

XIV Conferencia Latinoamericana de Informática  
17avas. Jornadas Argentinas de Informática  
e Investigación Operativa  
Buenos Aires, Setiembre 1988.

## APREND - UM SISTEMA PARA AQUISIÇÃO AUTOMÁTICA DE CONHECIMENTO A PARTIR DE EXEMPLOS.

FERNANDO A. de Carvalho GOMES  
Universidade Estadual do Ceará - UECE

GIUSEPPE MONGIOVI  
Universidade Federal da Paraíba - UFPB

HELIO de Menezes SILVA  
Universidade Federal da Paraíba - UFPB

Grupo de Inteligência Artificial  
Departamento de Sistemas e Computação  
Universidade Federal da Paraíba - UFPB  
Caixa Postal 10071  
Campina Grande - Pb  
56100  
Brasil

Telefone : (083) 321.7222 r.510  
Telex : 0832211

Sumário - Este trabalho descreve o APREND, um sistema para aquisição automática de conhecimento a partir de exemplos. O método utilizado associa dois processos: o ID3 e o D-S. O ID3 é usado para gerar uma árvore de decisão que é a representação de todo o conhecimento adquirido e que pode ser transformado para a forma de regras. O D-S é um algoritmo baseado na teoria das evidências de Dempster-Shafer e é usado para a determinação dos fatores de certeza das regras geradas pelo ID3.

## 1.0 INTRODUÇÃO

A aquisição de conhecimento é o processo de transferência de um certo domínio do conhecimento de um perito nesse domínio para o sistema, envolvendo também a transformação deste conhecimento numa forma que possa ser entendida.

Embora a tecnologia de desenvolvimento de sistemas especialistas tenha experimentado avanços consideráveis, a aquisição de conhecimento continua sendo o gargalo ("bottleneck") na construção de um sistema especialista (Feigenbaum 1981 e Buchanan 1985), isto deve-se principalmente à problemas de comunicação entre o engenheiro do conhecimento e o perito no domínio. Estes problemas de comunicação advêm da linguagem específica que cada domínio do conhecimento possui e, de acordo com (Waterman 1986): "o perito combina as partes do conhecimento básico de uma forma tão rápida, que torna-se difícil para ele descrever o processo de raciocínio inerente a uma conclusão".

Existem dois métodos de aquisição de conhecimento: cognitivo e automatizado (Boy 1988). Os métodos cognitivos são fundamentados essencialmente no engenheiro do conhecimento, que funciona como uma espécie de "filtro" entre a informação fornecida pelo perito e a representação do conhecimento na máquina. Os métodos automatizados são aqueles que são processados com o auxílio do computador e, portanto, sem a influência direta do engenheiro do conhecimento. Entre os métodos de aquisição automática de conhecimento destacam-se: indução por meio de árvores de decisão, entrevista automática com o perito e métodos gráficos de apresentação das estruturas do conhecimento.

A conscientização de que o cruciante problema da

aquisição de conhecimento deva ter soluções "rápidas e impessoais", veio estimular a investigação de métodos automatizados que aumentem a eficiência e confiabilidade do processo.

O processo de indução através de árvores de decisão, é um método simples, que não necessita de nenhum conhecimento prévio do domínio. Este método funciona a partir de exemplos, que são descritos por atributos e diagnósticos ou classes e registrados no cotidiano do perito no domínio. A indução se utiliza de teorias probabilísticas e como resultado fornece o conhecimento representado na forma de árvore de decisão, de onde pode-se obter regras do tipo SE...ENTÃO. A estas regras associa-se um fator de certeza ou fator de atenuação (FA), que representa a característica de incerteza do conhecimento sendo adquirido. Por exemplo, um FA de 0,8 indica que existe uma forte evidência (0,8 em 1,0) de que a regra se verifica.

O objetivo deste trabalho é descrever um sistema de aquisição automática de conhecimento, baseado na indução em árvores de decisão (Quinlan 1986a) e em um processo estatístico para o cálculo do FA das regras geradas. Esta associação fornece regras que constituem um conhecimento confiável a ser aplicado em sistemas especialistas baseados em regras.

Na seção 2.0 será feita uma descrição do método de aquisição automática de conhecimento utilizado pelo sistema. Na seção 3.0 será mostrado uma visão geral do sistema APREND e um exemplo completo da obtenção de regras para sistemas especialistas.

