakison and Lupyan (2008) have demonstrated the early development of the concepts of causality, animate objects and inanimate objects in neural networks. Asa H is able to learn the same feature sets and sequences used by Rakison and Lupyan (2008, figure 16 and appendix).

Learning Ion Source Control

Our plasma ion source is a low-pressure hot filament arc (Jones, 1997). The operator is able to control the ionizing ("primary") electron emission from the filament cathode by adjusting V_f, the filament heating voltage. The primary electrons are accelerated and the discharge is sustained by the variable discharge voltage V_d. The gas fill pressure, P, is controllable by a servo actuated needle valve.

The operator is able to monitor the source operation by measuring the voltage V_f and V_d as well as the corresponding filament and discharge currents I_f and I_d and the neutral gas pressure (obtainable from thermocouple or ionization type vacuum gauges). The operator also monitors the real-time performance or output of the source by measuring the plasma density, N_e, and (electron) plasma temperature, T_e.

At each time step V_f, V_d, I_f, I_d, P, N_e, and T_e are measured and updated settings for V_f, V_d, and P are output. The sequence of this vector is sought which maximizes the time integral of ion output $N_eT_e^{1/2}$. Figure 3 illustrates the ion output attainable by a human operator (xs) and by Asa H as a function of time/learning (circles).

Self Monitoring

In an effort to fine tune the learning process in more complex environments we have added self monitoring inputs and self adjusting outputs. These inputs watch the time spent:

Taking input

Giving output

Searching the case base

Feature extraction

Adding to memory

Sorting

Comparing

Extrapolating

Deduction

Value assessment/simulation

And updating cases.

Outputs then adjust the amount of time spent on each of these operations in an effort to adapt the learning process to the environment seen. (Adjustments are based upon utility improvements measured.)

As a consequence of self monitoring, in rapidly changing environments input is accepted frequently, at the expense of extrapolation, etc. In slow changing but complex environments input is accepted less frequently and more extrapolation is performed.

Conclusion

Our Asa H software agent (Jones, 2006) has been successful in experiments involving classification, robotics, protolanguage learning, and control. In each case Asa learns/creates a suitable task description appropriate to the application environment.

It is possible for Asa to do far more selfmonitoring than people do. We find, however, that this can actually be a distraction and ultimately reduce problem solving efficiency. Again, time spent needs to be adjusted based on utility improvement.

Figure Captions

Figure 1: The hierarchical agent architecture.

Figure 2: Simulated robot, R, free to wander among obstacles, O, encountering energy sources, S. The space is assumed to be discrete and toroidal.

Figure 3: Total ion yield per run as function of training time.

Literature Cited

Agrawal, R., Lin, K., Sawhney, H.S. and Shim, K. 1995. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. Proceedings of the 21st VLDB Conference, Zurich, Switzerland.

Amit, R. and Mataric, M. 2002. Learning movement sequences from demonstration. Proceedings of the International Conference On Development and Learning. MIT, Cambridge, MA.

Heiserman, D.L. 1981. *Robot intelligence with experiments*. Tab books, Blue Ridge Summit, PA, 322 p.

Heiserman, D.L. 1982. *Projects in machine intelligence for your home computer.* Tab books, Blue Ridge Summit, PA, 337 p.

Hunt, E.B., Marin, J. and Stone, P.J. 1966. Experiments in Induction. Academic Press, New York, 247 p.

Jones, H. 1968. *Sign Language*. Hodder and Stoughton, London, 192 p.

Jones, R. 1997. An Intelligent plasma/ion source controller. Kansas Academy of Science Transactions 100 (3-4) p. 85-93

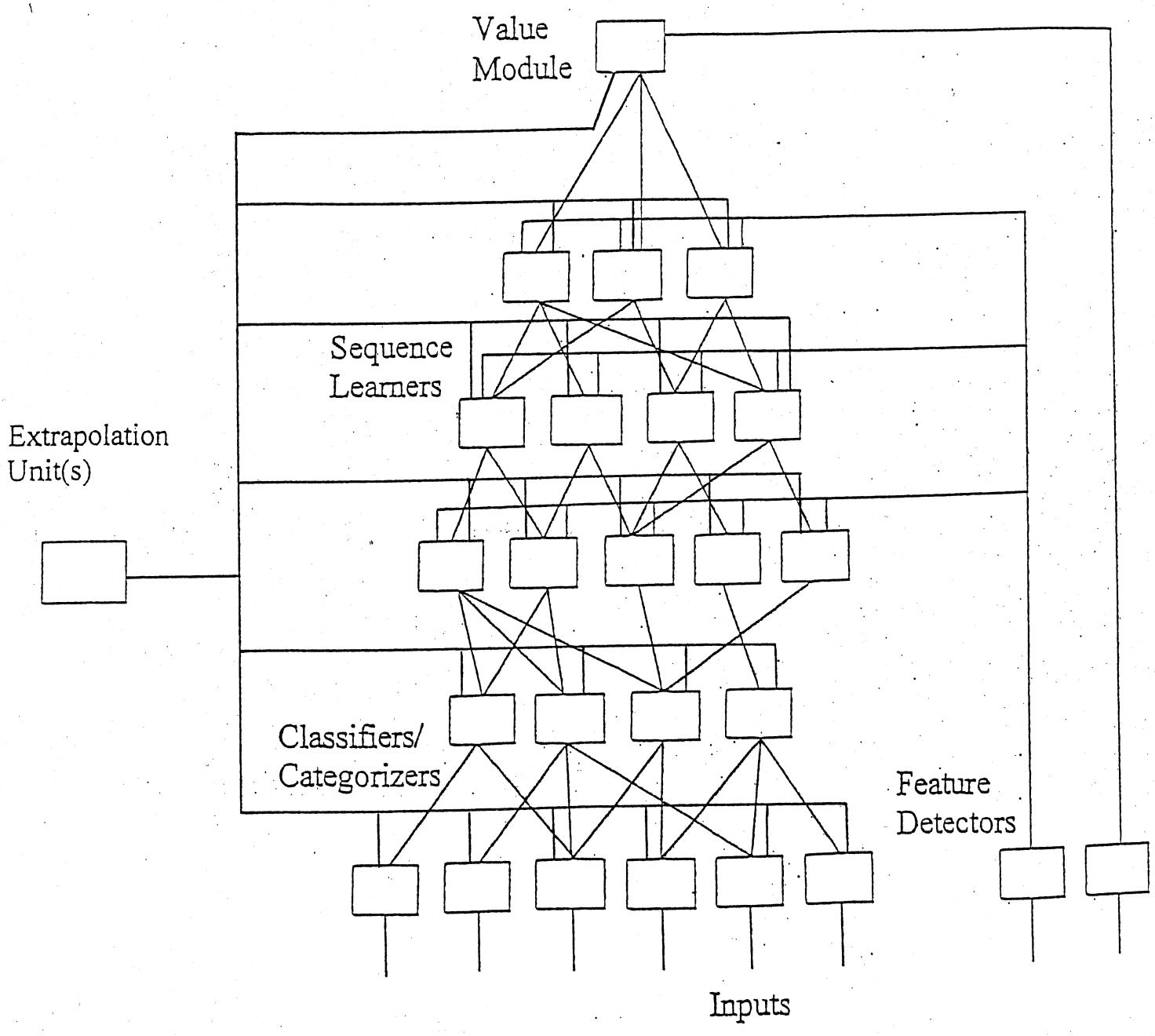
Jones, R. 2006. AsaH: A hierarchical architecture for software agents. Kansas Academy of Science Transactions 109 (3-4) p. 159-167

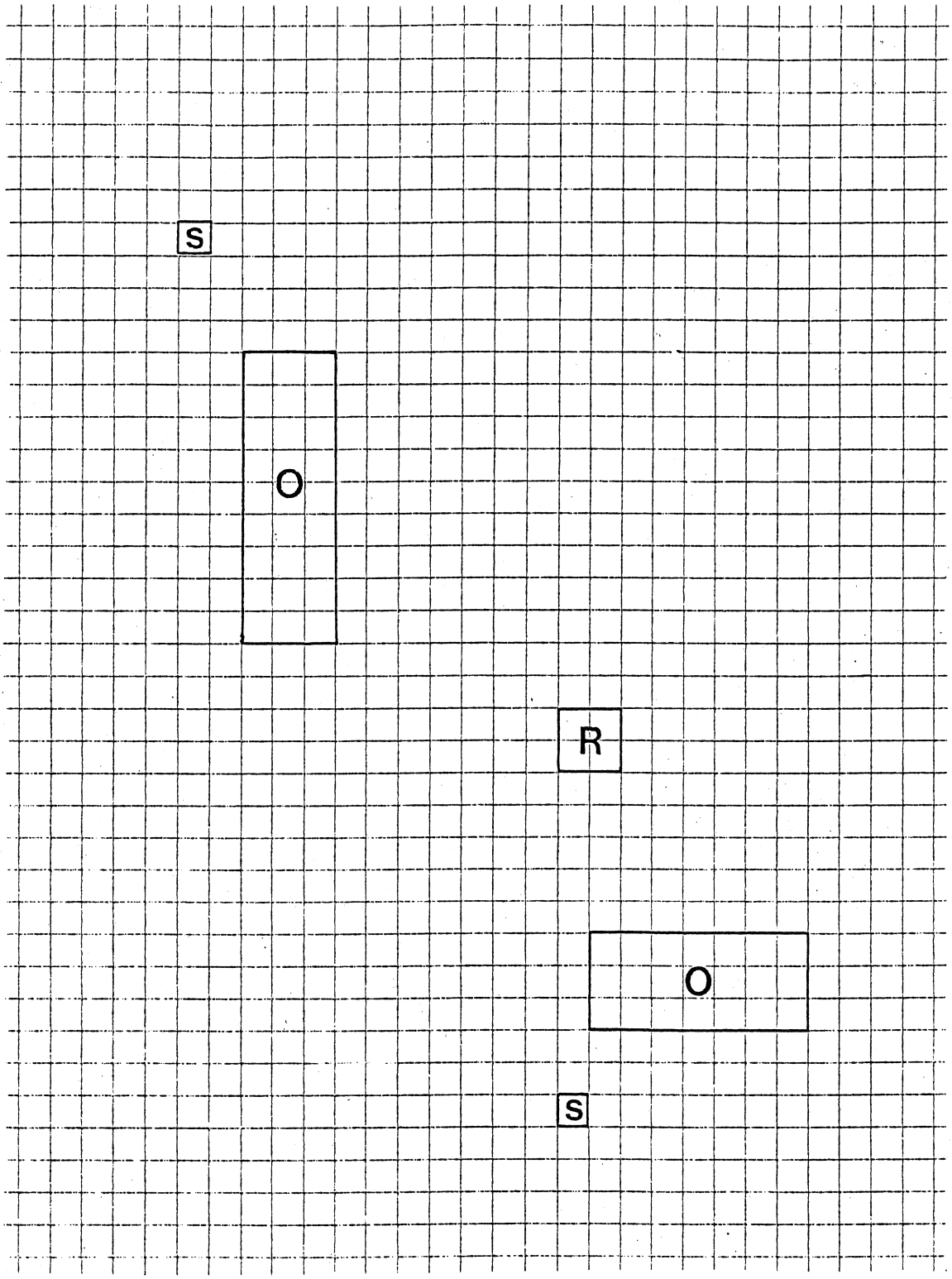
Jones, R. 2007 Integrating feature discovery algorithms into Asa software agents, Kansas Academy of Science Transactions 110 (3-4) p. 302

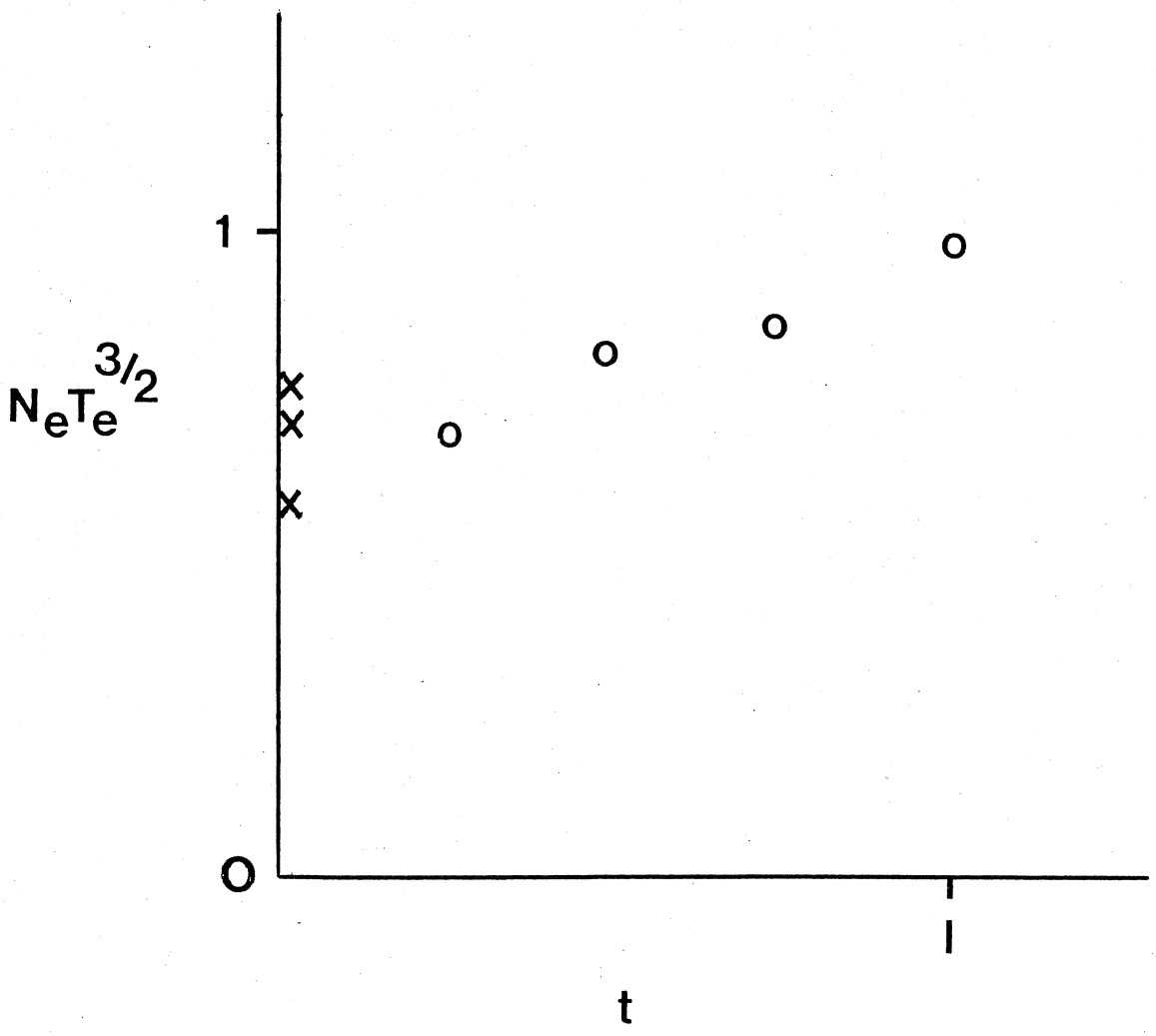
Kohonen, T. 1988 *Self Organization and Associative Memory*. Springer-Verlag, Berlin, 255 pp.

Rakison, D.H. and Lupyan, G. 2008. Developing Object Concepts in Infancy: An Associative Learning Perspective. Wiley-Blackwell, Boston, p 130.

Selfridge, O.G. 1958. Pandemonium: a paradigm for learning. In *Pattern Recognition*, Uhr, L. (ed.), John Wiley and Sons, New York, p 339-348.


Value
Module
Sequence
Learners
Extrapolation
Unit(s)
Classifiers/
Categorizers
Feature
Detectors
Inputs


S
O
R
O
S


1
$N_e T_e^{3/2}$
O
t
1

Capitalism is Wrong

R. Jones

Department of Physical Sciences

Emporia State University

Abstract

Its argued that capitalist economics has a theory of value which is not consistent with traditional human values and society's legal system.

We must begin by distinguishing two different sorts of quantities, scalars and vectors.

1. Scalars versus Vectors

A scalar is a quantity that can be described by just a number. Your age, or height or weight or the number of people in a room, these are all scalar quantities. A vector, on the other hand, is a quantity that requires two or more numbers in order to describe it. An example might be the journey (displacement) "3 blocks north and 5 blocks east." These two distances must remain distinct. Motion northward can not substitute for motion eastward or vice versa.

2. Value Monism versus Value Pluralism

Value monism is the idea that there is only one thing of value (pleasure perhaps?) or that at least it is possible to define a single "common currency" in terms of which the value of all things can be measured (money?) Value pluralism is simply the rejection of value monism, the idea that there are 2 or more things to be valued (e.g. your wife's love and your child's life) and that they can not be reduced to, or compared via, some common currency.

3. Utility

In capitalist economic theory a utility function maps any state of affairs (with its costs, rewards, opportunities, etc.) to a real number, a scalar. Frequently this scalar utility is simply money; in other situations it is some more generalized scalar quantity.

4. Needs

Human beings have multiple needs. We need air to breath, water to drink, and food to eat. (It should be noted that air is free, and water is often free, while food can be so expensive that many people starve to death.) To survive we have other needs as well. Many of these needs are "incommensurable," that is they can not be measured by a common standard; one can not be traded off for another. No amount of water can make up for having no food to eat. A plentiful supply of fresh air can not make up for a lack of water to drink.

5. Utility is a vector

Human values, in turn, arise from our needs and are also incommensurable. (Berlin, 1998, XVIII) Economic "utility" must be a vector quantity (Dasgupta, et al, 1999, and Liu, 1999, 6) it can not be a scalar "money". Business and capitalism are profoundly in error when they employ a mere scalar utility. A common currency is simply impossible. Keeney and Raiffa, state (1976, 19): "in complex value problems consequences . . . cannot be adequately described objectively by a single attribute (e.g. money)." It is not possible to put a dollar value on your wife's love or your child's life. These are examples of values we have which are incommensurable with money and with each other. They are separate components of the value VECTOR.

"Value monism" is wrong, the human value system is simply too complex to be boiled down to a single scalar (Pepper, 1998). You distort human values when you try to sum them all up as a single scalar. Just the instruction, "travel 8 blocks" is not adequate to represent "3 blocks north and 5 blocks east." When you try to put a price on love you

create prostitution. When you try to put a dollar value on a human being you begin slavery.

The "bottom line" can not accurately describe the worth of an enterprise. Business and capitalism are wrong in both the technical sense and the moral sense. Some economists defend scalar utility as simply an axiom. But a science does not get to try just ANY axioms. It only gets to use axioms that fit the facts. In the words of Beardon, et al "contrary to the widely held and inveterate belief of economists, there does exist a preference relation which is not representable by a utility function." (Beardon, et al, 2002) Human value systems are just more complex than business and capitalism allow for.

Von Neumann and Morgenstern (1944, 19-20) said "We have conceded that one may doubt whether a person can always decide which of two alternatives. . . he prefers. . . . It leads to what may be described as a many-dimensional vector concept of utility." But they evidently never published further details. In an effort to flesh out such a theory Dubra, Maccheroni, and Ok have published Expected Utility Theory without the Completeness Axiom (2001). Danan and Ziegelmeyer have reported on experiments which show that just such a theory is needed to successfully model human value judgments. I believe our everyday business and economic life must be modified to take account of these discoveries. Money can not be used as the sole yardstick.

A scalar simply contains too little information to do what capitalists want it to do. "Value pluralism" is required. If you insist on describing how successful a man is using a scalar then evolutionary biology tells us it should be something like U=(N-2)/L where

N are the number of offspring (children) you have and L is you lifespan. Notice how poorly money will correlate with this in the capitalist world.

A person's utility is really a vector, having components like fertility, life expectancy, IQ, etc. Suppose you wish to compare Jane and Mary. If Jane has all components of the utility vector greater that Mary then you can conclude that Jane is more successful that Mary. If, however, some components are greater for Jane while others are highest for Mary then no conclusion is possible. Wealth is just a component of such a utility vector. Assuming a scalar utility like money is an approximation. In simple problems it may work fine but in the real world it is inadequate. In comparing and choosing which college you wish to attend one school may have English, Economics and Math majors but no Earth Science major. A second college may have the Earth Science major, English, and Math majors but no Econ major. You just can not create a scalar utility that says which college is "better than" the other. Von Neumann admitted that utility might have to be a vector and was making an approximation when he assumed it to be scalar.

Business and capitalism require value monism while the human value system is characterized by value pluralism. If you have any doubt that human society exhibits value pluralism you need only observe the legal system. Upon being found guilty of some crimes it is only required that you pay a fine, where as in other cases you must pay with your life, or at least some years worth. No one would be satisfied to see a serial killer merely pay a fine each time he took another life. As Arrow says (1997, 757) "Regardless of our all-embracing market theories, we economists must recognize that there are goods that might be bought and sold but aren't . . . Judicial decisions and votes.

. . Use of the market and its language leads to results which offend our intuitions . . . The market is one system . . .Looking at policy issues from the point of any one system is likely to lead to unsatisfactory conclusions"

Occasionally a capitalist may admit that value pluralism is needed in real life and then claim that businessmen do consider more than just their treasured "bottom line". If there is any truth to this claim I would amend my criticism to say that capitalism makes too much use of value monism and too little use of value pluralism.

Value monism

I have argued elsewhere against value monism (http://www.robert-w-jones.com/, philosopher, axiology).

A strong argument against value monism is the fact that the monists can not agree on what is the single thing that is to be valued. Some suggestions have been:

pleasure

life/health/age

money

"utility"

"fitness"

goodness/godliness

time (but one could change this or one's age by relativistic means, and so....)

energy

etc....

References

Arrow, Kenneth. 1997. "Invaluable goods." Journal of Economic Literature, 35:2, 757-765.

Beardon, Alan, Juan Candeal, Gephard Herden, Esteban Indurain, and Ghanshyam Mehta. 2002. "The non-existence of a utility function and the structure of non-representable preference relations" Journal of Mathematical Economics, 37, 17-38.

Berlin, Isaiah. 1998. Concepts and Categories. Princeton, New Jersey: Princeton Univ. Press.

Dasgupta, Pallab, P.P. Chakraborti, and S.C. DeSarkar. 1999. Multiobjective Heuristic Search. Weisbaden, Germany: Vieweg.

Dubra, Juan, Fabio Maccheroni, and Efe Ok. 2001. "Expected utility theory without the completeness axiom" ICER Applied Mathematics Working Paper 2001/11.

Keeney, Ralph and Howard Raiffa. 1976. Decisions with Multiple Objectives. New York, New York: Wiley.

Liu, Boading. 2002. Theory and Practice of Uncertain Programming. New York, New York: Physica-Verlag.

Von Neumann, John and Oskar Morganstern. 1944. Theory of Games and Economic Behavior. Princton, New Jersey: Princeton Univ. Press.

Pepper, Stephen. 1958. The Sources of Value. Berkeley, California: Univ. of Calif. Press

Chaining Case-based Reasoners

R. Jones

Physical Science Department

Emporia State University

Emporia, KS 66801

Abstract

A fuzzy knowledge-based system is described wherein individual production rules are replaced by case-based reasoners. The knowledge engineering effort required to create such an expert system is considerably reduced as compared to more traditional methodologies.

Introduction

Knowledge-based systems have become an important class of computer software (Jackson, 1998). The most common sort of knowledge-based or "expert" systems are the "production" or "rule-based" systems consisting of a set of if-then condition-action rules like:

1. a man is an animal

and:

2. an animal is mortal

which, in combination with facts like:

3. Socrates is a man

are chained together (using conventional first order predicate logic) to produce conclusions, e.g.:

Socrates is mortal

Another sort of knowledge-based system is the "case-based reasoner" (CBR). CBRs contain a set of stored records or "cases" which are compared with new inputs (inquiries). If the inquiry "closely resembles" one of the stored records the output (prediction) associated (stored) with that case is assumed to apply to the new situation as well (Kolodner, 1993).

Chained Case-based Reasoners

The present system is a significant variation on traditional case-based reasoning. Conventional CBRs follow what might be described as a radical behaviorist-like model. That is, only inputs and outputs are employed, there are no intermediate ("internal" or "hidden") variables and no chaining to conclusions.

In our system, however, input values first generate intermediate results and these, in turn, are then combined in order to calculate further intermediate variables, and then, ultimately, output predictions. Any number of CBRs may be combined (chained) to produce a complete expert system (Jones, 2005) a kind of fuzzy semantic network.

A simple example

The system operation is probably best illustrated with the help of a simple ("toy") example which is intended to predict the likelihood that a given student will chose to attend some particular college. Three case bases are assumed to describe this decision process:

1. The student's grade average, parents' total annual income, and any special eligibility status will determine the amount of financial aid the student can expect to be offered.
2. The student's grade average, ACT test score, minority status, and number of extracurricular activities will determine whether or not the student will be accepted for admission to this particular school.
3. The school's annual tuition rate, financial aid amount offered, acceptance for admission, the size of the school, and its proximity to the student's home will determine how likely the student is to attend this particular school.

This example is intended for illustrative purposes only. It is not based on criteria employed at any real university.

Our software differs from the more conventional rule-based systems in that each "rule" is instead implemented as a complete case-based reasoner. Previous student applicant records are used to assemble the case bases of figures 1-3.

When a new student is under consideration this set of 3 CBRs can be used to determine the likelihood of his/her attendance. In the simplest possible implementation the input data are compared with the individual case bases of figures 1 and 2 to determine if the new student will gain admission and how much financial aid he/she might expect to receive. These intermediate results are then passed along to the CBR of figure 3 to determine the likelihood of the new student's attending this particular school.

For each case base some similarity measure is employed in order to locate the recorded case which most closely resembles the new input. The simplest method for doing this simply counts the number of attributes that the inquiry and the recorded case share in common. Vector dot products and other similarity measures are also available to define a match (Kolodner, 1993). The output recorded as part of that best matching case is then assumed to be applicable to the new student as well. Free CBR software is available for such nearest neighbor matching (Casual and Protos).

Alternatively, multiple cases can be used to interpolate to an output for the new inquiry. Outputs from n nearest matching cases can be retrieved and weighted by factors:

$$\frac{\text{sum-distance}}{\text{sum } (n-1)}$$

where "distance" is the distance between the input vector and that particular case and "sum" is the sum of the distances for all n records used in the interpolation (Stahl, 2001). Euclidean, Manhattan, or other distance measures can be employed in the calculation.

An Example Inquiry

Suppose a new student applies for college admission and has:

A grade point average of: 3.7

Her parents' income is: 45k

Her status is: 0

Her ACT score is: 28

Her number of extracurricular activities was: 3

Her minority status is: 0

The particular school's tuition is: 6k$

The particular school's size is: 4k

The distance that the applicant lives from this school is: 15 miles

The new student's descriptors do not exactly coincide with those of any previous applicants but figure 2 suggests that she most closely resembles students which were accepted, with the nearest match being with case 7, and figure 1 suggests she most closely resembles a student who received 5k$ per year in financial aid (case 7). When these results are then passed on to the case base of figure 3 we find that the new student most nearly resembles a student who did ultimately choose to attend the school. (i.e. case 7)

Conclusion

It has been possible to decompose a case-based reasoning system into a chain of smaller CBRs. This has the potential of speeding up the construction and use of large scale applications and making their detailed reasoning more transparent to the user. The knowledge engineering is considerably reduced in that the cases do not have to be decomposed into individual small scale rules but instead can be employed as recorded.

References

Jackson, Peter, 1998. Introduction to Expert Systems, Addison-Wesley, 542 p.

Jones, R., 2005. Chaining Case—Based Reasoners, Trans. Kansas Academy or Science, 108 (1/2), 74-75.

Kolodner, Janet, 1993. Case-Based Reasoning, Morgan Kaufmann, 321p.

Stahl, Gerry, 2001. Armchair mission to Mars: Using case based reasoning and fuzzy logic to simulate a time series model of astronaut crews, in Soft Computing in Case Based Reasoning, S. K. Pal, T. S. Dillon, and D. S. Young (Eds) Springer, 321-334.

Casual and Protos software for case-based reasoning
ftp://ftp.cs.utexas.edu/pub/porter
http://www.ai-cbr.org/tools.html

Figure 1

Inputs			Output	
grade point average (4.0 scale)	parents' income (k$)	special status (0 or 1)	financial aid offer (k$/yr)	Case #
3.8	20	0	5	1
3.75	22	1	5	2
3.1	22	0	3	3
3.0	20	1	5	4
2.5	24	0	0	5
2.6	25	1	0	6
3.75	50	0	5	7
3.85	51	1	5	8
3.0	55	0	0	9
3.2	70	1	5	10
2.5	60	0	0	11
2.4	80	1	0	12
3.8	100	0	5	13
3.85	150	1	5	14
3.0	110	0	2	15
2.6	100	0	0	16

Figure 2

Inputs				Output	
grade average (4.0 scale)	ACT score (#)	extra curricular activities (#)	minority status (0 or 1)	acceptance (0 or 1)	Case #
3.8	30	2	0	1	1
3.75	31	3	1	1	2
3.1	30	4	0	1	3
3.0	29	3	1	1	4
2.5	29	3	0	1	5
2.6	28	2	1	1	6
3.75	27	3	0	1	7
3.85	28	2	1	1	8
3.0	26	0	0	0	9
3.2	27	2	1	1	10
2.5	26	1	0	0	11
2.4	25	1	1	0	12
3.8	22	1	0	0	13
3.85	23	2	1	1	14
3.0	22	1	0	0	15
2.6	23	1	0	0	16

Figure 3

Inputs					Output	
tuition (k$/yr)	financial aid offer (k$/yr)	admission acceptance (0 or 1)	distance from home (miles)	size of school (# students)	likelihood of attendance (0-1)	Case #
3	2	1	15	4	1	1
3	3	1	10	4	1	2
3	0	0	40	4	0	3
3	2	1	100	10	0	4
3	3	1	25	10	1	5
6	2	1	100	4	0	6
6	5	1	20	4	1	7
6	5	1	200	4	0	8
6	0	0	25	10	0	9
6	2	1	20	10	1	10
6	3	1	20	10	1	11
15	5	1	40	10	0	12
15	0	1	100	10	0	13
15	3	1	50	10	0	14
15	5	1	40	15	1	15
15	0	0	30	15	0	16
15	3	1	50	15	0	17

Objective Analysis of Student Data

R. Jones
Physical Sciences Department
Emporia State University
Emporia, Kansas

Introduction

"How good is my data?" How much data do I need to take?" These are common student questions and are often difficult to answer. In the present paper we outline methods for objectively deciding such matters.

Linear Relationships

In our typical first year physics course and advanced laboratory course many of the relationships we have students investigate are linear:

$F = ma$	plot	$F \ vs \ a$
$F = kx$	plot	$F \ vs \ x$
$PV = NkT$	plot	$P \ vs \ \frac{1}{V}$
$Q = mL$	plot	$Q \ vs \ m$
$f = \mu N$	plot	$f \ vs \ N$
$V = IR$	plot	$V \ vs \ I$
$R = \frac{\rho L}{A}$	plot	$R \ vs \ L$
$I^2 R\Delta t = mc\Delta T$	plot	$\Delta T \ vs \ \Delta t$
$n\lambda = \frac{dy}{L}$	plot	$y \ vs \ n$
$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$	plot	$\frac{1}{d_o} \ vs \ \frac{1}{d_i}$
$B = \frac{\mu_o N}{a} I\left(\frac{4}{5}\right)^{\frac{3}{2}}$	plot	$B \ vs \ I$
$qV = \frac{hc}{\lambda}$	plot	$V \ vs \ \frac{1}{\lambda}$

In other cases where the equations are nonlinear we recast the relationship in linear form by suitable mathematical transformations:

$y = \frac{1}{2}gt^2$	plot	$y \ vs \ t^2$
$F = mrw^2$	plot	$F \ vs \ w^2$
$T = 4\frac{L^2 f^2 \mu}{n^2}$	plot	$T \ vs \ \frac{1}{n^2}$
$\tau = 2\pi\sqrt{\frac{m}{k}}$	plot	$\tau \ vs \ \sqrt{m}$
$\tau^2 = \frac{4\pi^2}{g}L$	plot	$\tau^2 \ vs \ L$
$n_1 \sin\theta_1 = n_2 \sin\theta_2$	plot	$\sin\theta_1 \ vs \ \sin\theta_2$
$\ln\frac{N}{N_o} = \frac{-x}{\lambda}$	plot	$\ln N \ vs \ x$

The Pearson Correlation Coefficient

Having cast everything in simple linear form we can investigate the quality of our data using the common Pearson "R" correlation coefficient:[1]

$$R = \frac{\frac{\sum_{i=1}^{N} X_i Y_i}{N} - \left(\frac{\sum_{i=1}^{N} X_i}{N}\right)\left(\frac{\sum_{i=1}^{N} Y_i}{N}\right)}{\sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \left(\frac{\sum_{i=1}^{N} X_i}{N}\right)^2}\sqrt{\frac{\sum_{i=1}^{N} Y_i^2}{N} - \left(\frac{\sum_{i=1}^{N} Y_i}{N}\right)^2}}$$

where X is the independent variable, Y is the dependent variable, and N X_i, Y_i data points have been measured.

Many computer plotting software utilities provide the Pearson R value (or, alternatively, R^2) along with a graph X_i *vs* Y_i. MS Excel is one common example which we use. Pocket calculators (like the TI-83) can also compute R as one of their statistical analysis functions.

Guilford[2] says that $R \geq .9$ indicates "very high correlation," a "very dependable relationship" and we usually adopt $R \geq .9$ as our definition of "good data." Work in the social sciences might require a relaxation of this standard.

If one wishes to minimize the amount of math involved, it is possible to supply the students with a graph like figure 1. Data points have been chosen to give a correlation coefficient of $R = .9$ for this graph. Students can be instructed to hold their data plots alongside this model. Qualitatively, if their data looks as good or better than figure 1 then $R \geq .9$ and they have "good data."

How much data should I take?

It is possible to use a similar method to decide when enough data has been taken. Suppose students have collected the following data:

X	Y
2	3.1
2.1	5
3	5.1
3.1	7

A calculation of the correlation coefficient gives $R = .79$. If we are striving for $R \geq .9$, we are led to take more data, perhaps arriving at:

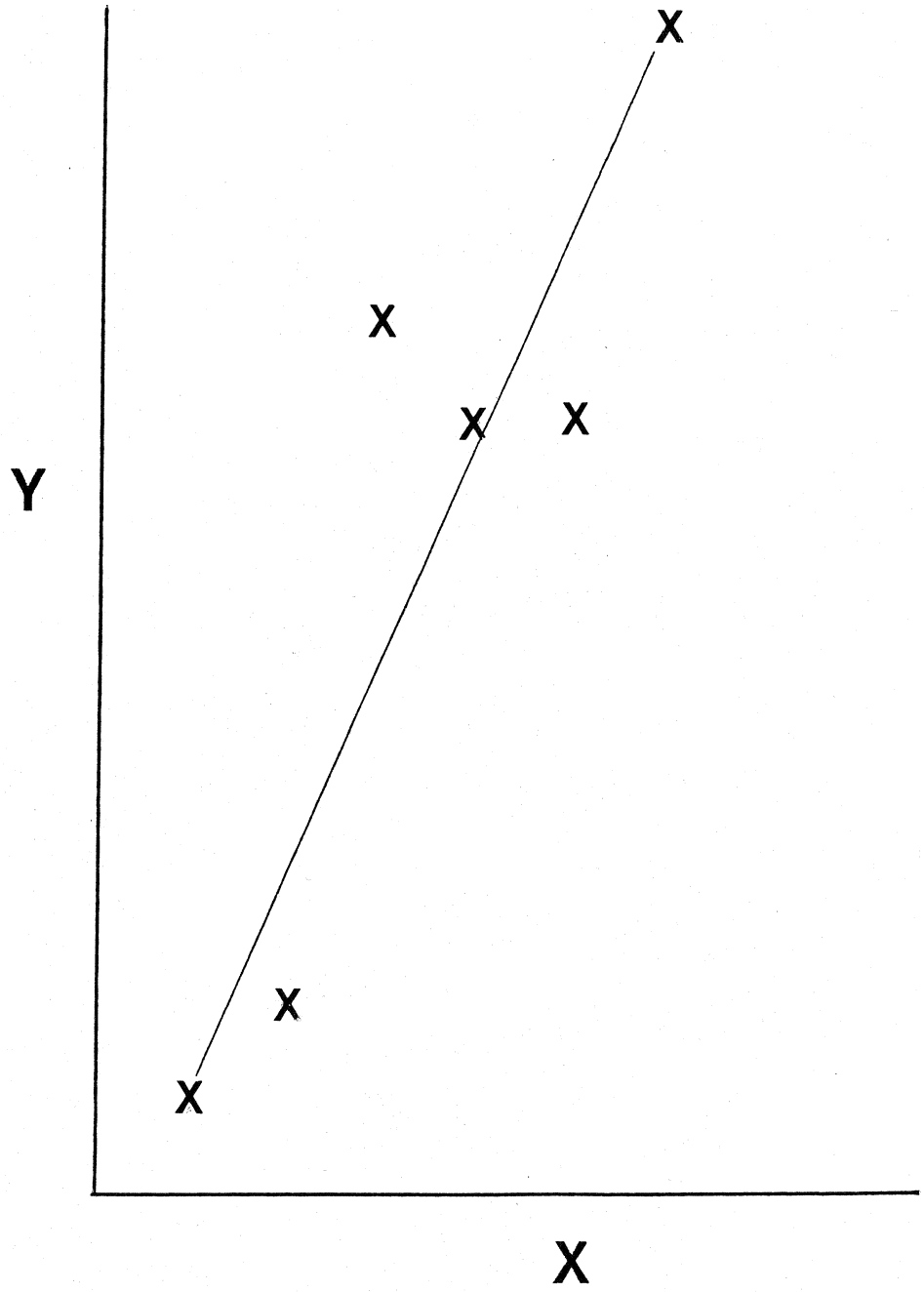
X	Y
2	3.1
2.1	5
3	5.1
3.1	7
1	1.5
1.1	2.5

The Pearson R value is now $R = .92$ and "enough" data has now been taken. It is worth pointing out to students that you only know how much data to take after you have begun to take measurements, not before. Data taking and analysis should be interspersed. One should not try to separate data taking from data analysis. Of course in real life one might also have to stop when you run out of time or out of resources.

1. R.C. Sprinthall (1994) "Basic Statistical Analysis" New York: Allyn and Bacon, pg 211.

2. J.P. Guilford (1956) "Fundamental Statistics in Psychology and Education" New York: McGraw-Hill, pg. 145.

Fig 1



Experiments in Machine Creativity

R. Jones
Physics
Emporia State Univ.
Emporia, KS 66801

Abstract

A variety of computer experiments have been conducted in order to explore several forms of potential machine creativity.

Introduction

Automated creativity is not new. Stephen Thaler believes that he has developed a "creativity machine" which can invent things and discover ideas (Thaler, 1996). Quillian's semantic memory program was shown to be capable of discovering "an idea that was implied by, but by no means expressed in, the data that were input." (Quillian, 1968 and Cohen and Feigenbaum, 1980) Rumelhart et al have shown that semantic networks can be learned by artificial neural networks (ANNs) and that by exploring unusual input attributes (descriptors) the ANN can itself invent new categories from the original input categories (Rumelhart, et al, 1986).

Autoassociative Neural Network Experiments

In the present work we train an autoassociative neural network on known categories (known plasma confinement schemes). Once trained the ANN can recall each of the input categories or generate interesting new confinement ideas.

The autoassociative neural network (ANN) consists of N input units, one for each attribute (parameter) relevant to each input category (vector) I_k. These are fully connected to M hidden units. Normally M<N in order to force the hidden layer's internal representation to compress the input data. (A significant role of theory is data compression.) We can add hidden units if the network fails to train properly. The hidden units are, in turn, fully connected to N output units. In an auto associator the outputs are duplicate copies of the inputs I_k. All variables in I_k are normalized so as to fall within the unit interval.

For each input vector I_k the hidden unit j receives an input:

$$h_j=\sum_{k=1}^{N} w_{jk}I_k+w_o \qquad (1)$$

and produces an output:

$$V_j = \frac{1}{1+\exp(-h_j)} \tag{2}$$

(We don't know that this function is optimum but we use it because it is mathematically convenient). W_o is an adjustable bias.

Output unit k then receives:

$$h_k = \sum_{j=1}^{M} w_{kj} V_j + w_o \tag{3}$$

and computes the output vector:

$$I_k^1 = \frac{1}{1+\exp(-h_k)} \tag{4}$$

(It is possible for us to introduce more than 1 hidden layer if we so desire. Three hidden layers are adequate to represent any continuous function. Five hidden layers are adequate to represent any discontinuous function.) The initial values of the weights W_{jk} and W_{kj} are random numbers in the range between 0 and 1. The I_k^1 values can be thresholded if binary output is desired.

The computed theoretical output vector I_k^1 is compared with the actual vector I_k and the error $I_k - I_k^1$ is "back propagated" in order to correct the weights W_{kj} and W_{jk}(Bishop, 1994).

The weights W_{kj} are corrected by amounts:

$$\Delta w_{kj} = C(I_k - I_k^1) \left(\frac{1}{1+\exp(-h_k)}\right) \left(1 - \frac{1}{1+\exp(-h_k)}\right) V_j \tag{5}$$

and the weights W_{jk} are corrected by amounts:

$$\Delta W_{jk} = \Sigma C(I_k - I_k^1) \left(\frac{1}{1+\exp(-h_k)}\right) \left(1 - \frac{1}{1+\exp(-h_k)}\right) W_{kj} \left(\frac{1}{1+\exp(-h_j)}\right) \left(1 - \frac{1}{1+\exp(-h_j)}\right) I_k \tag{6}$$

where c, the "learning rate", is a constant, typically 0.1

The weights are adjusted (corrected) at each calculation step by:

$$\text{new } W_{kj} = \text{old } W_{kj} + \Delta W_{kj}$$

$$\text{new } W_{jk} = \text{old } W_{jk} + \Delta W_{jk} \qquad (7)$$

and the process is repeated, using the same single I_k until $I_k - I_k^{\,l}$ is smaller than some desired error. This whole process is then repeated with each additional available training vector I_k, each vector representing a different input category. Adjustable noise can be added to help escape from local minima.