## 2.0 O METODO UTILIZADO

O método proposto para obtenção de regras a partir de exemplos é uma combinação do algoritmo ID3 (Quinlan 1986a) e um processo para o cálculo dos FA's das regras geradas, baseado na teoria das evidências de Dempster-Shafer (Shafer 1976).

O ID3 gera regras de modo eficiente, através de indução em árvores de decisão, porém não calcula os FA's das regras geradas. A determinação informal dos FA's é um processo subjetivo e normalmente leva a resultados discrepantes. Portanto se faz necessário um processo formal de geração dos FA's. O processo de Dempster-Shafer (D-S) poderia ser usado para gerar as regras com os seus respectivos FA's, porém com uma complexidade explosiva e com alguns inconvenientes na fase de formação das regras. Então a solução encontrada foi combinar as vantagens dos dois processos: o ID3 para gerar as regras e o D-S apenas para calcular os FA's das regras já formadas, porém com uma complexidade muito reduzida.

### 2.1. Indução em Árvores de Decisão:

O algoritmo ID3 é um membro da família TIDT ("Top-Down Induction of Decision Trees") e é originário do CLS, um algoritmo desenvolvido por Hunt nos anos cinquenta e apresentado em (Hunt 1966).

O ID3 é o algoritmo que mais tem apresentado resultados práticos, por exemplo, o sistema de diagnóstico de falhas de placas de circuitos impressos da ITT Europa, o sistema para separação gas-óleo da British Petroleum e o controle de processos para usinas nucleares da Westinghouse (Carter 1987). A grande vantagem do ID3 com relação a outros métodos é a escolha do

universo de exemplos que serão trabalhados. Nesse aspecto, o ID3 utiliza apenas alguns exemplos para chegar a uma árvore que classifique todos os exemplos observados. Este pequeno conjunto de exemplos chama-se janela.

A tarefa do ID3 é formar uma árvore de decisão que classifique corretamente todos os exemplos observados.

No primeiro passo, o algoritmo irá fornecer uma árvore de decisão que, não só classifica corretamente todos os elementos do conjunto janela, como também grande parte dos exemplos observados. Em seguida irá pesquisar na totalidade dos exemplos, todos aqueles que não são classificados pela árvore fornecida. Então, uma seleção destes elementos será realizada de forma a inclui-los no conjunto janela. A partir da nova janela, o algoritmo formará uma nova árvore de decisão e o processo de classificação se repetirá até que todos os exemplos observados sejam classificados. Ao final do processo, cada caminho da árvore de decisão resultante, representa uma regra. O conjunto destas regras constitui o conhecimento correspondente aos exemplos observados.

## 2.2 O Processo D-S:

A característica principal do D-S é a de considerar a totalidade dos exemplos observados não como um universo absoluto, mas como parte de um universo maior. Por exemplo, cem casos anotados por um médico especialista, não podem ser considerados como o universo de todos os casos possíveis. Já outros métodos estatísticos consideram os cem exemplos como um universo absoluto. Esta característica do D-S produz FA's com valores que estão mais próximos da realidade.

O processo D-S apresenta uma desvantagem. Para se calcular o FA de um atributo com relação a uma classe, necessita-se percorrer a totalidade dos exemplos anotados. Entretanto esta ineficiência praticamente desaparece quando as regras já estão formadas. É o caso de o processo ser usado apenas para o cálculo dos FA's.

### **3.0 DESCRIÇÃO DO SISTEMA APREND**

#### **3.1 Visão Geral**

O sistema é dividido em três fases distintas: a edição da tabela de exemplos, a geração de regras e a determinação dos FA's.

A fig. 3.1 mostra um gráfico de estrutura correspondente ao problema a ser resolvido, apresentando dois níveis de detalhamento.

O módulo "EDITAR EXEMPLOS" edita os elementos da tabela de exemplos, que dizem respeito ao domínio do conhecimento que se quer implantar no sistema especialista. O objetivo principal deste módulo é obter exemplos que se passaram no cotidiano do perito no domínio. O módulo "GERAR REGRAS" corresponde ao algoritmo ID3. Ele chama o módulo "INCL. JANELA" nas fases de formação da janela e inclusão de elementos não mapeados.

O módulo "DETERMIN. FA", corresponde ao D-S.

Tanto a nível conceitual, como a nível de implementação, foram usados alguns recursos importantes para diminuir a complexidade dos algoritmos envolvidos.