Training on subsequent I_k data sets reduces the network's ability to reproduce previous data input. For this reason, repeated training "passes" are made through the complete training set until $I_k - I_k^{\,l}$ is acceptably small for all data sets employed.

ANN Creativity

We will first illustrate how ANN operates with a "toy" example. In this "simplest possible example" ANN is told only that I. a "stellarator" has good plasma confinement, no open field lines, no plasma rotation, and a closed magnetic geometry. ii. A "tandem mirror" has good confinement, a potential barrier Δv_p, no plasma rotation, and no closed geometry (i.e. "open" geometry) iii. A "centrifugal trap" has poor plasma confinement, no closed magnetic field, (open magnetic geometry), and plasma rotation. It knows nothing more. The input data are represented in vector form, table 1.

An ANN trained on the 3 vectors I_k of table1 functions as a content addressable memory. If we input only the knowledge that the system of interest has a closed magnetic field, i.e. inputting the vector:

$$I_k = 0 \quad 0 \quad 0 \quad 0 \quad 1$$

we obtain as output:

$$I_k = 0 \quad 0 \quad 0 \quad 1 \quad 1$$

i.e. the ANN thinks we are describing a stellarator, a system characterized by a closed magnetic field and good confinement.

On the other hand if we ask ANN for a system with good plasma confinement i.e. if we input:

I_k= 0 0 0 1 0

we may (and did) get as output:

I_k= 1 0 0 1 1

i.e. ANN suggests that $_{\text{both}\,\Delta V_p}$ and closed magnetic fields are associated with good confinement and should be employed simultaneously. This "new idea" is, in fact, "magneto electrostatic confinement", a new approach to controlled fusion (Jones, 1992).

Relatively little knowledge is represented in this toy example. ANN is told, for instance, that good confinement can be related to either electrostatic potential barriers or toroidal magnetic fields. Only a few other facts are provided in this austere database.

Science is both theory and experiment driven. ANN produces new associations based on its representation of the input data. If experiments do not support this extrapolation then the observed results (associations) can be input as additional (new) training data and the process iterated. If, for instance, nonmonotonic V_p and closed B led to poor confinement we would input the new training vector:

I_k= 1 0 0 0 1

in addition to the vectors of table 1.

Our toy problem was given very little information but it was exactly the information needed with which to synthesize the new idea of magnetoelectrostatic confinement. In any realistic ANN database much larger amounts of data will be input. Much of it will be a mere distraction leading, at best, to other new ideas, at worst to nothing useful. ANN can synthesize new ideas but there is no guarantee that they are correct, only experiment can determine that.

The toy example does little more than merely add two ideas together. A more realistic example is given in table 2. In this database there are 8 input vectors, each is a different sort of common plasma source. 8 different attributes characterize these devices: V, voltage, I, discharge current, B, magnetic confining field, HC, hollow cathode , N, relative plasma density (medium=.5, high=1, low=.1), V_{wall}, wall bias voltage, C, number of cathodes (none=0 single=1, many=4), M, magnetic multidipoles.

An ANN trained on the data vectors of table 2 was stimulated by the input:

I_k= 0 0 0 0 0 -1 0 0

and gave the output:

$$I_k^{\,l} = \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad -1 \quad 0 \quad 1$$

suggesting that a negative wall bias be employed with multidipoles. This is the "Emporia source" geometry explored by Jones (Jones, 1987).

In the present examples the user has identified the attributes, and the semantic categories. The value given to each attribute of a category simply represents its frequency of occurance (association with) that category. 1=always present, 0=never present, .5=present, ~ ½ the time, etc. Alternatively, the user can supply attributes only and an ART (adaptive resonance theory) network can discover semantic categories directly from data/examples.

In general ANN is not simply adding two semantic networks and regurgitating all of their attributes. An attribute present in one semantic net may be deleted from the "composite" output. This was seen in the example on page 34 of Rumelhart (Rumelhart, 1986).

The examples cited so far were given very limited input information and therefore their outputs are similarly limited. Table 3 represents the largest example we could display on a single sheet of paper but it is still rather small compared to some databases we have employed. Each input vector represents a description of an approach to controlled fusion. An ANN trained on these semantic descriptions is being searched in an effort to discover new confinement schemes. Finer grained (more detailed) attributes are also useful.

ART Network Experiments

It is common for plasma confinement devices to exhibit a variety of operational regimes. Bumpy tori exhibited 3 primary confinement modes, termed C, T, and M. (Colchin, 1983). In addition to these the T mode could be further divided into submodes having either positive or negative ambipolar potential.

Tokamak devices, on the other hand, are attributed with L, H, Z, P (Kaye, 1985) and VH modes (Turnbull, 1995) among others. Distinctly different modes are usually explained by distinctly different physics and a systematic search for high performance typically translates into an exploration of the detailed device mode structure. It is important, then, to be able to identify the various operating regimes accessible with a given experimental device.

Our recent experiments in stellarator (Jones, 1995) and magnetoelectrostatic confinement devices (Jones, 1985) display evidence of complex mode structure. In fact, the preliminary

application of neural network analysis to a stellarator database resulted in the discovery of a previously unsuspected mode of operation in which plasma confinement improved as the magnetic confining field was reduced (Jones, 1995).

Artificial neural network programs are capable of modeling complex input-output relationships learned on the basis of empirical examples alone. In the experiments which we wish to model it is possible to vary the magnetic field used to confine the plasma, B, the neutral gas fill density (Hydrogen), N_n, and the RF power input from antennae used to sustain the discharge, P. We are able to characterize the behavior of the experiment with a set of (average) "output" (performance) parameters: the plasma (electron) temperature, T_e, the plasma density, N_e, and the confinement time, t. Each of these 6 parameters is averaged over an experimental run and then normalized so that each quantity falls within the unit interval. The concatenation of the 6 scaled quantities then constitutes a vector, V^k, which describes a given experimental run. The job of the neural network is to group all of these 6 dimensional vectors into categories ("modes"). This has been accomplished with an ART2 network (Carpenter and Grossberg, 1987 and Freeman, 1994).

The ART Network Algorithm

Each category is defined by a prototype (exemplar) P_i, (Initially the P_i are either defined by the user or randomly selected from the set V^k.) As each run vector V^k is processed by the network we use the maximum of the dot product

$$P_i \cdot V^k$$

to find the prototype that best categorizes that particular V^k.

We use

$$P_i \cdot V^k \geq \propto \sum_j V_j^k \qquad (8)$$

to test whether P_i is sufficiently similar to V^k. Here $\propto$ is a positive number:

$$\propto \leq \frac{1}{\sqrt{p}} \qquad (9)$$

with p the dimension of the vectors V^k.

If V^k is sufficiently similar to the existing P_i then we test to see if:

$$P_i \cdot V^k \geq p \qquad (10)$$

a vigilance parameter, $o \leq p \leq 1$.

If acceptable V^k belongs to that category P_i and we modify P_i to more closely resemble

V^k.

$$P_i = \frac{(1-\beta)\ P_i + \beta V^k}{\| (1-\beta) P_i + \beta V^k \|} \qquad (11)$$

where β is a small positive number.

If V^k is not sufficiently similar to P_i then we define a new category with V^k as its prototype P_i. This procedure is itterated for all V^k.

ART Creativity

If the network has previously learned a particular input vector it will quickly identify (catagorize) that vector when it is presented with it again. If the network does not immediately recognize an input vector it will search through its stored patterns (category examples) looking for a match. If no match is found, then the network will learn (store) the new pattern, defining a new category. If, on the other hand, the input matches a stored (known) pattern to within a specified amount (determined by the vigilance parameter) then the existing category will be adjusted toward the input vector pattern. An ART network trained on raw experimental data from the EBT device rediscovered the known C, M, and T modes in this fashion.

Another way to use the algorithm is to input known mode exemplars (initial P_is) and adjust the vigilance to see what other modes the network might suggest. Table 4 consists of the mode prototypes P_1, P_2, and P_3 input by the user to describe stellarator experiments. Table 5 shows the mode structure discovered by the network when presented with the existing experimental stellarator database (Jones, 1995). P_1, P_2, and P_3 have been changed slightly by the network and two new modes P_4 and P_5 have been postulated. Experiments are now underway to see if P_4 and P_5 do, indeed, indicate physics which differs from that of the P_1, P_2, and P_3 regimes.

Future Work: Analogy Machines

A form of analogical reasoning can be performed on feature vectors A, B, C, and D. A is to B as C is to D. The vector difference B-A gives the magnitude and direction of the change starting from category A needed to get to category B. Adding B-A vectorially to category C then produces the desired analog vector D; i.e. D=C+B-A.

As a toy example in an attribute space having 3 dimensions (feline, young, canine) we can perform the analog reasoning cat is to kitten as dog is to "?". In this space "cat" is the vector A=(1, 0, 0), "kitten" a vector B=(1, 1, 0), and "dog" is C=(0, 0, 1). The difference vector B-A is then (1, 1, 0)-(1, 0, 0)=(0, 1, 0); B differs from A in being young. B-A is then added to C; (0, 0, 1)+(0, 1, 0)= (0, 1, 1), i.e. D is a "young, canine", that is a puppy.

In a vector space of categories C+B-A can be compared with all of the known categories using the vector dot product operation. The existing category that is found to be nearest to C+B-A is the solution D. If C+B-A is not close enough to any EXISTING category then a new D can be defined; i.e. automated category discovery. The ART neural network algorithm can decide if the two are "close enough".

Conclusions

A number of computer experiments have been conducted which give evidence that several forms of machine creativity are practical.

Literature Cited:

Bishop, C.M. 1994. Neural networks and their applications. Rev. Sci. Instr. 65(6): 1803-1832.

Carpenter, G.A. and S. Grossberg. 1987. ART2: self-organization of stable category recognition codes for analog input patterns. Applied Optics 26: 4919-4930

Cohen, P.R. and E.A. Feigenbaum. 1980. The Handbook of Artificial Intelligence, Vol. III, William Kaufmann, Inc. pp.36-39.

Colchin, R.J. 1983. Plasma properties in the elmo bumpy torus. Plasma Physics 25: 597-615.

Freeman, J.A. 1994. Simulating Neural Networks with Mathematica. Addison-Wesley pg 243.

Jones, R. 1985. Magnetoelectrostatic confinement in RSX-1. lett. Nuovo Cimento 42: 357-362.

Jones, R. 1987. A new low pressure gas discharge device, Trans. Kansas Acad. Sci. 90: 120-126.

Jones, R. 1992. Plasma confinement with electric and magnetic fields, Trans. Kansas Acad. Sci. 95: 122-127.

Jones, R. 1995. Plasma parameter optimization using a neural network. Trans. Kansas Acad. Sci. 98: 144-148.

Kaye, S.M. 1985. A review of energy confinement and local transport scaling results in neural heated tokamaks. Phys. Fluids 28: 2327-2343.

Quillian, M. R. 1968. Semantic Memory in M. Minsky, editor, Semantic Information Processing, MIT Press, pp216-270.

Rumelhart, D. E., P. Smalensky, J. L. McClelland, and G. E. Hinton. 1986. In Parallel Distributed Processing, MIT Press, pp 7-57.

Thaler, S. 1996. Neural nets that create and discover, PC AI magazine, pp. 16-21.

Turnbull, A.D. 1995. High beta and enhanced confinement in a second stable core VH mode advanced tokamak. Phys. Rev. Letters 74: 718-721.

table 1

	ΔV_p	open field lines	rotating plasma	good confinement	closed geometry
Tandem Mirror	1	1	0	1	0
Stellarator	0	0	0	1	1
Centrifugal Trap	0	1	1	0	0

table 2

	V	I	B	HC	n	V_{wall}	C	M
Glow Discharge	.5	.1	.05	0	.1	0	1	0
Arc	.1	.5	.05	0	.5	0	1	0
Hollow Cathode Discharge	.1	.5	.2	1	.5	0	1	0
Electrostatic	.5	.1	0	0	.1	-1	1	0
Tufted Anode Discharge	.5	.2	0	0	.3	1	1	0
Penning Discharge	.5	.2	1	0	.2	0	4	0
Multidipole Device	.5	.1	.1	0	.2	0	1	1
RF Discharge	0	0	.1	0	.2	0	0	.05

table 3

	B_z	Torus	I_z	Pulsed	B_v	n	rotation	electrodes	internal rings	ΔV_p	mirrors
Tokamak	1	1	1	1	1	1	0	0	0	0	0
ICF	0	0	0	1	0	4	0	0	0	0	0
Stellarator	1	1	0	0	.1	1	0	0	0	0	0
RFP	.2	1	1	0	.05	1	0	0	0	0	0
Int. Ring	.3	1	0	0	0	1	0	0	1	0	0
Surmac	0	1	0	0	0	1	0	0	4	0	0
Mirror	1	0	0	0	0	1	0	0	0	0	1
Centrifugal	1	0	0	0	0	1	1	1	0	0	1
TM	1	0	0	0	0	1	0	0	0	1	1
Multi Mirror	1	0	0	0	0	1	0	0	0	0	4
EBT	1	1	0	0	0	1	0	0	0	0	4
pinch	.2	0	4	1	0	2	0	0	0	0	0
X trap	0	1	1.5	1	0	1.5	0	0	1	0	0
Lavrentyev	0	.05	0	0	1	1	0	1	1	0	1
Wall Conf.	1	1	0	1	0	3	0	0	0	0	0
E. Static	0	0	0	0	0	2	0	1	0	1	0

Table 4

Parameter	Mode		
	P_1	P_2	P_3
B (Gauss)	700	700	1700
N_n (mm)	5×10^{-5}	10^{-5}	5×10^{-5}
P (Watts)	50	50	200
T (eV)	14	50	25
Ne (cm^{-3})	10^{11}	4×10^{10}	4×10^{11}
t (ms)	0.4	0.3	0.6

Table 5

Parameter	Mode				
	P_1	P_2	P_3	P_4	P_5
B (Gauss)	700	700	1800	750	1800
N_n (mm)	4×10^{-5}	10^{-5}	4×10^{-5}	10^{-5}	10^{-5}
P (Watts)	50	50	200	250	500
T (eV)	13	53	24	50	60
Ne (cm^{-3})	1.4×10^{11}	3.9×10^{10}	4.1×10^{11}	3×10^{11}	5×10^{11}
t (ms)	0.4	0.3	0.6	0.4	0.4

Experiments With A Hierarchical Ensemble Classifier

R. Jones
Physical Sciences Department
Emporia State University

We describe the design and use of a hierarchical ensemble classifier.

Each of the component classifiers that make up the layered hierarchy is trained on a data subset which is small compared to the complete training database.

Introduction

Ensemble classifiers have become one of the most active areas in machine learning research (Dietterich, 1997 and Hansen and Salamon, 1990). Each of the component classifiers that make up the ensemble can be trained on a subset of the available data simplifying and speeding up the learning process. A further speed up is possible if components can be implemented on multiple processors run in parallel.

A Hierarchical Ensemble Classifier

The present work employs a hierarchical arrangement of universal approximators (figure 11) and bears some relationship to the ensemble classifier described by Chou (Chou, 1999). The training database is first divided into a number of smaller datasets by a standard clustering algorithm, figure 1. This is best illustrated using the toy database example of figure 2, which was clustered by the k-means algorithm (excluding Q) using commercial Unistat 4.5 software.

In this example In1 and In2 are inputs and Out1 and Out2 are responses which yield reward Q. Three clusters were formed centered at (-.5, +.25, 0, .75), (1, -1, 0, 0), and (-.1429, +.4286, 0, -.8571). The data subset that constitutes the first cluster (figure 3) is used to train ANN1 (artificial neural network 1) using commercial Brainmaker 3.5 software. The data subset forming the second cluster (figure 4) is used to train ANN2 and the data subset from the third cluster (figure 5) is used to train ANN3. Network training takes In1, Out1, In2, and Out2 as ANN inputs and Q as the sole ANN output; i.e., each of ANN1, ANN2, and ANN3 is a network having 4 inputs and 1 output.

Once ANN1, ANN2, and ANN3 have been trained the full data set (figure 2) is passed through all three ANNs producing the nonlinearly transformed data set of figure 6. This representation is then clustered by the k-means algorithm forming 2 clusters centered at (1.0756, 1.5794, 1.4755) and (0.2217, 1.2378, 0.3784). The data subset from the first of these clusters (figure 7) is then used to train ANN4 (figure 8) and the remaining data from figure 6 is used to train ANN5 (figure 8).

With ANN1, ANN2, ANN3, ANN4, and ANN5 all trained the full data set (figure 2) is passed through the first 2 layers of the hierarchy to produce the doubly transformed (and compressed) representation of figure 9. A final single classifier, ANN6, is then trained on this compact 2 component representation, figure 10.

Classifier Use

With all 3 layers of the hierarchy trained the performance system is completed (figure 11), and can be deployed. At time step 1 In1 is input. Out1 can then be adjusted until a good Q value is obtained (subject, possibly, to performance time limitations). At time step 2 In2 is added (to In1 and the frozen Out1) and Out2 is then adjusted so as to optimize Q. The network of figure 11 successfully reproduces the complete database of figure 2.

The toy example of figure 2 was chosen for illustrative purposes only. In more realistic examples many more inputs, outputs, and time intervals have been employed. In large scale applications it is also possible to employ more than the 3 hierarchical layers illustrated in figure 11.

Literature Cited

T. G. Dietterich, Machine-Learning Research: Four Current Directions, AI Magazine 18(4) 97-136, 1997

L. Hansen and P. Salamon, Neural Network Ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence 12(10) 993-1001, 1990

Yu-Yu Chou, PhD Thesis, Hierarchical Multiple Classifier Learning System, Univ. of Washington, Electrical Engineering, 1999 pp 115

Brainmaker software version 3.5, California Scientific Software, Nevada City, CA, 1993

Unistat software version 4.5, Unistat Limited, Unistat House, London, UK, 1997

Figure Captions

fig 1: learning phase 1, for layer 1

fig 2: complete original dataset

fig 3: dataset for training ANN1

fig 4: dataset for training ANN2

fig 5: dataset for training ANN3

fig 6: nonlinearly transformed and compressed data set

fig 7: dataset for training ANN4

fig 8: learning phase 2, for layer 2

fig 9: nonlinearly transformed and compressed data set

fig 10: learning phase 3, for layer 3

fig 11: classification/performance phase

Fig 1

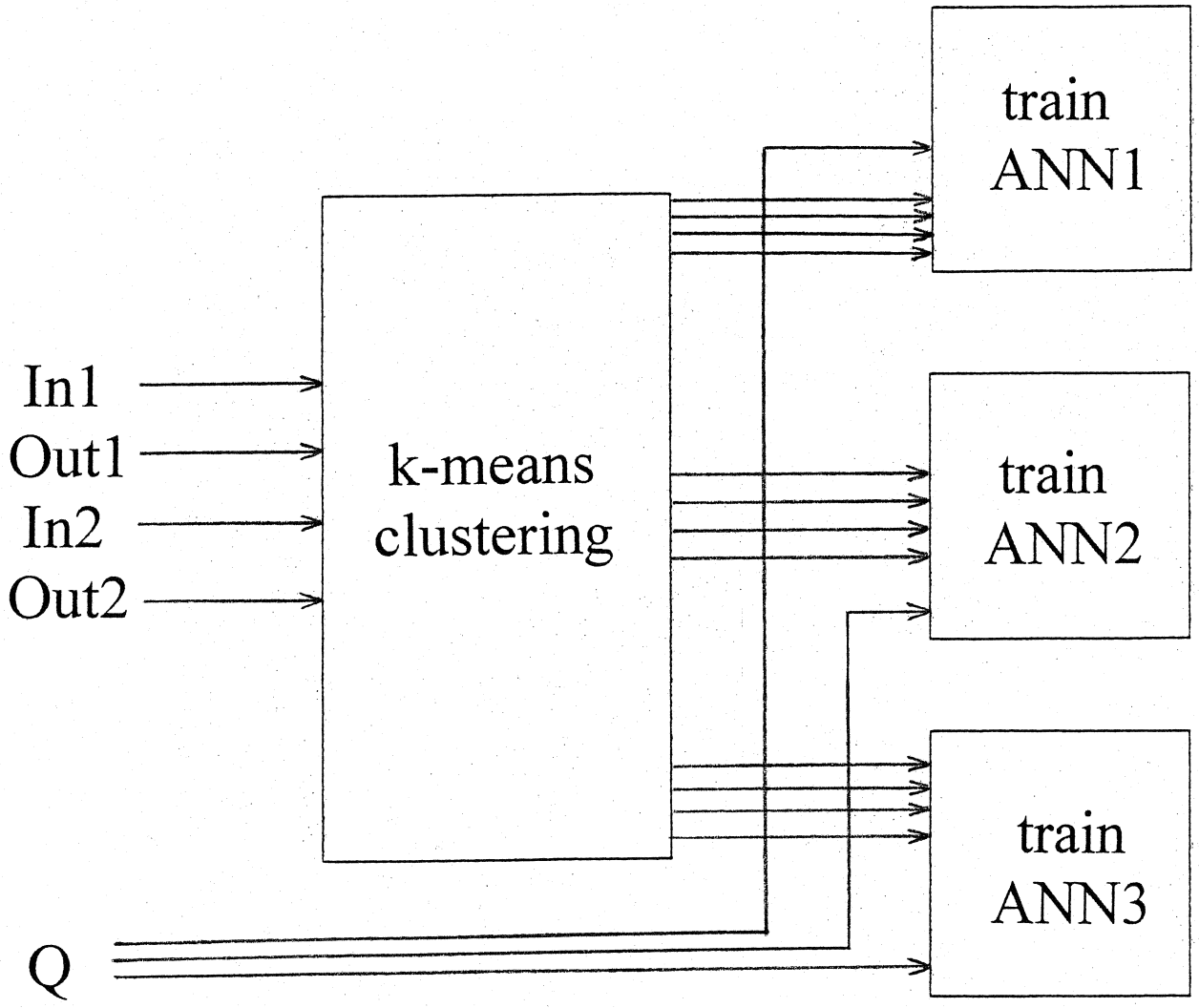

train
ANN1
In1
Out1
In2
Out2
k-means
clustering
train
ANN2
train
ANN3
Q

fig 2

Original Dataset

IN1	OUT1	IN2	OUT2	Q
-1	1	-1	1	2
1	-1	1	-1	2
-1	1	1	-1	2
1	-1	-1	1	2
-1	1	-1	-1	1
1	-1	-1	-1	1
-1	1	1	1	1
1	-1	1	1	1
-1	-1	-1	1	1
-1	-1	1	-1	1
1	1	-1	1	1
1	1	1	-1	1
1	1	-1	-1	0
-1	-1	1	1	0
1	1	1	1	0
-1	-1	-1	-1	0
-1	1	0	0	1
1	-1	0	0	1
1	1	0	0	0
-1	-1	0	0	0

Fig. 3

Data Set for Training ANN1

IN1	OUT1	IN2	OUT2	Q
-1	1	-1	1	2
-1	1	1	1	1
-1	-1	-1	1	1
1	1	-1	1	1
1	1	1	1	0
-1	-1	1	1	0
-1	1	0	0	1
-1	-1	0	0	0

Fig 4

Data Set for Training ANN2

IN1	OUT1	IN2	OUT2	Q
1	-1	1	-1	2
1	-1	-1	1	2
1	-1	-1	-1	1
1	-1	1	1	1
1	-1	0	0	1

fig 5

Data Set for Training ANN3

IN1	OUT1	IN2	OUT2	Q
-1	1	1	-1	2
-1	1	-1	-1	1
-1	-1	1	-1	1
1	1	1	-1	1
1	1	-1	-1	0
-1	-1	-1	-1	0
1	1	0	0	0

Fig 6

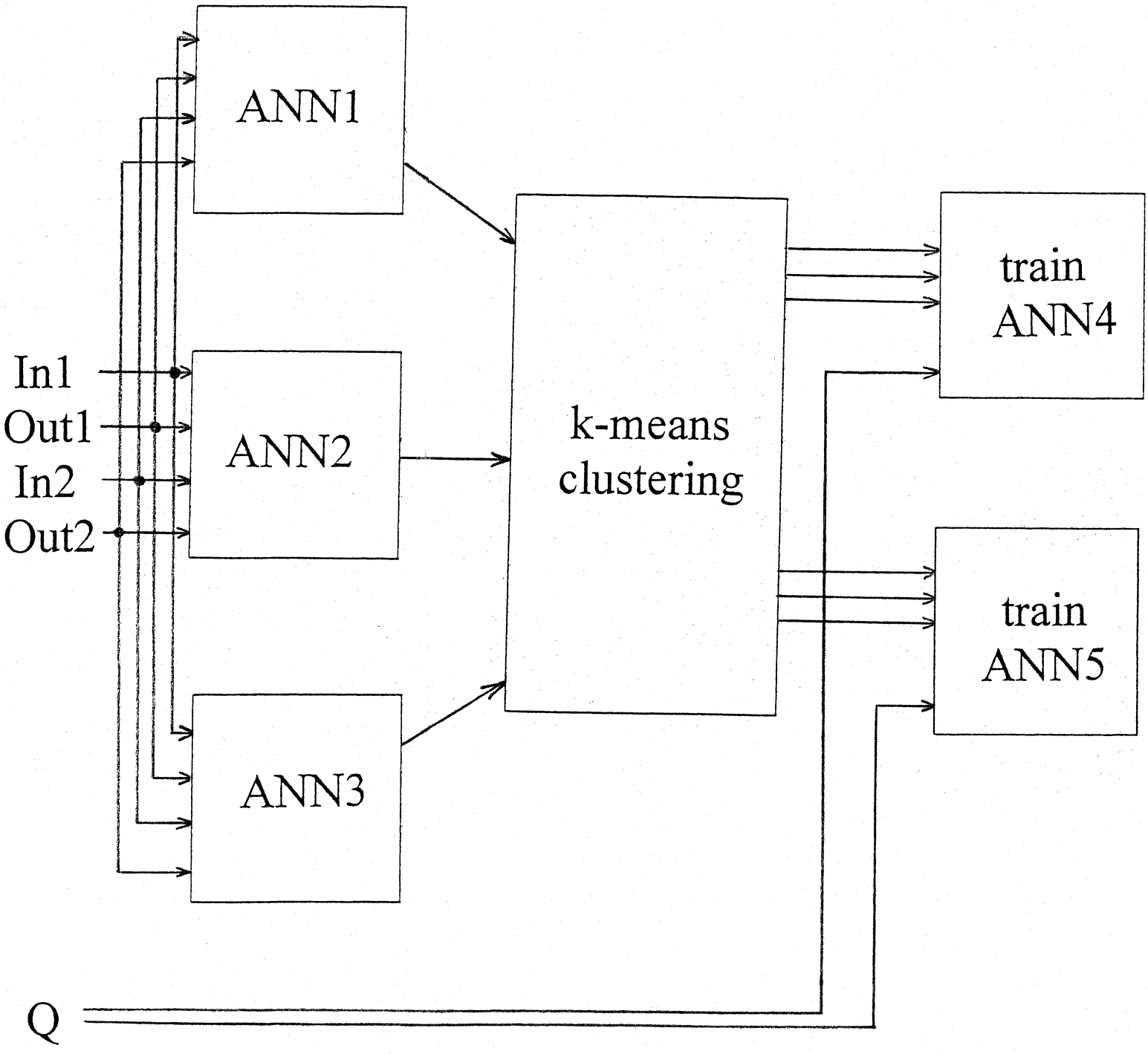
Transformed Data Set

ANN1 Output	ANN2 Output	ANN3 Output	Q
1.8009	1.8892	1.0485	2
.0758	1.9004	.8678	2
.3912	1.0977	1.8029	2
.0807	1.9002	.0568	2
.9269	1.5042	1.1876	1
.1593	1.056	.2262	1
1.193	1.7417	1.773	1
.0324	1.0279	.2243	1
.9127	1.2112	.1139	1
.0363	1.1722	1.1642	1
.9035	1.2388	.1383	1
.0944	1.0127	.8087	1
.2609	1.0101	.1276	0
.1583	1.6731	.5929	0
.1027	1.0452	.2345	0
.0675	1.0855	.1764	0
1.0661	1.6643	1.6056	1
.0822	1.1001	.1759	1
.1779	1.0171	.1998	0
.1803	1.117	.568	0

Fig 7

ANN1 Output	ANN2 Output	ANN3 Output	Q
1.8009	1.8892	1.0485	2
0.3912	1.0977	1.8029	2
0.9269	1.5042	1.1876	1
1.193	1.7417	1.733	1
1.0661	1.6643	1.6056	1

fig 8


ANN1
ANN2
ANN3
In1
Out1
In2
Out2
k-means
clustering
train
ANN4
train
ANN5
Q

Retransformed Data Set

ANN4 Output	ANN5 Output	Q
1.9002	1.9996	2
1.377	1.8151	2
1.9009	1.7755	2
1.9146	1.8043	2
1.1001	1.9996	1
1.6246	0.6403	1
1.0586	1.9996	1
1.5975	0.7619	1
1.9036	0.861	1
1.3675	1.1569	1
1.9036	1.1627	1
1.3514	0.817	1
1.7298	0.2018	0
1.6185	0.1969	0
1.6009	0.587	0
1.6936	0.5187	0
1.063	1.9996	1
1.7051	0.5822	1
1.637	0.2736	0
1.2793	0.1876	0

Fig 10

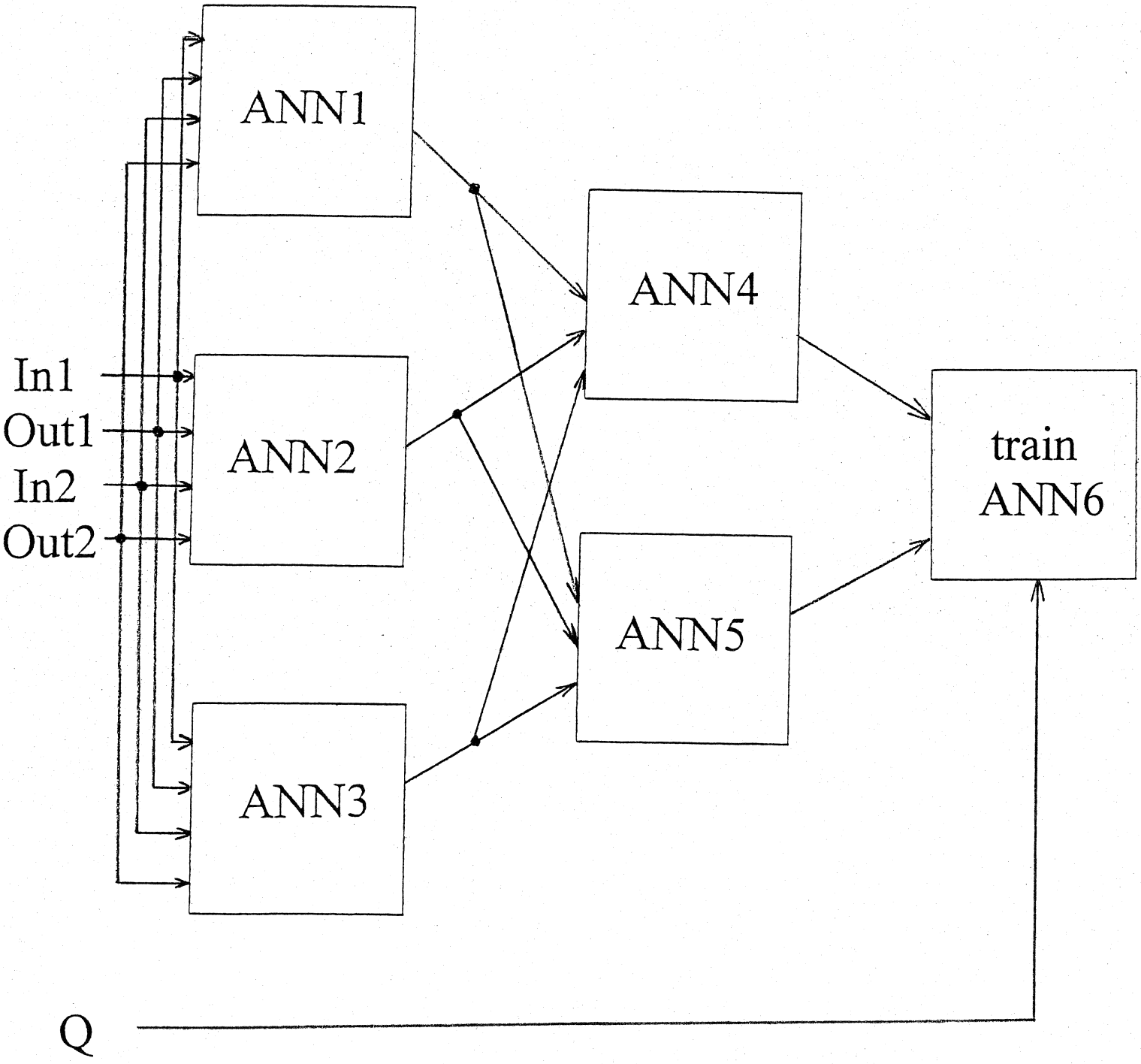
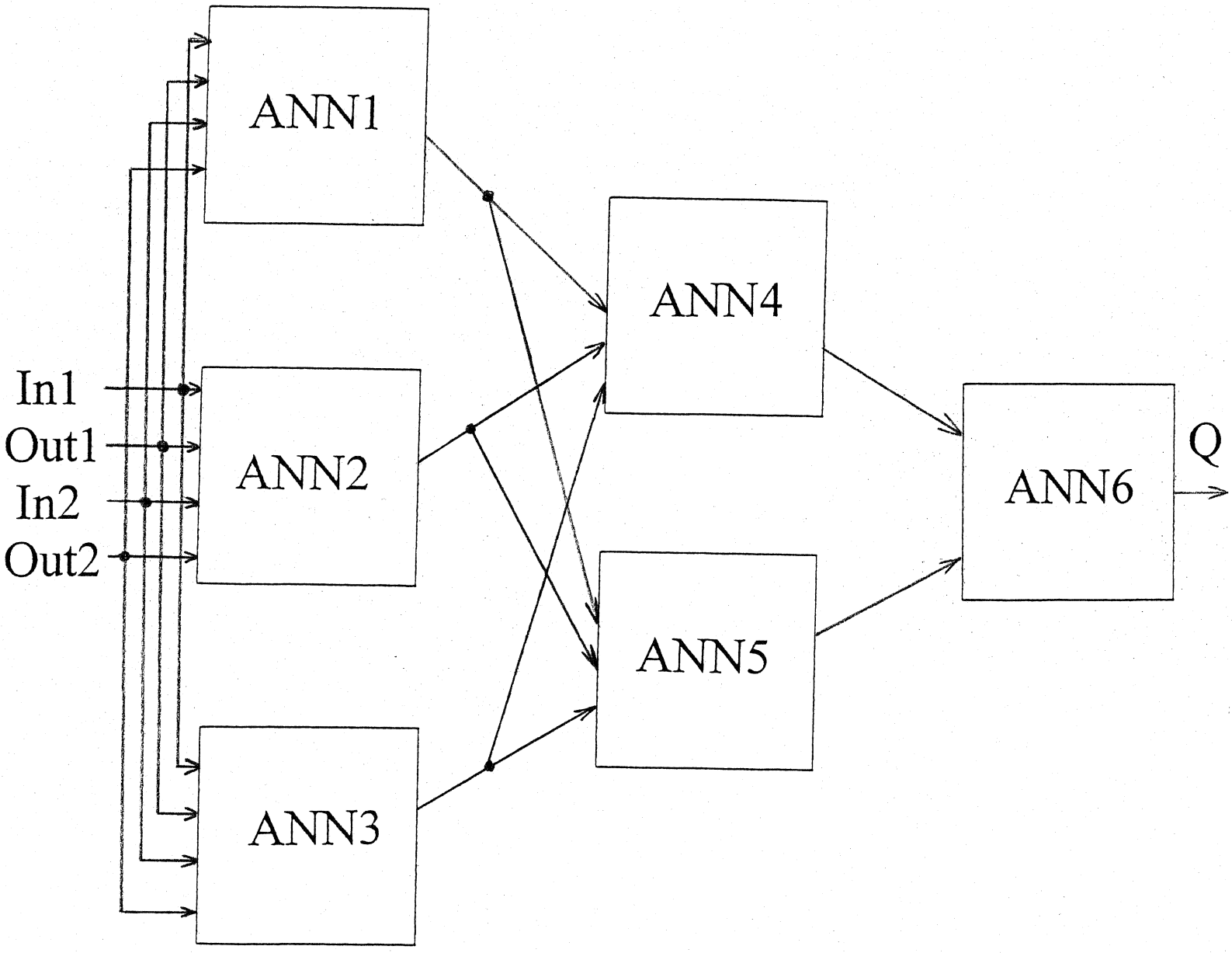


Fig 11



Neural Network Categorization of Experimental Data

R. Jones
Physics
Emporia State Univ.
Emporia, KS 66801

A neural network program is used to sort langmuir probe characteristics by category and data quality.

Introduction

Computerized data acquisition systems have the potential for generating huge databases. Our computer interfaced Langmuir probe diagnostic (El Jaeid, 1988) can, for instance, log far more probe characteristic curves than we could possibly hope to analyze by hand. In order to make better use of this data we have created a program which sorts the probe characteristics into a few fundamental categories and by quality. This software tool allows the experimenter to sample the database and automatically archive "typical" data sets in real time.

Langmuir Probe Characteristics

Electrostatic ("Langmuir") probes continue to be one of the most important and most widely used tools for making measurements in plasmas. Langmuir probes are routinely employed in order to measure plasma density, n_e, electron temperature, T_e, and plasma potential, V_p.

The theory of current collection by biased electrodes of various different shapes (i.e. flat metal discs, cylindrical wires, or conducting spheres) is well understood in the limit of a magnetic field free plasma (Hoyaux, 1968 and Schott, 1968).

A typical Langmuir probe consists of a metallic electrode which can be inserted into and electrically biased with respect to the plasma. As the probe is biased more and more negatively the electron current which is collected can be observed to decrease, figure 1. The electron temperature is obtained from the bias voltage required to reduce the probe current by a factor of 1/e whereas the density is proportional to the magnitude of the current itself.

Sorting Probe Characteristics by Category

In a Maxwell Boltzmann distributed plasma the Langmuir probe characteristic should be roughly exponential in the region of voltages $<V_p$ (figure 1). Various other probe characteristic shapes are possible, however. In the presence of a non-maxwellian plasma the probe characteristic

curve may resemble figure 2 and we may not wish to proceed with a calculation of the electron temperature from such a data set.

Noisy data may also be produced like that of figure 3. Again, conventional mathematical data reduction may not be warranted.

In the presence of magnetic fields probe characteristics like those of figures 4 and 5 are possible (Tagle, et al, 1987 and Guenther, et al, 1990). In each case subsequent mathematical analysis must be modified according to the theories of Guenther and Tagle.

Probe data may also be off scale, figure 6. This may occur if gas pressure is too low (the data will then likely be off scale along the V axis, figure 6) or if the discharge current is too high (the data will then likely be off scale along the I axis). Any of these (or other) categories can be learned by an artificial neural network program provided that a number of examples from each category of probe data are available for training the network.

The Neural Network

The artificial neural network has 100 inputs, 50 probe voltages and the 50 corresponding currents drawn by the probe. These are fully connected to 25 hidden units, (One can add or delete hidden units if the network does not train properly). The hidden units are, in turn, fully connected to 7 output units, one for each of the 7 defined output categories.

The 100 input quantities are all normalized so as to fall within the unit interval 0-1 and concatenated to form a 100 dimensional input vector, I_k. Each characteristic curve is then described by a vector I_k and an appropriate category (output vector) corresponding to figures.

The output vector used is ,

O_l = 1 0 0 0 0 0 0

for conventional Maxwellian plasmas,

$O_2 = 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0$

for nonMaxwellian plasma with high energy tail,

$O_3 = 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0$

for a magnetized JETLIKE plasma,

$O_4 = 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0$

for a magnetized T-10 like plasma,

$O_5 = 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0$

for a noisy plasma, and

$O_6 = 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0$

or,

$O_7 = 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1$

for offscale data sets.

For each input vector I_k the hidden unit j receives an input:

$$h_j = \sum_{k=1}^{100} w_{jk} I_k + w_o \qquad (1)$$

and produces an output:

$$U_j = \frac{1}{1+\exp(-h_j)} \qquad (2)$$

(We don't know that this function is optimum but we use it because it is mathematically convenient). W_o is an adjustable bias.

Output unit i then receives:

$$h_i = \sum_{j=1}^{25} w_{ij} u_j + w_o \qquad (3)$$

and computes the output vector:

$$O_i^1 = \frac{1}{1+\exp(-h_i)} \tag{4}$$

The initial values of the weights W_{jk} and W_{ij} are random numbers in the range between 0 and 1.

The computed theoretical output vector O_i^1 is compared with the actual vector category O_i and the error O_i-O_i^1 is "back propagated" in order to correct the weights W_{ij} and W_{jk} (Bishop, 1994).

The weights W_{ij} are corrected by amounts:

$$\Delta w_{ij} = c(o_i-o_i^1)\left(\frac{1}{1+\exp(-h_i)}\right)\left(1-\frac{1}{1+\exp(-h_i)}\right) u_j \tag{5}$$

and the weights W_{jk} are corrected by amounts:

$$\Delta w_{jk} = \sum_{i=1}^{7} c\,(o_i-o_i^1)\left(\frac{1}{1+\exp(-h_i)}\right)\left(1-\frac{1}{1+\exp(-h_i)}\right) w_{ij} \left(\frac{1}{1+\exp(-h_j)}\right)\left(1-\frac{1}{1+\exp(-h_j)}\right) I_k \tag{6}$$

where c, the "learning rate", is a constant, typically 0.1 or 0.2. (c may be reduced as training progresses.)