#### **3.2 Adquirindo Conhecimento**

O conhecimento é adquirido a partir de um conjunto de

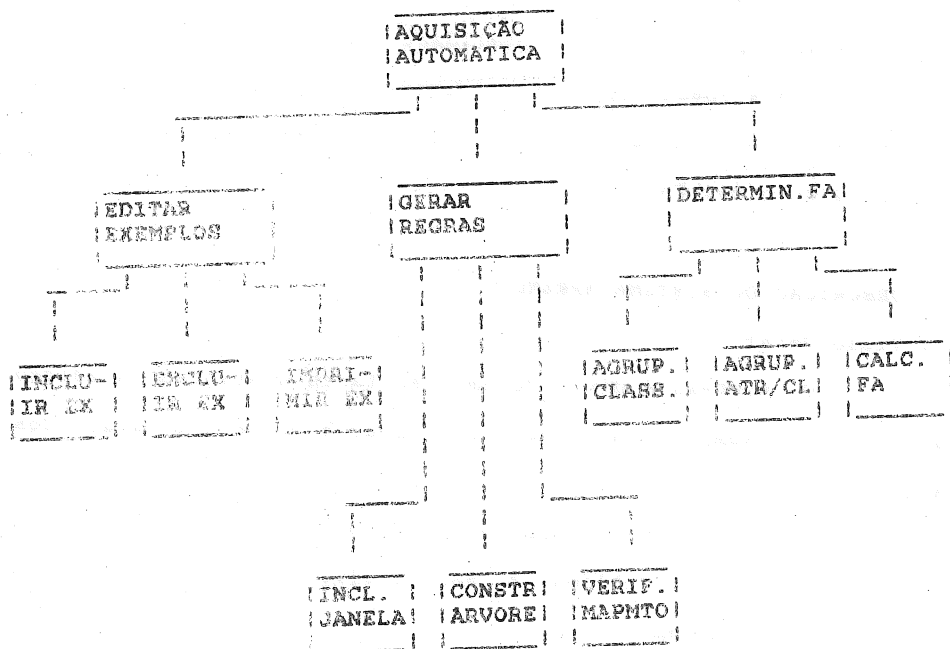


Figura 3.1. Visão Geral do Sistema

casos. Cada caso, também chamado de objeto ou exemplo, é descrito como uma coleção de atributos. Cada atributo indica alguma característica importante de um objeto e será limitado a um conjunto de valores discretos, mutuamente exclusivos. A classe, é o diagnóstico do caso.

A seguir descreve-se informalmente o sistema APREND, através da apresentação de um exemplo ilustrativo.

O exemplo ilustrativo consiste na observação por vários anos, das condições meteorológicas das manhãs de sábado, durante um longo período de observações.

Consideram-se os seguintes atributos, adaptados de (Quinlan 1986a):

**ATRIBUTO****VALORES**

Tempo.....[ensolarado, nublado, chuvoso]  
Temperatura.....[frio, agradável, quente]  
Umidade.....[alta, normal]  
Vento.....[forte, fraco]

Quando juntos, pode-se caracterizar muito bem as condições meteorológicas de uma manhã de sábado, por exemplo:

Tempo : ensolarado;  
Temperatura : agradável;  
Umidade : normal;  
Vento : fraco.

Para formar um objeto, resta definir a classe deste objeto, que possui um conjunto de valores observados de acordo com o domínio do conhecimento que se quer adquirir. Esta observação poderia ser, no caso dos atributos já definidos, a decisão de ficar em casa ou sair para praticar um esporte. Assim sendo, a classe terá os valores [Casa, Futebol, Basquete].

Objetos que parecem ser atípicos são particularmente interessantes, pois indicam uma situação que, mesmo difícil de acontecer, vem influenciar na construção da árvore de decisão, como veremos adiante.

**Utilização de ID3**

Para a escolha dos elementos do conjunto janela, leva-se em conta a proporção da ocorrência de cada classe em todos os exemplos observados.

Suponha que foram observados um total de 1.000 casos. Para o conjunto janela, inicialmente, foram utilizados 11

elementos, conforme a fig. 3.2.

| No. | Tempo      | Temperatura | Umidade | Vento | Classe   |
|-----|------------|-------------|---------|-------|----------|
| 1   | ensolarado | agradável   | normal  | forte | basquete |
| 2   | nublado    | frio        | alta    | fraco | casa     |
| 3   | nublado    | agradável   | alta    | fraco | basquete |
| 4   | chuvoso    | frio        | alta    | forte | casa     |
| 5   | nublado    | quente      | normal  | fraco | futebol  |
| 6   | ensolarado | quente      | normal  | fraco | futebol  |
| 7   | nublado    | agradável   | alta    | forte | basquete |
| 8   | chuvoso    | frio        | alta    | fraco | casa     |
| 9   | nublado    | agradável   | normal  | forte | futebol  |
| 10  | ensolarado | quente      | alta    | fraco | casa     |
| 11  | chuvoso    | quente      | alta    | forte | basquete |

Figura 3.2 Conjunto Janela

A raiz da árvore de decisão corresponde ao atributo de menor entropia, isto é, o que apresenta menor incerteza com relação às classes. O conceito de entropia é originário da teoria da informação e sua fundamentação pode ser encontrada em (Aczél 1978).