The weights are adjusted (corrected) at each calculation step by:

$$\text{new } w_{ij} = \text{old } w_{ij} + \Delta w_{ij} \tag{7}$$

$$\text{new } w_{jk} = \text{old } w_{jk} + \Delta w_{jk}$$

and the process is repeated, using the same single I_k, O_i until O_i - O_i^1 is smaller than some desired error. Adjustable noise can be added to help escape from local minima. This whole process is then repeated with I_k, O_i data from additional probe characteristics taken from each of the defined categories.

Training and Using the Network

The experimenter identifies 25 (or more) representative data sets corresponding to each of

the 7 categories. The neural network is then trained on these (175 or so) characteristics.

Once trained the neural network can be used to categorize new data sets in real time as they are taken. Select data can be referred to the experimenter for viewing or sent on for the appropriate data reduction software (providing temperature, density, or other output).

Literature Cited

Bishop, C.M. 1994. Neural networks and their applications. Rev. Sci. Instrum. 65: 1803-1832.

El Jeaid, I. 1988. Interfacing a Langmuir Probe to the Cromenco Microcomputer, Emporia State Univ., MSC project report

Guenther, K., A. Hermann, M. Laux, P. Pech, and H. Reiner. 1990. Characteristics of electric probes in tokamak SOL: influences of the magnetic field. J. Of Nucl. Mater. 176: 236-239.

Hoyaux, M.F. 1968. ArcPhysics. Springer-Verlag, New York. P. 258.

Schott, L. 1968. Chapter 11 in W. Lochte (ed). Plasma Diagnostics. Holtgreven, North Holland.

Stangeby, P.C. 1982. Effect of bias on trapping probes and bolometers for tokamak edge diagnostics. J. Physics 15D: 1007-1029.

Tagle, J.A., P. Stangeby, and S. Erents. 1987. Errors in measuring electron temperatures using a single Langmuir probe in a magnetic field. Plasma Phys. Cont. Fusion. 29: 297-300.

Figure Captions

Fig. 1. Langmuir probe current drawn, I, as a function of bias voltage, V. I_e=electron current as a function of bias voltage. I_i=ion saturation current.

Fig. 2. A Langmuir probe characteristic taken in a non-Maxwellian plasma.

Fig. 3. A Langmuir probe characteristic taken in a noisy (turbulent) plasma.

Fig. 4. A Langmuir probe characteristic taken in a magnetic field in the JET device.

Fig. 5. A Langmuir probe characteristic taken in a magnetic field in the T-10 experimental device.

Fig. 6. A Langmuir probe characteristic taken with low neutral gas fill pressure.

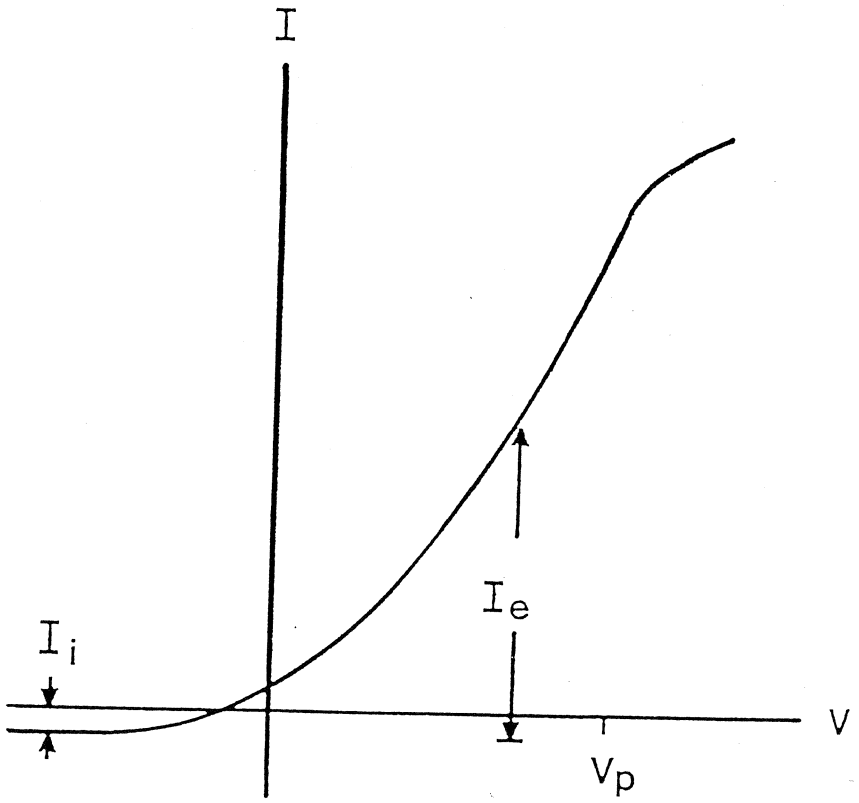

I
I_e
I_i
V
Vp

fig 2

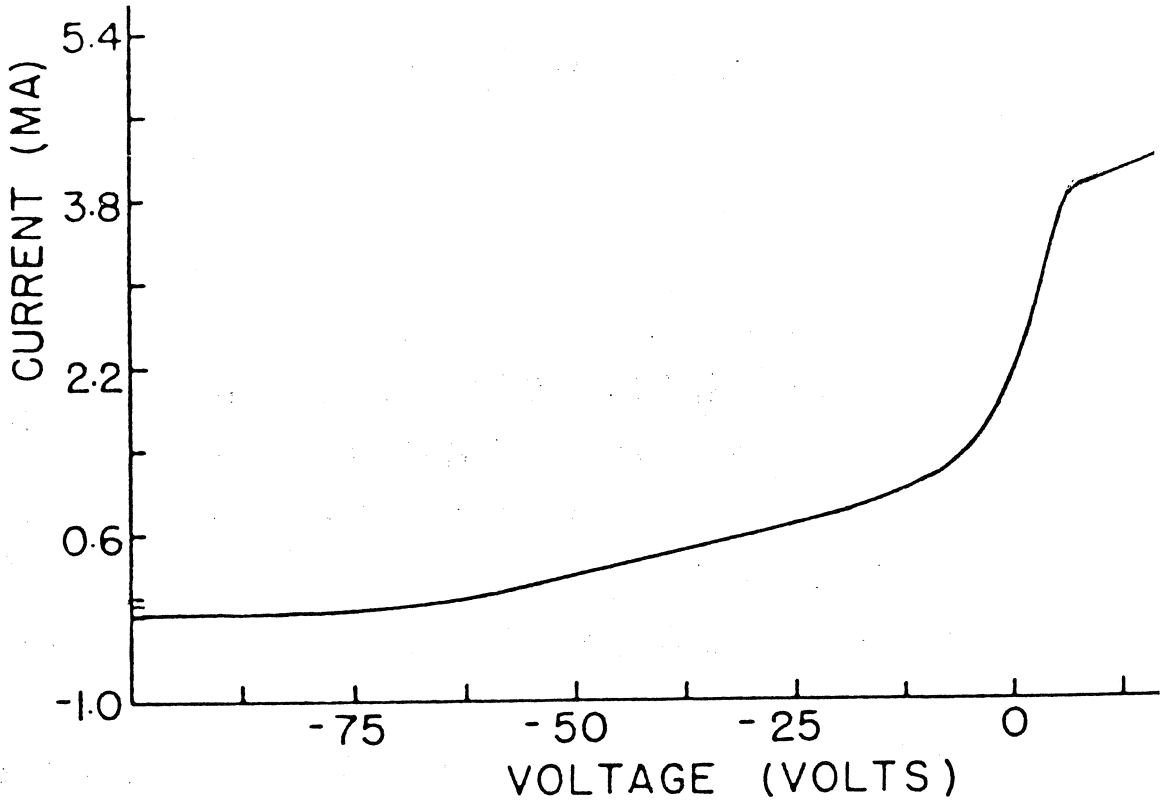


Fig 3

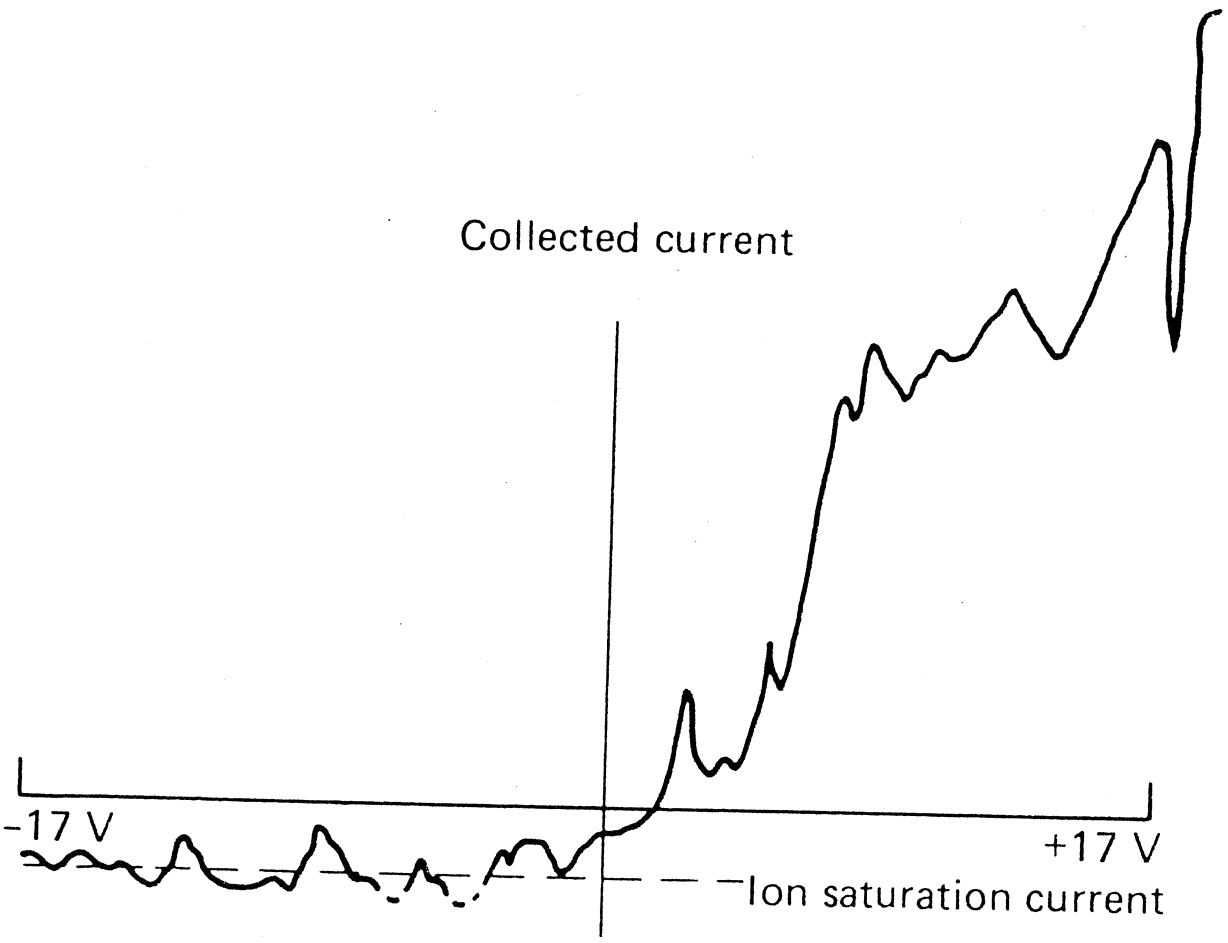


fig 4

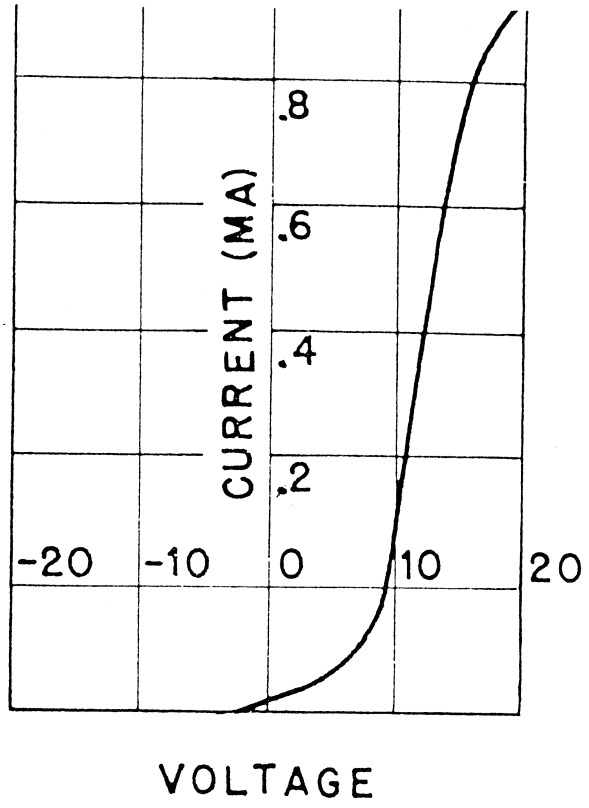


fig 5

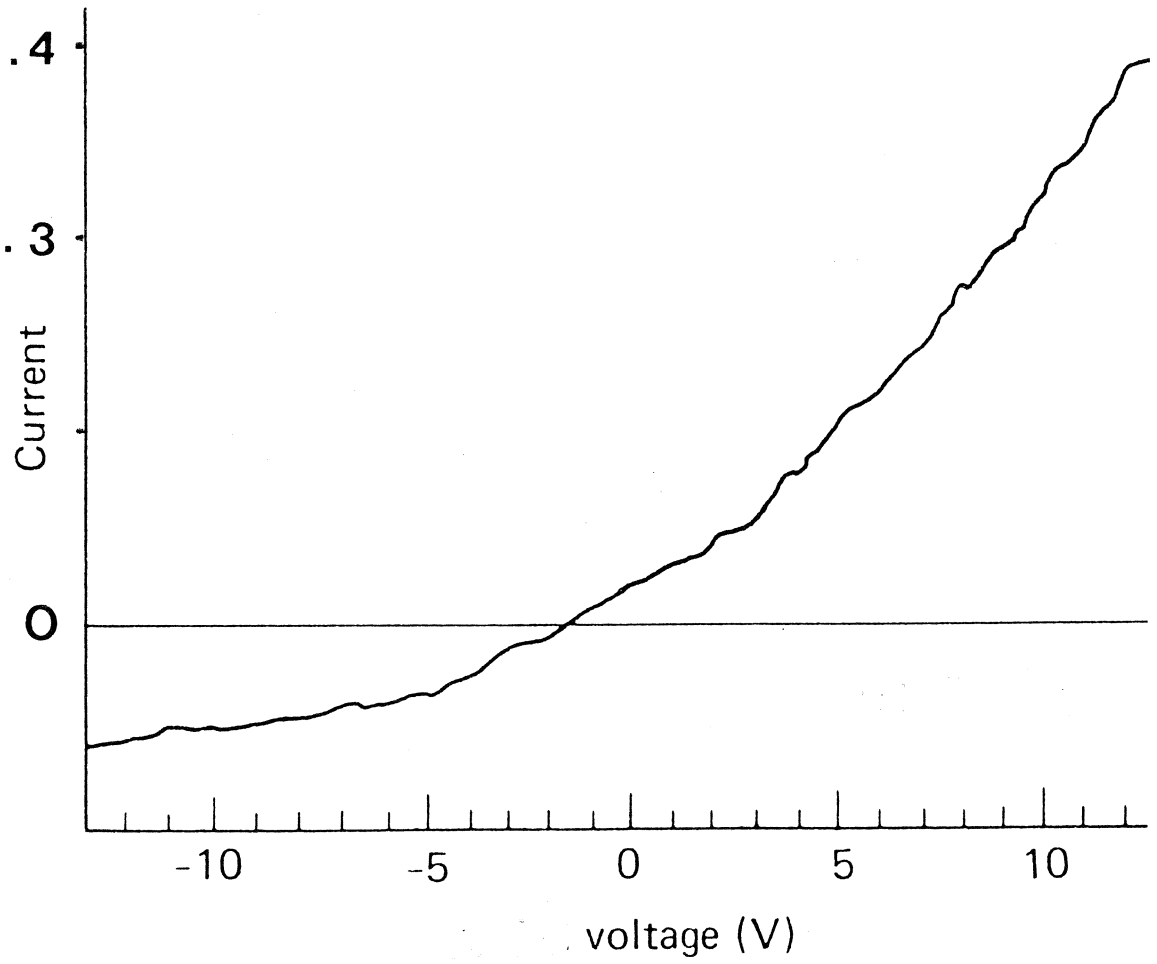
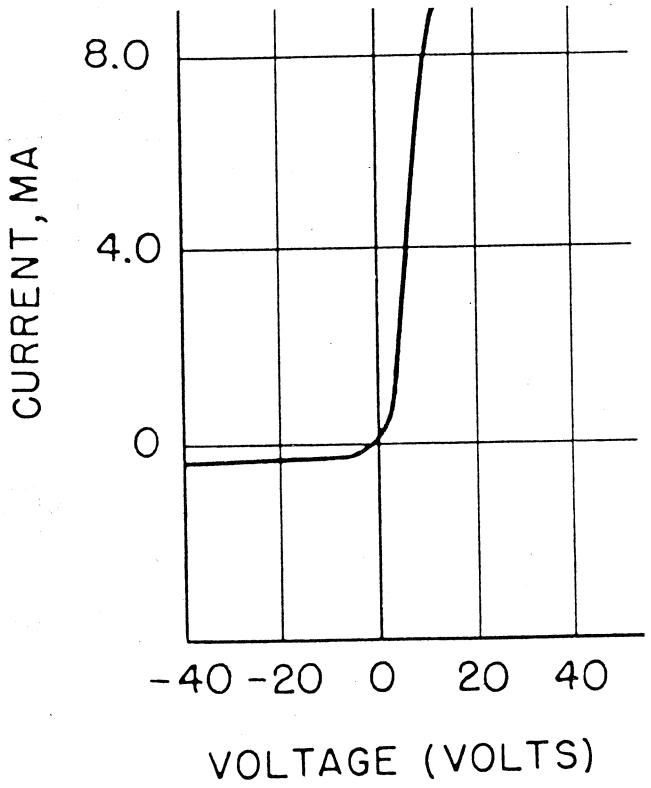


fig 6



Quiet Plasma

In Gas Mixtures

R. Jones

Physics Department

Emporia State University

Small amounts of argon gas introduced into helium plasmas results in reduced levels of plasma turbulence.

Introduction

Most plasmas exhibit substantial amounts of turbulence or "noise." Such oscillations modify the characteristics of a plasma and contribute to such things as the plasma's electrical resistivity and transport. In most cases it would be desirable to reduce the plasma noise, that is to "quiet" the plasma in some way.

In the early 1970s Bollinger (Bollinger, et al, 1972) reported that plasmas formed in certain gas mixtures were unusually quiet. There was little effort to reproduce these results, however (save for Michelsen and Hirshfield, 1974) nor was the work continued by Siedl's group.

Spurious Quiescence/ measurement defects

While working with the same device that Bollinger and Seidl employed Jones (Jones, 1975) occasionally also observed quiet discharges. There observations were, however, traced to conductive coatings building up on the Langmuir probes.

The plasma noise was measured using the circuit shown in Figure 1. The plasma density is roughly proportional to the D.C. voltage (V) measured across the resistor (R) and the noise (oscillation) level is given by the voltage fluctuation with time (figure 2), dV.

The probe tip is illustrated in figure 3. Over time a silver-gray coating would build up on the ceramic insulator (figure 3). This coating was composed of metal evaporated from the discharge filament (figure 4), sputtered from the vacuum chamber or component parts, and remnants of vacuum pump oil and vacuum grease. It was found to be electrically conducting and formed a conducting path from the probe tip to ground (figure 1).

As a consequence of the build up of the coating on the probe insulator the spurious electrical path results in a gradual increase in the D.C. voltage measured across R. This erroneously enhances the apparent plasma density while leaving the fluctuation level, dV, unchanged. The apparent reduction in noise, measured as dV/V was traced to the "dirty probe."

New Research

Dispite the doubts which have been raised there was still some reason for continued interest in Bollinger's result. The addition of small amounts of hydrogen to CO_2 plasmas was seen to quiet laser discharges and improve their performance (Trtica and Miljanic, 1994). Having quieter plasmas would also make experimental measurements easier and more accurate. For this reason we decided to repeat Bollinger's experiments over a wide range of conditions

Experiment

The experiments were performed in a low pressure, hot cathode, magnetized arc, figure 4. In this device electrons from the hot filament, C, are accelerated toward

grounded anodes, a, so as to strike and ionize a low pressure fill gas. Following Bollinger's work discharges were in helium into which we added small quantities of argon from a second gas feed line. Probes were cleaned regularly so that no contaminant built up. Electrical measurements verified that there were no substantial leakage currents to ground.

Adding a percent or two of argon to a helium discharge results in a measurable increase in the plasma density. This makes sense theoretically. In the steady state the ion loss along the magnetic field just balances the gas ionization rate:

$$\frac{1}{t} = n_n \sigma v_e \qquad (1)$$

Where n_n is the neutral gas density, σ is the electron-gas ionization cross section, and v_e is the (average) ionizing electron velocity. The average time for ion loss is:

$$t \cong \tfrac{1}{3} L / c_s \qquad (2)$$

Where L is the plasma column length and c_s is the ion acoustic speed:

$$c_s = \sqrt{\frac{T_e}{M_i}} \qquad (3)$$

Where Te is the plasma temperature and Mi is the ion mass.

With two gas components contributing equally to the plasma density:

$$n_{n_{He}} \sigma_{He} v_e L \frac{\sqrt{M_{He}}}{\sqrt{T_e}} = n_{n_{Ar}} \sigma_{Ar} v_e L \frac{\sqrt{M_{Ar}}}{\sqrt{T_e}} \qquad (4)$$

or:
$$n_{n_{He}} \sigma_{He} \sqrt{M_{He}} = n_{n_{Ar}} \sigma_{Ar} \sqrt{M_{Ar}} \qquad (5)$$

So:
$$\frac{n_{n\,Ar}}{n_{n\,He}} = \frac{\sigma_{He}\sqrt{M_{He}}}{\sigma_{Ar}\sqrt{M_{Ar}}} \approx \frac{8 \times 10^{-17}}{2.53 \times 10^{-15}} \approx 3 \times 10^{-2}$$

As small amounts of argon are added the plasma density increases and the noise, falls, figure 2. If large amounts of argon are added, however, (> 20%) the noise increases again.

Discussion

Although we have confirmed Bollinger's basic result we did make several new findings. In our device the anode is far from the filament cathode. We did not require the close anode-cathode fit which Bollinger specified (Bollinger, etal., 1972). We also extended our experiment into the region of cold cathode discharge operation. We found that as the cathode became cold new sorts of oscillations came to dominate the discharge (sparking at the cathode) and these could not be quieted by the addition of argon.

Literature Cited

Bollinger, D., D.Boyd, H. Liu and M. Seidl. 1972. Production of quiescent plasmas in a magnetic field by a beam-plasma discharge. Phys. Rev. Letts. 28 (12) : 723-726.

Jones, R. 1975. Nonlinear development of absolute and convective beam-plasma instability. Ph.D. thesis. Stevens Institute of Technology, Hoboken, N.J. 124p.

Michelson, P. and J.L. Hirshfield. 1974. Nonlinear excitation of ion acoustic waves. Phys. Fluids. 17 (12) : 2205-2207.

Trtica, M. and S.S. Miljanic. 1994. Investigation of gas chemistry in the sealed TEA CO2 laser. Chem. Physics. 181: 409-416.

Figure Captions

Figure 1: Langmuir probe circuit composed of probe wire, P, bias battery, b, resistor, R, grounded metal vacuum chamber, c, and oscilloscope, V.

Figure2: Plasma density measurement as a function of time in pure helium discharge, a, and in helium with 2% admixture of argon, b.

Figure 3: Langmuir probe consisting of wire tip, a, grounded coaxial conductor, c, and ceramic insulator, b. Dashed line indicates how conductive short circuit "leakage path" can develop.

Figure 4: Discharge device consisting of negatively biased filament cathode, c, grounded anode, a, and permanent magnets, m. This assembly sits in the metal vacuum chamber of figure 1.

Fig 1

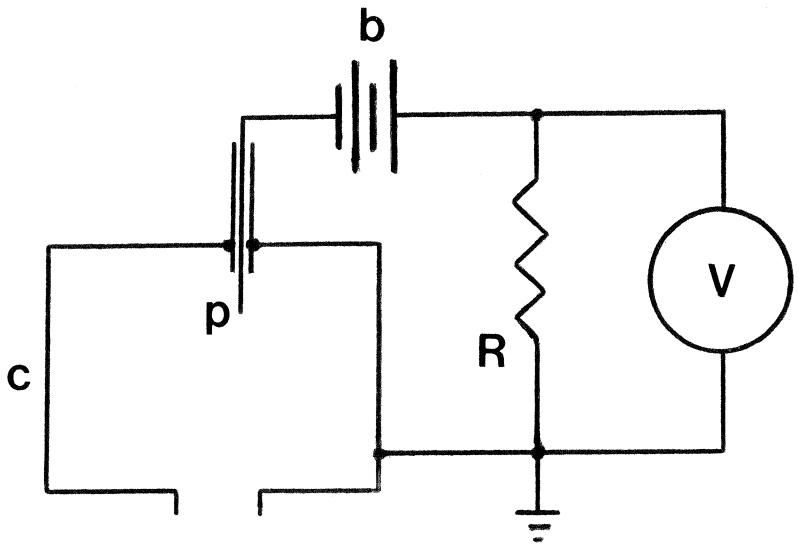


Fig 2

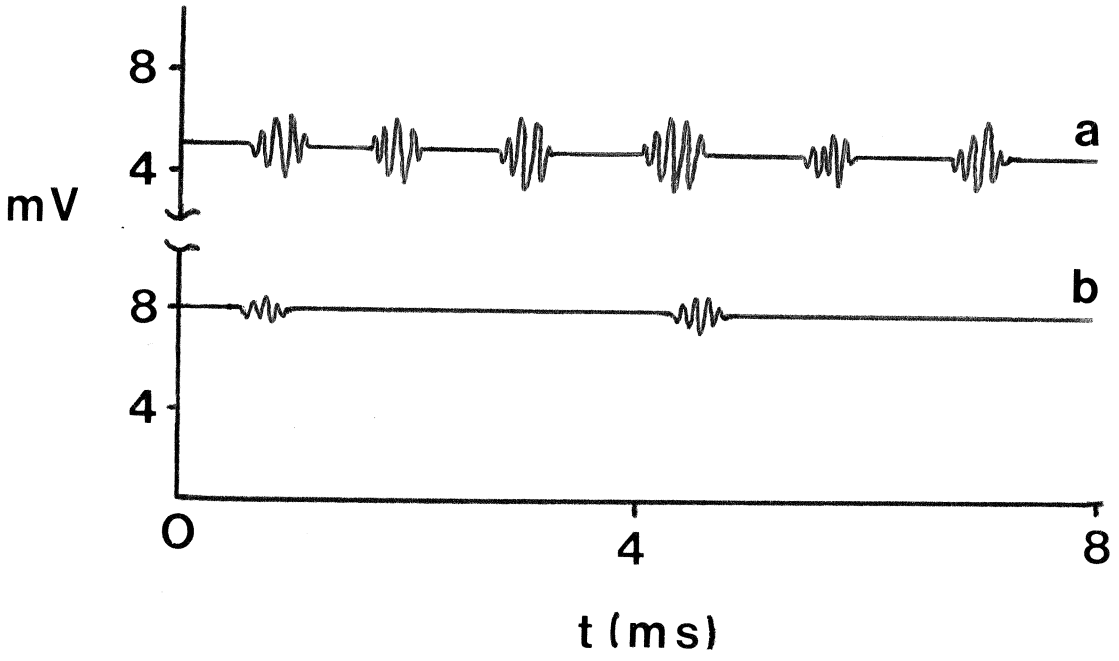
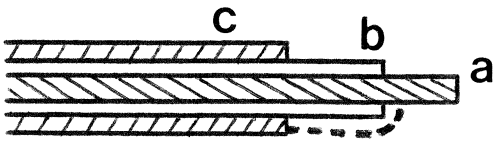
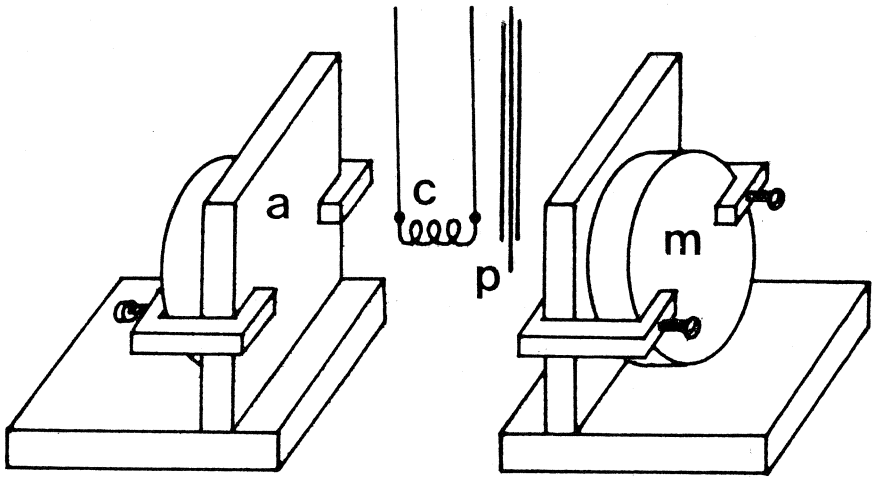


Fig 3




a
c
p
m

Plasma Optimization Using Data Mining

R. Jones
Physics Department, Emporia State University
Emporia, Kansas 66801

Data mining has been employed to model the performance of a plasma confinement device. Plasma density, temperature, and confinement are predicted as functions of the control parameters: magnetic field strength, neutral gas pressure, and discharge voltage.

Introduction

Our understanding of plasma transport in magnetic confinement devices is sketchy. Neither analytic nor computational models are sufficiently well developed to describe accurately the behavior of real experimental systems. Artificial neural networks (Bishop, 1994), on the other hand, are capable of modeling complex input-output relationships learned on the basis of empirical examples alone.

As an initial effort to apply data mining technology to plasma optimization and modeling we have employed data taken from a small research stellarator (Jones, 1985).In the experiment which we wish to model, it is possible to vary the magnetic field used to confine the plasma, B; the neutral gas fill density (Hydrogen), N_n; and the voltage (V) used to sustain the discharge. The experiment is like that of Roth (1976) but in a small stellarator. The discharge is sustained by biasing a probe or limiter with respect to the chamber wall.

The "output" of our experiment is a set of three parameters, the plasma electron temperature, T_e; measured in electron volts, the plasma density, N_e (the number of electrons and ions per cubic centimeter of volume); and the confinement time, t, in seconds that the plasma is confined in the device (Jones, 1985). We are interested in optimizing (maximizing) a figure of merit $T_e\ N_e\ t$ but we chose a three component output vector $O_i' = (T_e, N_e, t)$ in order to track the variation of all three output quantities individually.

In the future we may wish to expand this "dependent variable set" in order to include ion temperature and fluctuation level but (T_e, N_e, t) seemed a reasonable compromise at the present time.

Mathematical Method

The neural network consisted of 3 input units, one each for B, N_n, and V. These are connected fully to 10 hidden units. The hidden units in turn, are, connected fully to 3 output units, one each to represent T_e, N_e, and t. B, N_n, V, T_e, N_e, and t are all normalized so as to fall within the unit interval.

For each vector I_k the hidden unit j receives an input:

$$h_j = \sum_{k=1}^{3} w_{jk} I_k + w_o \quad (1)$$

and produces an output:

$$v_j = \frac{1}{1+\exp(-h_j)} \quad (2)$$

w_0 is an adjustable bias. We typically have employed values of $w_0 = -1$.

Output unit I then receives:

$$h_i = \sum_{j=1}^{10} w_{ij} V_j + w_0 \quad (3)$$

and computes the output vector:

$$O_i = \frac{1}{1+\exp(-h_i)} \quad (4)$$

The initial values of the weights w_{jk} and w_{ij} are random numbers in the range between 0 and 1.

The computed theoretical output vector O_i then is compared with the actual measured vector O_i' and the error $O_i' - O_i$ is back propagated in order to correct the weights w_{ij} and w_{jk} (Bishop, 1994).

The weights w_{jj} are corrected by amounts:

$$\Delta W_{ij} = C(O_i' - O_i)\left(\frac{1}{1+\exp(-h_i)}\right)\left(1 - \frac{1}{1+\exp(-h_i)}\right) V_j \quad (5)$$

and the weights w_{jk} are corrected by amounts:

$$\Delta W_{jk} = C(O_i' - O_i)\left(\frac{1}{1+\exp(-h_i)}\right)\left(1 - \frac{1}{1+\exp(-h_i)}\right)$$

$$\text{x } W_{ij}\left(\frac{1}{1+\exp(-h_j)}\right)\left(1 - \frac{1}{1+\exp(-h_j)}\right) I_k \quad (6)$$

where C is constant, typically 0.2.

The weights are adjusted (corrected) at each calculation step by:

$$\text{new } w_{ij} = \text{old } w_{ij} + \Delta w_{ij} \quad \text{and} \quad \text{new } w_{jk} = \text{old } w_{jk} + \Delta w_{jk} \quad (7)$$

and the process is repeated, using the *same single* I_k, O_i' vector pair, until $O_i' - O_i$ is smaller than some desired error. This process then is repeated with *each additional* available training vector pair (I_k, O_i').

Training on subsequent I_k, O_i' data sets reduces the network's ability to reproduce previous I_k, O_i' pairs. For this reason, repeated training "passes" are made through the complete training set until $O_i' - O_i$ is acceptably small for *all* data sets employed. (The acceptable error $O_i' - O_i$ is selected based on a knowledge of the typical *experimental* accuracy obtainable for O_i').

The neural network was trained on ~50 data sets with B ranging from 500 to 2000 Gauss N_n varing from $\sim 10^{-5}$ to 10^{-3}mm and V varying from ~200 to 3000 Volts.

DISCUSSION AND RESULTS

Once trained, the neural network model was tested by inputing a number of vectors (B, N_n, V) which had not been employed during the training process. The number of training sets required to achieve a given accuracy of representation depends strongly on the smoothness of the function being modeled (Poggio and Frederico, 1991).

After having been developed in this way the neural network model can be explored by inputing any desired (B, N_n, V) vector and recording the resulting O_i. If the object of the work is to maximize the product T_eN_et then we learn that in fact several *different* optimal discharge "modes" are represented in the neural network model. In one mode that we have identified, we determine that T_eN_et is maximized for maximum input B. N_n, as might be expected, however, should be of intermediate value, not maximal.

Unlike Roth (1976) who found a monotonic increase in plasma density in a similar experiment:

$$N_e = CV^{1.1}N_n^{.4} \qquad (8)$$

Our data mining program suggests that there exist optimal settings of the discharge voltage and gas fill pressure, N_n, which leads to maximization of N_e, τ, and $Tn\tau$, figures 1-3.

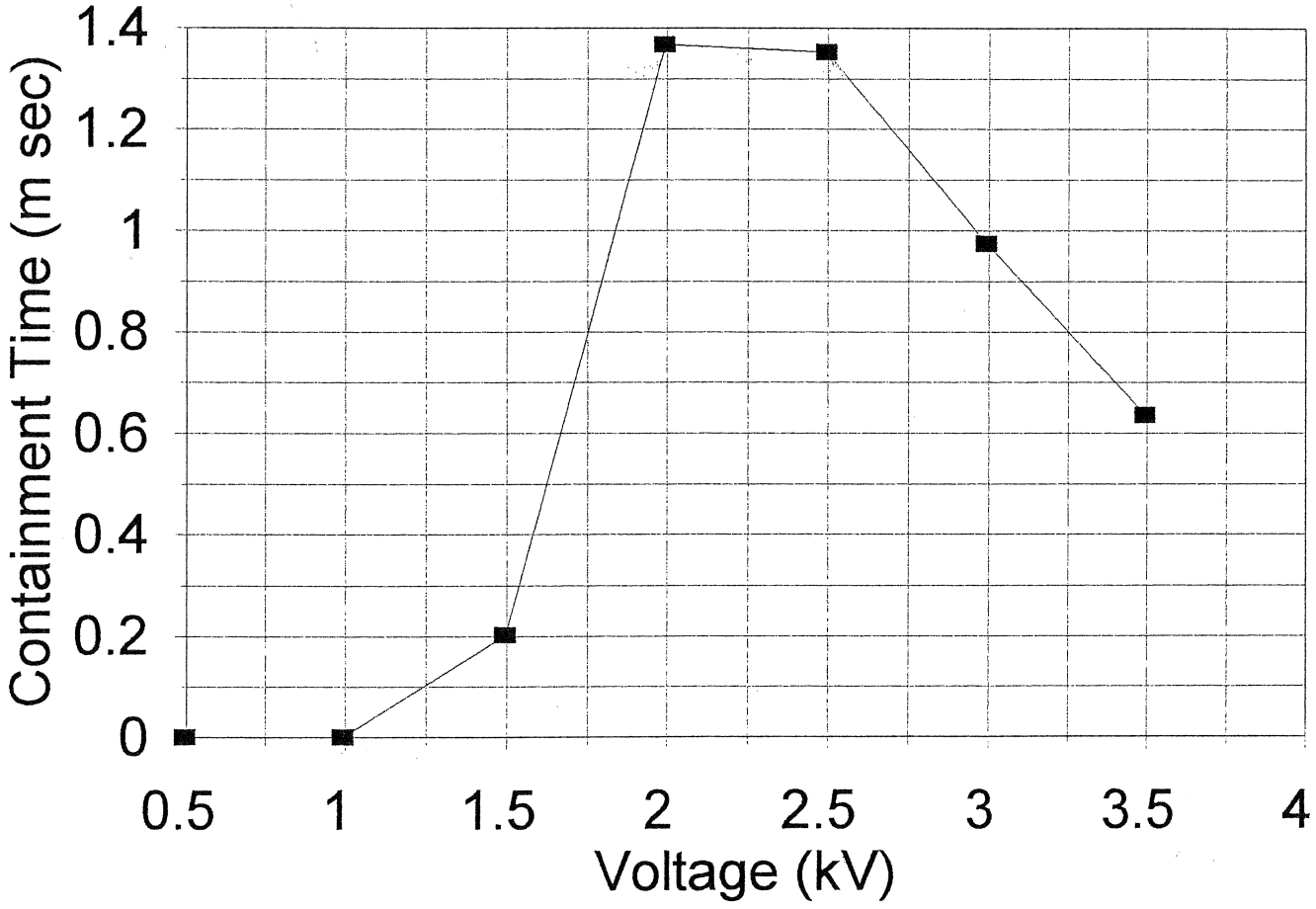
Literature Cited

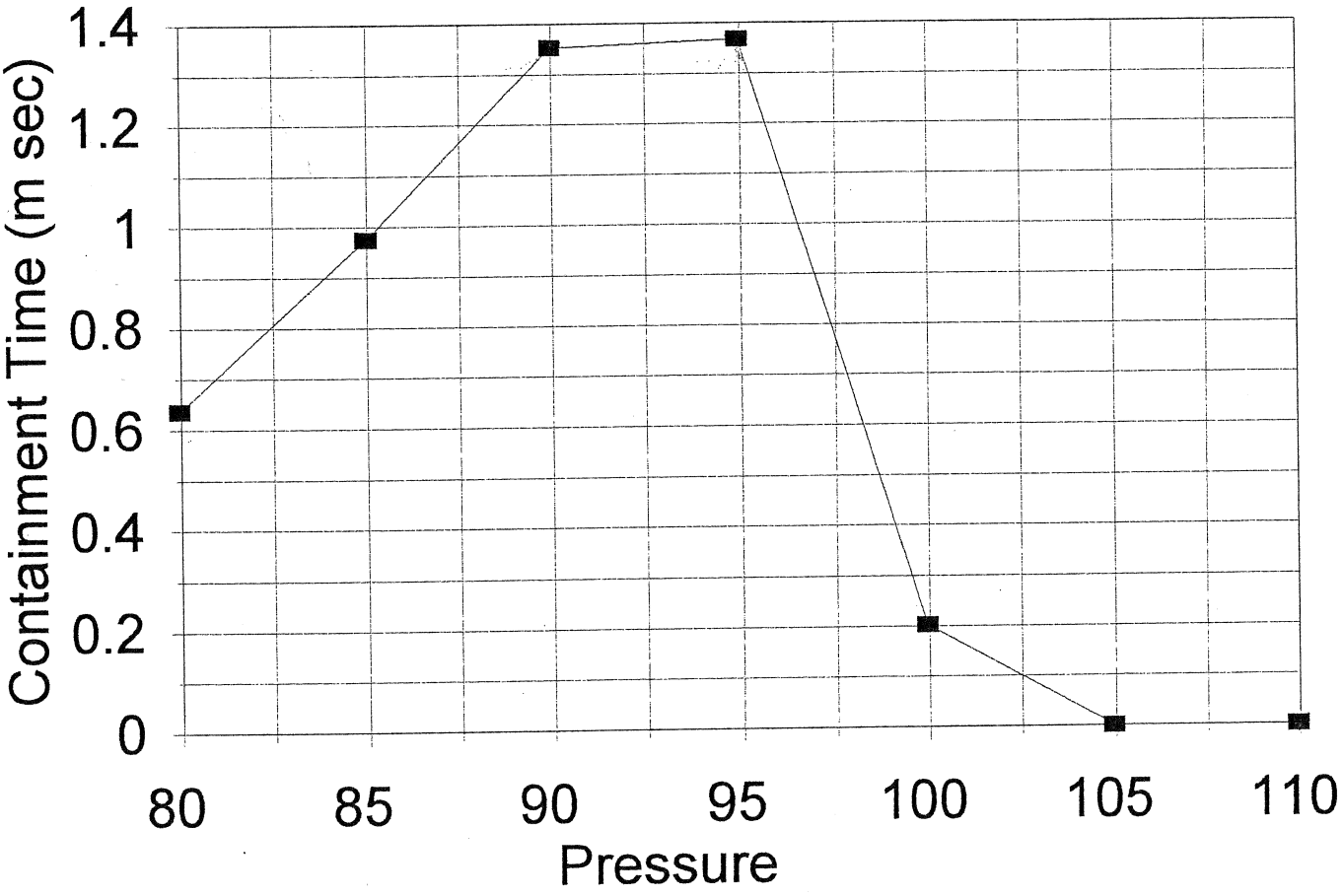
Bishop, C. M. 1994. Neural networks and their applications. Rev. Sci. Instrum. 65: 1803-1832.