A entropia de um atributo A é a média ponderada das entropias dos valores de A. Matematicamente é dada por:

$$H(C|A) = \sum_{i=1}^N p(v_i) \cdot H(C|v_i) \quad (3.1)$$

onde:

N - número total de valores do atributo A;

$p(v_i)$  - é a probabilidade de o valor  $v_i$  aparecer no conjunto janela;

$H(C|v_i)$  - é a entropia do valor  $v_i$  de A, representada pela expressão:

$$H(C|v_i) = - \sum_{j=1}^M p(c_j|v_i) \cdot \log p(c_j|v_i), \quad (3.2)$$

onde:

M - número total de classes.

$p(c_j | v_i)$  - é a probabilidade de se ter um exemplo com o valor  $v_i$  concluindo  $c_j$ .

No conjunto janela da fig. 3.2, calcula-se a entropia de cada atributo. Por exemplo, a entropia do atributo Tempo, é obtida como segue:

$$H(\text{Classe}=\text{casa} | \text{Tempo}=\text{nublado}) = - \frac{1}{5} \log \frac{1}{5} = 0,46,$$

$$H(\text{Classe}=\text{basquete} | \text{Tempo}=\text{nublado}) = - \frac{2}{5} \log \frac{2}{5} = 0,53,$$

$$H(\text{Classe}=\text{futebol} | \text{Tempo}=\text{nublado}) = - \frac{2}{5} \log \frac{2}{5} = 0,53.$$

então, pela expressão (3.2) tem-se:

$$H(C | \text{Tempo}=\text{nublado}) = 0,46 + 0,53 + 0,53 = 1,52.$$

Calculando-se a entropia para os outros valores do atributo Tempo, encontra-se:

$$H(C | \text{Tempo}=\text{ensolarado}) = 1,59 \text{ e } H(C | \text{Tempo}=\text{chuvoso}) = 0,92.$$

Desta forma, pela expressão (3.1), a entropia do atributo Tempo será:

$$H(C | \text{Tempo}) = \frac{5}{11} 1,52 + \frac{3}{11} 1,59 + \frac{3}{11} 0,92 = 1,37.$$

Calculando-se a entropia dos outros atributos, tem-se:

$$H(C | \text{Temperatura}) = 0,85,$$

$$H(C | \text{Umidade}) = 0,92,$$

$$H(C | \text{Vento}) = 1,41.$$

O atributo de menor entropia é Temperatura, que será, conforme a fig. 3.3, a raiz da árvore de decisão.

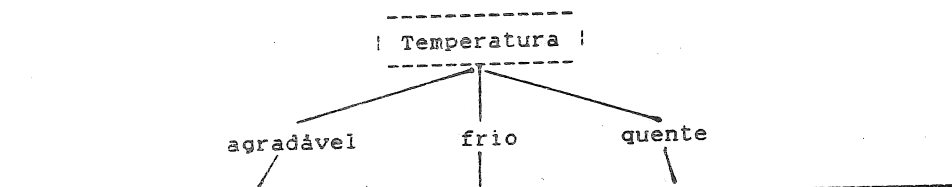


Figura 3.3 Raiz da árvore de decisão

Para se obter as sub-árvores correspondentes aos valores do atributo raiz, procede-se de forma análoga à anterior. Nesse caso, usa-se uma sub-janela que é formada apenas pelos elementos da janela anterior que contêm o valor do atributo raiz correspondente à sub-árvore específica.