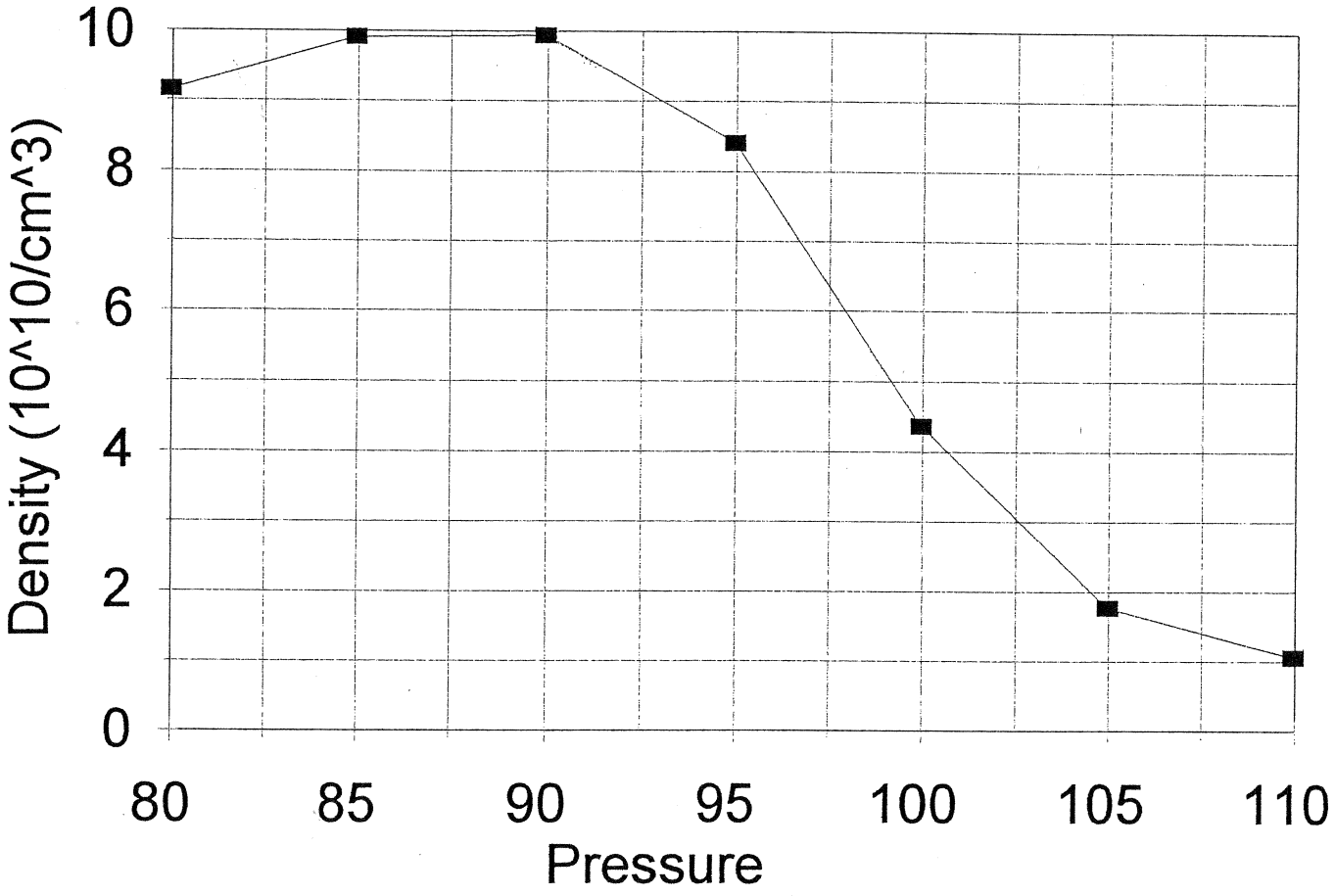
Jones, R. 1985. Magnetoelectrostatic confinement in RSX-I. letter Nuovo Cimento. 42:357-362.

Poggia, T. and G. Frederico. 1991. Hyper BF: a powerful approximation technique for learning. Machine Intelligence III:271-285.

Roth, J. R. 1976. Preliminary scaling laws for plasma current, ion kinetic temperature, and plasma density in the NASA Lewis bumpy torus plasma. NASA Tech. report TM X-73434, 28pgs.


Containment Time (m sec)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
3
3.5
4
Voltage (kV)


Containment Time (m sec)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
80
85
90
95
100
105
110
Pressure


Density (10^10/cm^3)
10
8
6
4
2
0
80
85
90
95
100
105
110
Pressure

The origin of large scale fluctuations in a Roth torus

R. Jones
Physics Dept,
Emporia State University
Emporia, KS 66801

Abstract- The Roth torus plasma confinement device exhibits large scale fluctuations which are attributed to unipolar arcing. These are undesirable and make experimental measurements difficult at best.

Introduction

Toroidal magnetic confinement systems are among the most common and most successful devices for confining hot fusion plasmas. Roth has suggested that electrode biasing of such plasmas can improve both their temperatures (J.R. Roth, 1983) and confinement (J.R. Roth, 1992) by an order of magnitude.

We decided to explore this idea in our small stellarator device (R. Jones, 1985). The device is illustrated in figure 1 where the plasma is electrically biased in the manor described by Roth by inserting an electrode L and biasing it with respect to the grounded metal machine vacuum chamber.

Experimental Result

The initial results were difficult to assess. With typical bias voltages of 1000-5000 Volts DC (either polarity) the plasma was extremely noisy. Both Langmuir probes and electrostatic energy analyzers (R. Jones, 1978 and 1979) showed plasma fluctuations that were approximately 100% of the plasma density itself. Few meaningful measurements were possible under these conditions. Optical diagnostics like hydrogen alpha light emission exhibit similar large fluctuations.

We have been conducting a study to try to identify the origin of these large scale oscillations and believe we have now identified their cause. We frequently observe localized flashes on the metallic interior of our device. These occur near sharp edges, corners, and scratches and are only seen in the biased plasma experiments. They are not typical of our previous RF plasmas.

Microscopic examination of the regions which exhibit the light flashes show the pits and tracks characteristic of unipolar arcing (Holliday, 1966). We have tested this hypothesis in

several ways. Firstly, we have cleaned the metallic surfaces chemically, mechanically, and by discharge cleaning. In each case the flashes are reduced as is the amplitude of the plasma oscillations. Secondly, we have inserted clean and oxidized metallic limiters into the plasma. The "dirty" (oxidized) samples produce flashes and enhance the fluctuation level seen on our diagnostics as expected if the origin is unipolar arcing.

Conclusions

We have identified unipolar arcing in our Roth torus experiment. We find that this makes it difficult to observe the increased plasma temperatures and confinement Roth predicted. Some reduction in this fluctuation level is possible if the discharge chamber is repeatedly cleaned. Unfortunately our machine is intricate and is not easy to clean nor is it easy to keep oxide deposits from accumulating.

References

J.R. Roth, Electrostatic biasing and radially inward transport in a toroidal plasma, IAEA technical committee meeting on tokamak plasma biasing, Montreal, 8-10 Sept. 1992, pg 1-56.

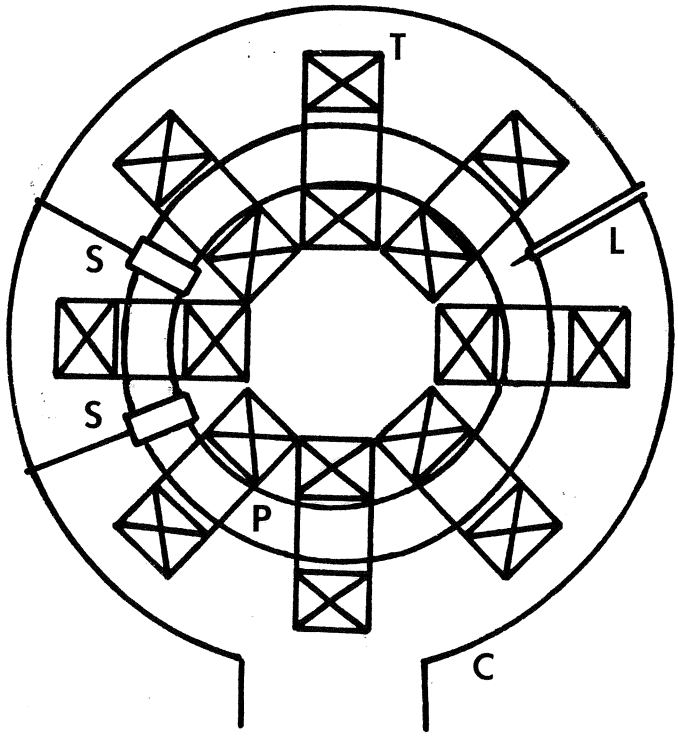
J.R. Roth, The electric field bumpy torus, Advanced Bumpy Torus Concepts, Oak Ridge National Lab, Oct. 1983, N.A. Uckan, Ed., Pg. 327-362.

R. Jones, reviews of Scientific Instruments, vol. 49, pg. 21, 1978 and vol. 50, pg. 392, 1979.

J.H. Holliday and G.G. Isaacs, Arc initiation at metal surfaces in a hydrogen Penning discharge, British Journal of Applied Physics, Vol. 17, pg 1575-1583, 1966.

Figure Caption

Plasma device with toroidal magnetic field coils, T, toroidal plasma column, P, vacuum chamber, C, probe, L and limiters S.


T
L
S
S
P
C

Plasma Diagnostics

R. Jones

Introduction

Plasma diagnostics have been the subject of numerous texts. The purpose of the present article , then, is to introduce the subject and to present those techniques best suited to the austere research environment found in University laboratories in general and developing countries in particular.

The Langmuir Probe

A simple metal electrode inserted directly into the plasma and biasable with respect to (machine) ground is known as a "Langmuir Probe." The current versus voltage curve obtained from the electrode is called the Langmuir probe characteristic.

A detailed theory of the Langmuir probe is still being developed to this day and we will offer only an introductory discussion here. Anyone who is interested (or confused) is refered to the text "Electric Probes In Stationary and Flowing Plasmas"[1].

The "typical" (or ideal) Langmuir probe characteristic is shown in figure 1 and can be divided into 3 regions, the "electron saturation" region A, the "active region" B, and the "ion saturation" region C. The "positive" probe current I in this diagram indicates electron current flow to the probe while -I indicates ion flow. Since the ion mass far exceeds the electron mass ions generally move much more slowly and the ion current density is small compared to the electron current density. "V_p" indicates the so called "plasma potential" or "space potential" obtainable directly from the probe characteristic . The point of transition between regions A and B may not be a sharp one and V_p is often estimated by extrapolating the slopes from each region (dashed lines) in order to obtain a well defined intersection point. "$I_{e_{sat}}$" is the probe current corresponding to a probe bias potential $V = V_p$ and is called the "electron saturation current."

With the probe biased to the same potential as the plasma itself (i.e. V_p) the average electron current flowing to the probe will be:

$$I_{e_{sat}} = A\, n_e\, e\, \bar{V}_e \tag{1}$$

where the average electron velocity directed toward the probe (of area A) is:

$$\bar{V}_e = \int_o^\infty V_e f_e(V_e)\, dV_e \tag{2}$$

with $f_e(V_e)$ the (normalized) Maxwellian velocity distribution function of plasma electrons:

$$f_e(V_e) = \left(\frac{2\pi KT_e}{m_e}\right)^{-3/2} \exp\left(\frac{-\frac{1}{2}mV_e^2}{KT_e}\right) \tag{3}$$

having temperature T_e. K is Boltzmann's constant, e is the electronic charge, and T_e may be expressed in ^0K or in electron volts (1 eV = 11,605^{0}K). Performing this integration we obtain:

$$I_{e_{sat}} = A\, n_e e \left(\frac{KT_e}{2\pi m_e}\right)^{\frac{1}{2}} \tag{4}$$

from which we can calculate n_e the plasma density (in particles per cubic centimeter) once T_e has been determined.

V_f (figure 1) is called the probe "floating potential" and is the probe voltage at which the total probe current I (ion plus electron) just equals zero. This is the voltage an unbiased probe will acquire when it is inserted into a plasma.

The plasma electron temperature is obtainable from the probe characteristic in the region B. In region B the probe voltage, V, is negative with respect to the potential of the plasma (V_p) and slow electrons of energy

$$\tfrac{1}{2}mV_e^2 < e(V_p - V) \tag{5}$$

are unable to climb the potential barrier V_p-V and reach the probe. The probe electron current in this region will then be:

$$I_e = Ae\, n_e \int_{V_{min}}^{\infty} V_e f_e(V_e)\, dV_e \tag{6}$$

where

$$\tfrac{1}{2}mV_{min}^2 = e\,(V_p - V) \tag{7}$$

If we transform equation 6 to energy variables:

$$I_e = \frac{Ae\, n_e}{m} \int_{eV}^{\infty} f_e(E_e)\, dE_e \tag{8}$$

and we observe that the electron energy (not velocity) distribution function is proportional (in region B) to the derivative of I_e with respect to the probe potential (i.e. (i.e. $edV \propto dE_e \alpha V_e dV_e$):

Integrating 6 we obtain:

$$I_e(V) = Ae\, n_e \left(\frac{KT_e}{2\pi m_e}\right)^{\frac{1}{2}} \exp\left[\frac{e(V-V_p)}{KT_e}\right] \tag{9}$$

If we neglect the ion current to the probe (which is small) then $I \approx I_e$ and this equation suggests that a plot of ln(I) versus V (in region B) should give a straight line (if the electrons are Maxwellian distributed) of slope e/KT_e. T_e in electron volts can then be read directly from a semilog replot of the

probe characteristic (region B).

For simplicity we have assumed a simple one dimensional (plane) geometry. In practice various probe geometries are employed. Cylindrical probes can be fashioned by simply baring the tip of the center conductor of a semiridged coaxial cable. (BA 50250 copper coax available from Sealectro Limited, Walton Rd., Farlington, Portsmouth, Hampshire, U.K. is vacuum tight and can be inserted through a vacuum wall via an Edwards High Vacuum 08-C006-00-000 Rotary Vacuum Seal). Since A can be enhanced by the sheath dimensions a flat probe (with $A >> \lambda^2_{Debye}$) is preferred for absolute density measurement. A cylindrical probe is, however, less sensitive to primary (nonthermal) electrons.

Probe cleanliness is important for measurements of all kinds and is usually insured by discharge cleaning. This involves biasing the probe very positive so as to heat it red hot. If the plasma is not too tenuous a very negative potential can also be used for ion sputter cleaning.

Multiple probe techniques have been evolved to solve particular problems. Double probes reduce the peak probe current and so minimize perturbations to the plasma as well as probe heating[2]. Triple probes allow real time measurements of the plasma parameters in pulsed discharges[3]. Langmuir probes are also frequently used (at various D.C. biases) to detect RF oscillations, either as fluctuations in the probe current or voltage.

A floating hot probe will emit thermionic electrons and tend to charge itself positively until its potential becomes comparable to the plasma space potential[4]. An ohmically heated probe (i.e. a filament immersed in the plasma) then will "float" at the plasma potential when made sufficiently hot and at the usual floating potential (figure 1) when cold. Such "emissive" probes are useful in making D.C. electric field measurements.

Energy Analyzers:

Although electron and ion beams are detectable with flat Langmuir probes the signal due to thermal ions is swamped by the plasma electron current. The simplest energy analyzer[5], then, is a flat probe ("collector") preceeded by a biasable selector ("discriminator") grid. Additional grids are usually added to control the plasma-analyzer sheath, suppress secondary emission at the collector, or improve resolution.

A typical ion analyzer will have a metal case and entrance grid which are floating, a second ("control") grid which is biased sufficiently negative to repel most of the electrons which were able to enter through the plasma-analyzer sheath, and a collector plate which is (voltage) swept in the manner of a Langmuir probe. Alternately, the collector can be held at fixed bias while a control grid is swept.

To minimize edge effects the grid diameter is made large compared to the spacing between the grids and collector. To assure a continuous sheath the entrance grid must have a mesh which is fine compared to the plasma Debye length.

Analyzer resolution is limited by mechanical constraints, field penetration of the grids, space charge effects, and possibly sputtering and secondary emission (important if ion energies exceed ~10 eV). Secondary emission is usually suppressed by adding a (negatively biased) grid just in front of the (secondary emitting) collector.

Placed at floating potential the entrance grid admits roughly equal currents of electrons and ions. Space charge is not neutralized beyond the discriminator grid, however, and if the current flowing in this (subsequent) region exceeds the Child Langmuir value:

$$J = \frac{CW^{3/2}}{d^2} \qquad (10)$$

(where W is the particle energy and d is the grid-collector spacing) then a virtual cathode (or anode) will form and distort the transmitted signal.

The performance of the analyzer can also be limited by field penetration, that is, the effect of a nearby electrode on the potential present between the wires of the control grid. Better resolution requires higher values of $\frac{r}{a}$ and $\frac{d}{a}$ where r is the grid wire radius and a is the spacing between grid

wires. Sensitivity depends upon keeping J, the analyzer current density, large, so $\frac{r}{a}$ can not be raised (this would reduce grid transparency). $\frac{d}{a}$ then should be increased. This can be done by using a fine mesh (which is limited, ultimately, by heating and mechanical resistance to warping) or a larger analyzer. Increasing d is also seen to conflict with the space charge limitation, equation 10.

Energy analyzers can also be biased so as to reject both electrons and ions and to detect energetic neutrals by the secondaries they generate at the collector[6].

Magnetic Probes:

1. Pulsed currents and magnetic fields can be measured with external "pick-up coils" or probes. A simplified "Rogowski belt" or "coil" is illustrated in figure 3a. The Rogowski coil consists(usually) of a large aspect ratio torus, evenly wound, with each turn (locally) perpendicular to the torus. By Ampere's law such a "belt" gives an output voltage which is proportional to $\int \dot{\vec{B}} \cdot \vec{d\ell}$ where $\dot{\vec{B}}$ is generated by the total time varing current enclosed by the toroid. Such a coil is typically attached to an integrator as shown in figure 3a. For sufficiently large resistance R the output voltage is then

$$\frac{1}{C}\int \frac{M}{R}\frac{di}{dt}dt = \frac{M}{RC}\, i \qquad (11)$$

where i is the enclosed (plasma) current and M is the mutual inductance between the plasma loop and the coil. Only the total enclosed current can be so inferred. Current profile measurements require the fabrication of small diameter probes which can be inserted into and moved about within the plasma. M is typically determined by calibration with a known A.C. current flowing in a wire inserted through the coil. For a sufficiently large number of even toroidal windings M is independent of the spatial location of this calibration current (or the wire size) within the toroid.

The coil windings must be closely wound so as to enclose a tube of flux but the belt itself can be deformed into any convenient shape which encloses the current to be measured. The minor cross-section of the belt can also be noncircular.

A large integration time constant, RC, must be employed so as to meet conflicting requirements of sensitivity, rate of exponential decay of the output voltage, and the inherent "frequency" of the current being measured, i(t). Additional gain is often incorporated in the form of a "Miller integrator" replacing the simple passive filter circuit of figure 3.

Any variable external magnetic fields (such as pulsed plasma confinement fields) can induce error signals in a simple Rogowski coil. "Back winding" is used as a first order correction for this effect, figure 3b. An improved system corrects for

placement of the back wound conductor, figure 3c. The potentiometer is set to minimize the output in the absence of plasma but with the ambient (confining) fields on.

2. A simple wire ("diamagnetic") loop can be used to measure the time rate of change of magnetic flux and so infer the plasma dia- or para- magnetism. The energy content of high pressure plasmas can be obtained in this way. Such loops are also employed to detect magnetic fluctuations. As with Rogowski coils diamagnetic loops may be mounted externally or inserted directly into the plasma. A typical loop might consist of a plastic spool $\sim$1 cm in diameter by $\sim$1 cm long wound with 10-100 turns of copper magnet wire ($\sim$ #30) insulated and wrapped with a $\sim$10 micron sheet of metal foil to provide electrostatic shielding. The shield is grounded and the coil output is amplified with an instrumentation (operational) amplifier. Calibration techniques are described in Lovberg [7].

Microwave Interferometry:

In a plasma dielectric microwave radiation suffers a phase shift

$$\Delta\phi = \int_0^L \left[\frac{\omega}{C} - \frac{\sqrt{\omega^2 - \omega_{pe}^2}}{C}\right] dx \qquad (12)$$

where ω is the microwave frequency and ω_{pe} is the plasma frequency.

$\Delta\phi$ can be measured by an interferometer, figure 4, from which the plasma density can then be calculated.

In a real (inhomogeneous) plasma $n_e(x) = n_e f(x)$ and with $\omega_{pe} << \omega$:

$$n_e = \frac{C \quad \Delta\phi \quad \omega \, m_e}{e^2 2\pi \int_o^L f(x)\,dx} \tag{13}$$

$f(x)$ can be obtained by radial Langmuir probe scans.

Horns are commonly used for launching the signal across the plasma. The transmitted signal and the reference are added in a magic tee and detected by a crystal diode and high gain amplifier. The reference signal amplitude and phase are varied to obtain a null. This is then repeated in the absence of plasma to obtain $\Delta\phi$.[8]

A square wave modulated microwave signal is useful to reduce errors due to D.C. shifts in the signal amplitude. Cavity modes are to be avoided by keeping $\frac{2\pi C}{\omega}$ small compared to the chamber dimensions. Steel wool absorbers placed around the microwave horns will help to reduce this problem.

For magnetoplasmas the plane of polarization is parallel to the magnetic field and the direction of propagation is perpendicular to the field.

Plasma Wave Diagnostics:

Mapping of plasma wave dispersion can give information on the plasma parameters and has the advantage of operating even in very tenuous plasmas. Tracing out the ion acoustic dispersion relation, for instance, allows one to determine the plasma density (via ω_{pi}) and obtain information about the plasma temperature through

$$\frac{\omega}{k} = \sqrt{\frac{\gamma_e KT_e + \gamma_i KT_i}{M_i}} \qquad (14)$$

where γ_e and γ_i are the ratio of specific heats for ions (commonly $\gamma_i \approx 3$) and electrons ($\gamma_e \approx 1$).

ω is obtainable directly by spectrum analyzer while k is most often obtained by 2 probe interferometry. Here a plasma wave is launched from one probe and detected by a second, moving, probe. The transmitted signal is added to a reference signal (of comparable amplitude) and k is computed from the shift in probe spacing needed to make the sum of the two signals go through successive maxima or minima.

The cost and complexity of plasma wave diagnostics is most often justifiable when the research group involved already has an ongoing interest in plasma wave phenomena.

Plasma Spectroscopy:

Monochromators and spectrographs can be used to examine the light emitted by a plasma and, in conjunction with various theoretical assumptions, provide plasma density and temperature information.[9]

Most methods for the spectroscopic determination of plasma temperature require only relative measurements of light intensity. Density estimates, on the other hand, can be obtained either from Stark profiles or from absolutely calibrated intensity measurements.

These methods are generally restricted to high plasma densities in order to obtain ample light intensity or to fulfil certain theoretical assumptions needed for data reduction. (High temperature plasmas can also present difficulties but this is probably of less concern to us here.) The use of optical diagnostics is then usually limited to pulsed plasma experiments or the like.

Typical steady state bench top plasma devices may provide ample radiation for (broad spectrum) observation by simple photo diodes, phototransistors, or photo cells. Over a narrow range of plasma parameters (and neglecting the possible stray light attributable to hot filaments, reflection, etc.) the output of these devices can be calibrated against Langmuir probes to give plasma density estimates. Such techniques are useful in that the resulting optical (luminosity) density diagnostic is

non-invasive and will not disturb a plasma as a Langmuir probe measurement would. The time variation of such photodetectors has also been used in the study of plasma waves in the Elmo Bumpy Torus.

References

1. Electric Probes In Stationary and Flowing Plasmas
Chung, Talbot, and Touryan Eds, Springer, 1975.

2. F. Johnson and L. Malter, Phys. Rev. 80, 59 (1950)
and S. Ariga, J. Phys. Soc. Japan, 31, 1221 (1971).

3. S.Chen and T. Sekiguchi, J. Appl. Phys., 36, 2363 (1965).

4. C.R. Hoffman, Plasma Physics, 13, 689 (1971)

5. S. Stephanakis and W. Bennett, Rev. Sci. Instrum. 39,
1714 (1968).
R. Jones, Rev. Sci. Instrum, 49, 21 (1978), 50, 392 (1979)
and R.L. Stenzel, et al, Rev. Sci. Instrum, 53, 1027 (1982)

6. J.E. Osher, Rev. Sci. Instrum, 53, 1685 (1982).

7. R.H. Lovberg in Plasma Diagnostic Techniques, Huddlestone
and Leonard Eds, Academic Press, 1965.

8. C. Wharton in Plasma Diagnostic Techniques, Huddlestone
and Leonard Eds, Academic Press, 1965.

9. H.R. Griem, Plasma Spectroscopy, McGraw-Hill, 1964.

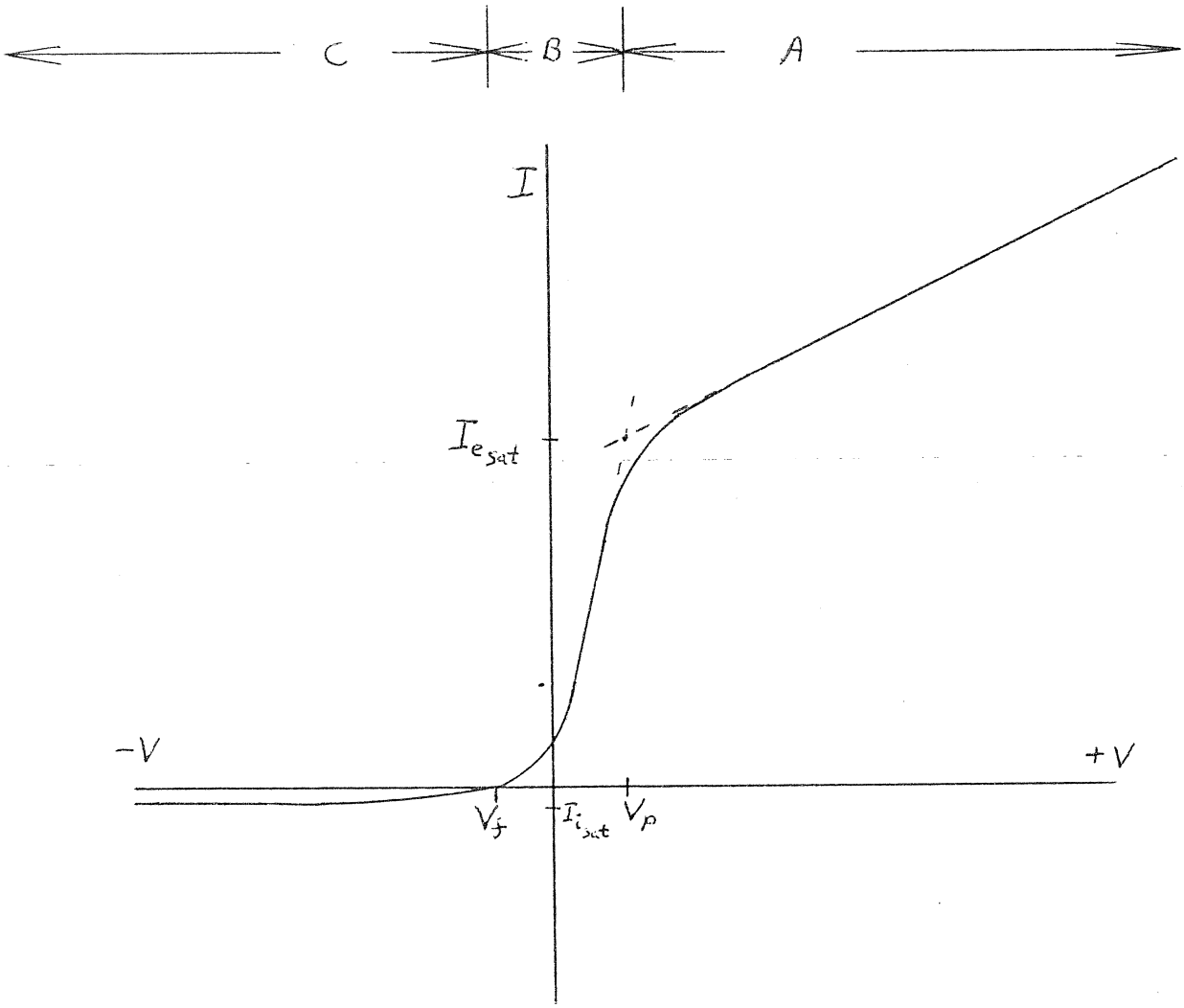
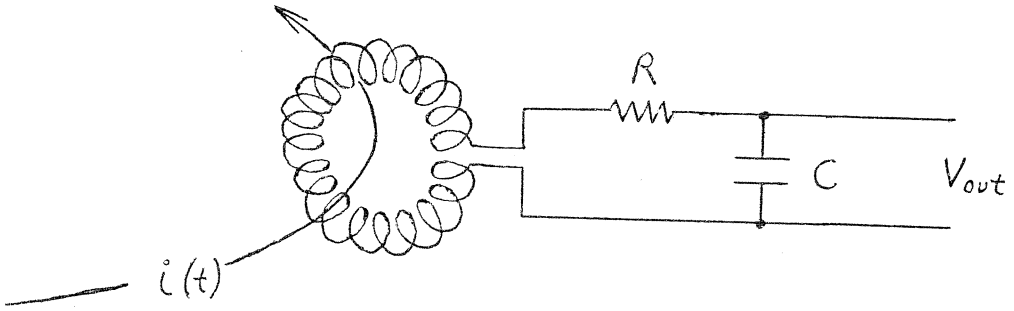
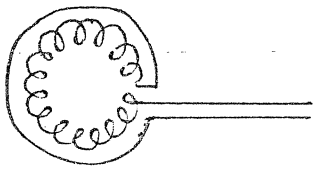


Figure 1

3a



3b



3c

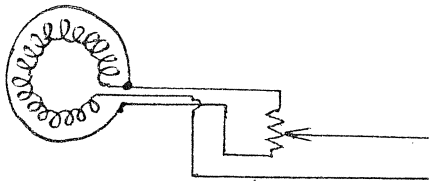


figure 3

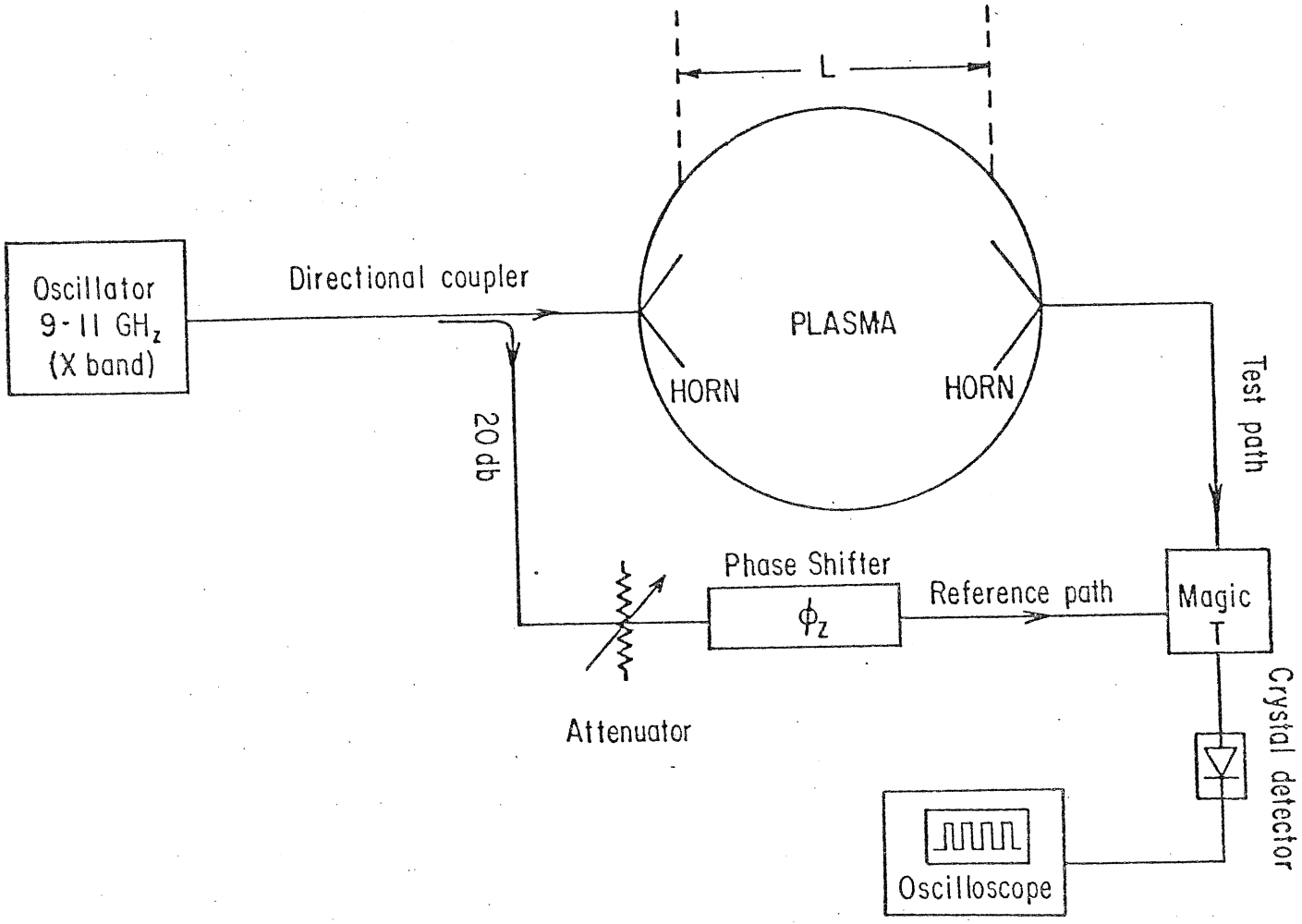

L
Oscillator
9-11 GHz
(X band)
Directional coupler
PLASMA
HORN
HORN
Test path
20 db
Phase Shifter
ϕ_z
Reference path
Magic
T
Attenuator
Crystal detector
Oscilloscope

Figure 4

Turbulent Thermal Insulation With Clump Regeneration And Wall Confinement

R. Jones

DuPree's theory of clump regeneration suggests a considerable reduction in the parasitic electron overheating that will be associated with turbulent thermal insulation.

The usual inertial confinement Scaling for open magnetic systems:

$$\tau_E \simeq \tau_P \simeq \frac{V}{AC_s} \equiv \frac{L}{2C_s} \qquad (1)$$

where V is the plasma volume, A is the loss area, and C_s is the ion sound speed, can be combined with the Lawson criteria:

$$T \simeq 10^4 eV \qquad (2)$$

$$n\tau_E \simeq 10^{14} cm^{-3} S$$

in a $\beta \leq 1$ device:

$$\beta \equiv \frac{8\pi nT}{B^2} \qquad (3)$$

to predict the length of a breakeven experiment[1]. With B limited by present day superconducting magnet technology[2].

$$B \leq 2 \times 10^5 G$$

such a Lawson device would have a length $L \geq 1$ Km.

This excessive length is readily improved by the addition of cold material end plugs[3]. In the presence of the resulting axial pressure balance electron thermal conduction losses dominate[4].

$$\tau_E \simeq \left(\frac{L}{V_e}\right)^2 \upsilon \qquad (4)$$

where V_e is the electron thermal speed and υ is the electron collision frequency(measured parallel to the confining magnetic field). For typical breakeven parameters,

$B = 2 \times 10^5 G$, $n \simeq 10^{17} cm^{-3}$, $\tau_E \simeq 10^{-3} S$, $T \simeq 10^4 eV$, and classical electron collisions, $\upsilon \propto \frac{n}{T^{\frac{3}{2}}}$:

$$L \simeq \tfrac{1}{2} \text{ Km},$$

only a factor of 2 improvement. A practical device would then require the turbulent enhancement[4,5] of υ.

Turbulent thermal insulation would, however, result in a concurrent turbulent plasma heating which must not become excessive[6]:

$$P \leq \frac{nT}{\tau_E} \qquad (5)$$

where P is the turbulent heating power density. If $P = W\upsilon$ where W is the turbulent wave energy density and υ is approximated by[7]:

$$\upsilon \simeq \omega_{Pe} \frac{W}{nT} \qquad (6)$$

the breakeven length becomes[8]:

$$L \simeq 200 \text{ meters}$$

and the maximum allowable turbulence level is about:

$$\frac{W}{nT} \simeq 10^{-5}$$

$$\frac{W}{nT} \simeq \left(\frac{dn}{n} \right)^2$$

$$\frac{dn}{n} \simeq \frac{e\phi}{KT}$$

In DuPree's theory of turbulence[9,10], however, clumps are regenerated and the decay of injected turbulence is

extended to ~5 times the effective particle collision time[11]. This reduced turbulent heating rate has been observed in computer simulation[11,12,13] and would allow a further reduction in reactor length.

Equating the turbulent heating rate to the electron energy loss rate:

$$\frac{nT}{\tau_E} = \frac{W\nu}{5} \tag{7}$$

Solving equations 4, 6 and 7 with $n \approx 10^{17}cm^{-3}$, $T \cong 10^4$ eV, $\tau \simeq 10^{-3}s$: $L \gtrsim 100$ meters.

It is also useful to know how $n\tau$ scales with reactor size. Combining equations 4, 6, and 7 with the condition $\beta \equiv 1$ we obtain:

$$\tau \propto \left(\frac{L}{B}\right)^{\frac{4}{3}} n^{\frac{5}{6}} \tag{8}$$

If T is held fixed then for $\beta \equiv 1$ we have $n \propto B^2$ so:

$$\tau \propto L^{\frac{4}{3}} B^{\frac{1}{3}} \tag{9}$$

or:

$$n\tau \propto L^{\frac{4}{3}} B^{\frac{7}{3}} \tag{10}$$

B is, of course, fixed by engineering constraints leaving $n\tau$ determined by $L^{4/3}$.

Further length reduction can be obtained by going over to wall confined plasmas[14], $\beta > 1$. For a plasma with $n \approx 10^{19}cm^{-3}$, $B = 2 \times 10^5$G, $T \approx 10^4$eV, $\tau_E \approx 10^{-5}$S, and $\frac{W}{nT} \cong 5 \times 10^{-5}$, Solving equations 4 and 6 gives:

$L \gtrsim 1$ meter

when equation 7 is just satisfied.

With $\beta > 1$ radial heat flow is governed by the equation[15].

$$\tau_E = \frac{3nr^2}{2K} \qquad (11)$$

where K, the thermal conductivity, may be turbulently enhanced by a factor of ~ 20 since

$$\frac{\nu}{\nu_{classical}} \cong \frac{\omega_{pe}\frac{w}{nT}}{nT^{-3/2}} \approx 20$$

Solving 11 gives a reactor plasma radius:

$$r \gtrsim 5 \text{ cm}$$

and a plasma energy $nT\pi r^2 L$ of $\gtrsim$ 300 MJ. Such a compact breakeven device would be more than competitive with any of the main line reactor concepts currently under study.

References:

1. Dawson, J.M., Hertzberg, A., Kidder, R.E., Vlases, G.C., Ahlstrom, H.G., and Steinhauer, L.C., Proc. of the V Inter. Conf. on Plasma Phys. and Controlled Fusion Research, Tokyo (1975)

2. Intermagnetics General Corp, Guilderland, New York, U.S.A. currently manufactures superconducting magnets with $B \geq 1.8 \times 10^5$ Gauss. In laboratory magnets critical fields well in excess of 4×10^5 Gauss have been measured.

3. Malone, R.C. and Morse, R.L. (1977) Phys. Rev. Letts., 39, 134

4. Jones, R. (1977) Nuovo Cim., 40B, 182

5. Jones, R. (1980) Plasma Phys., 22, 753

6. Boozer, A. (1978) Comments Plasma Phys. Cont. Fus., 4, 31

7. Sagdeev, R.Z. and Galeev, A.A. (1969) Nonlinear Plasma Theory, Benjamin, New York

8. Jones, R. (1982) Plasma Phys., 24, 445

9. DuPree, T.H. (1966) Phys. Fluids, 9, 1773

10. DuPree, T.H. (1972) Phys. Fluids, 15, 334

11. DuPree, T.H., Wagner, C.E., and Manheimer, W.M., (1975) Phys. Fluids, 18, 1167

12. Hershcovitch, A. and Politzer, P.A. (1979) Phys. Fluids, 22, 1497

13. Hershcovitch, A. and Politzer, P.A. (1982) Phys. Fluids, 25, 522

14. Sakharov, A.D., (1958) Plasma Physics and Controlled Nuclear Fusion Research, Vol. I.

15. Waltz, R.E., (1978) Nuclear Fusion, 18, 901.

****