Por exemplo, para o valor agradável da fig. 3.3, a sub-janela é a da fig. 3.4.

| No. | Tempo      | Temperatura | Umidade | Vento | Classe   |
|-----|------------|-------------|---------|-------|----------|
| 1   | ensolarado | agradável   | normal  | forte | basquete |
| 3   | nublado    | agradável   | alta    | fraco | basquete |
| 7   | nublado    | agradável   | alta    | forte | basquete |
| 9   | nublado    | agradável   | normal  | forte | futebol  |

Figura 3.4 Sub-janela para Tempo=agradável

A fig. 3.5 mostra a árvore de decisão que classifica corretamente cada elemento do conjunto janela da fig. 3.1. Supondo que essa árvore é a árvore de decisão final, cada caminho dessa árvore (raiz-folha) representa uma regra. Por exemplo, o caminho mais a esquerda, na árvore da fig. 3.5, fornece a regra:

SE Temperatura é agradável &  
 Umidade é alta  
 ENTÃO Jogue Basquete

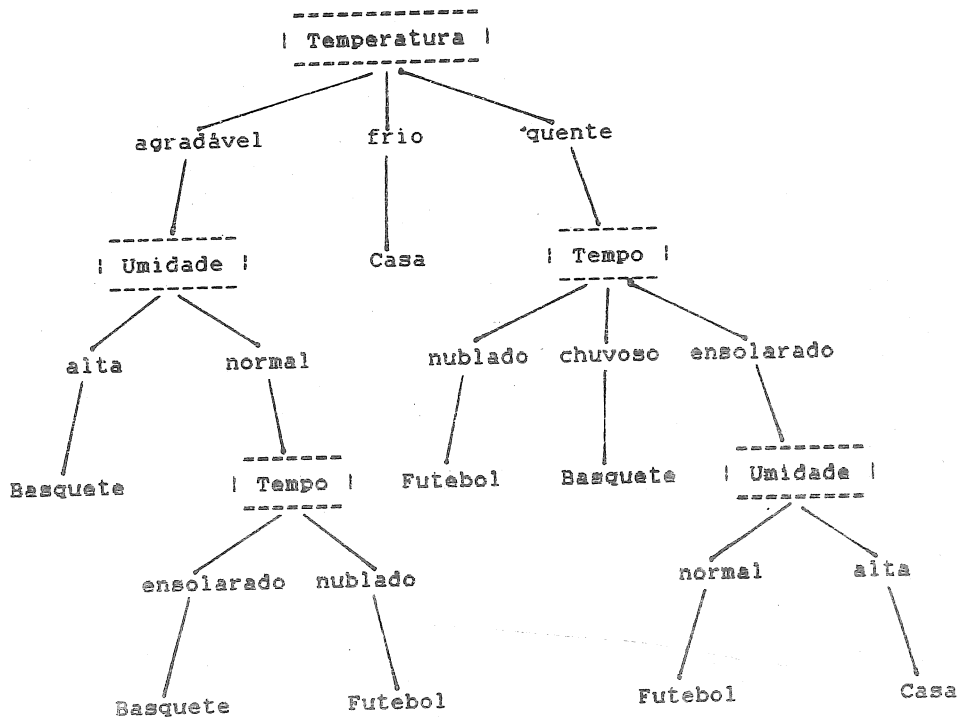


Figura 3.5 Árvore de Decisão

Se dois objetos no conjunto janela possuem os mesmos valores para todos os atributos, mas com valores diferentes para a classe, deve-se rotular o nó relativo a estes dois exemplos como Nulo, isto constitui o que se chama de "ruído" e um processo de eliminação do ruído é proposto em (Quinlan 1986b).

#### Utilização do D-S

De posse das regras geradas pelo ID3, calcula-se o FA de cada regra, através do D-S.

Inicialmente deve-se conhecer alguns conceitos da Teoria de Dempster-Shafer. A evidência que uma condição da regra (atributo = valor), contribui para que a conclusão (classe)

ocorra, chama-se de "medida de crença do valor do atributo com relação a uma classe", ou  $C(v,c)$ , onde "v" é o valor do atributo e "c" é o valor da classe. A "medida de crença do valor do atributo com relação a uma classe" permite calcular o FA deste atributo com relação a uma regra que conclua a mesma classe. Além da "medida de crença", necessita-se calcular também a "medida de crença-contrária", ou  $C(v,-c)$ , que é a evidência de que a condição da regra contribui para que não ocorra a conclusão. Estes conceitos podem ser vistos em (Shafer 1976).

A expressão que fornece o fator de certeza (FA) de um atributo com relação a uma classe, é:

$$FA(v,c) = \frac{C(v,c) \cdot (1 - C(v,-c))}{1 - C(v,c) \cdot C(v,-c)}$$

Para se encontrar a medida de crença de um atributo com relação a uma classe utiliza-se a expressão:

$$C(v,c) = p_1 \cdot (1 - P)$$

onde:

$p_1$  - é a probabilidade de ocorrer o valor v e a classe c, ao mesmo tempo, no conjunto dos exemplos;

P - é a área sob a curva de Gauss que dá a probabilidade de que a proporção de ocorrer a classe c, não mudará, ao separarmos todos os exemplos que possuem o valor do atributo escolhido.

Para se encontrar o FA final de uma regra, basta combinar, dois a dois, os fatores de certeza dos atributos que formam esta regra, usando a seguinte expressão:

$$FA = FA_1 + FA_2 - FA_1 \cdot FA_2.$$

Para exemplificar a utilização do método, calcular-se-á, com dados simulados do exemplo anterior, o FA para a regra:

SE Temperatura é agradável &  
Umidade é alta  
ENTÃO Jogue Basquete.

A "medida de crença" em que Temperatura = agradável influencia Classe = basquete (utiliza-se agora a totalidade dos exemplos, ou seja: 1.000), é:

$C(\text{Temperatura=agradável, basquete}) = 0,78 \cdot (1 - 0,38) = 0,48$ ,  
a seguir calcula-se a "medida de crença-contrária" em que Temperatura ≠ agradável influencia Classe=basquete, acha-se:

$C(\text{Temperatura=agradável, -basquete}) = 0,22 \cdot (1 - 0,38) = 0,14$ ,  
com estes valores, tem-se:

$$FA(\text{Temperatura=agradável, basquete}) = \frac{0,48 \cdot (1 - 0,14)}{1 - 0,48 \cdot 0,14} = 0,44.$$

Agora encontra-se o FA do outro atributo da regra (Umidade = alta), que é:

$$FA(\text{Umidade=alta, basquete}) = 0,53.$$

Combina-se os dois FA's. Então o FA, resultará:

$$FA = 0,44 + 0,53 - 0,44 \cdot 0,53 = 0,74,$$

portanto a regra será:

SE Temperatura é agradável &  
Umidade é alta  
ENTÃO Jogue Basquete  
FA = 0,74

esta é uma regra que é aplicável em sistemas especialistas.

#### 4.0 Conclusões

Neste trabalho foi abordado o problema geral da aquisição de conhecimento. Em particular, salientando-se a importância dos métodos de indução para gerar árvores de decisão a partir de exemplos. Foi mostrado que, entre os métodos de indução, o algoritmo ID3 é o que tem apresentado melhores resultados práticos. Para lidar com a natureza incerta do conhecimento, as regras geradas pelo ID3 foram complementadas com fatores de certeza, calculados por meio de um algoritmo baseado na teoria das evidências de Dempster-Shafer.

As idéias apresentadas foram utilizadas no projeto do sistema APREND. Para melhor ilustrar essas idéias, o sistema foi descrito informalmente através do acompanhamento de uma exemplo ilustrativo.

#### Referências

- Aczél, J & Daróczy, Z, On Measures of Information and Their Characterizations, Academic Press, New York, 1975.
- Boy, G., Faller, B. & Sallantin, J., Acquisition et Ratification de Connaissances; Crim, Montpellier, 1988.
- Buchanan, B.G. ,Lenat, E. & Waterman, D.A., Building Expert Systems, Addison Wesley, Reading, Mass., 1985.
- Carter, C. & Catlett J., Assessing Credit Card Applications Using Machine Learning, in IEEE Expert (Fall), 1987.
- Feigenbaum, E.A., Expert Systems in 1980s, in State of the Art Report on Machine Intelligence, A. Bond, Ed., Maidenhead: Pergamon-Infotech, 1981.
- Hunt, E.B., Marin, J. & Stone, P., Experiments in Induction,

Academic Press, New York, 1966.

- Michalski, R.S. & Chilauski, R.L., Knowledge Acquisition by encoding expert rules versus computer induction from examples: a case study involving soybean pathology, in Fuzzy Reasoning and its Applications, Gaines, B.R., London, 1981.
- Quinlan, J.R., Induction of Decision Trees, In Machine Learning, Kluwer Academic Publishers, Boston, MA, 1986a.
- Quinlan, J.R., The Effect of Noise on Concept Learning, in Machine Learning: An Artificial Intelligence Approach, V.2, Michalski, R et al. (Eds.), Morgan Kaufmann, Palo Alto, CA, 1986b.
- Shafer, G., A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
- Waterman, D.A., A Guide to Expert Systems, Addison Wesley, 1986.