SURVEY SOBRE TÉCNICAS DE TOMADA DE DECISÃO

Thiago Cristino dos Santos

UNICAMP

ABSTRACT

A ciência cognitiva é uma área de conhecimento que se dedica à muito tempo em questões referentes a processos relacionados com o conhecimento, como aprendizagem, classificação e aquisição de conhecimento.

Estas questões são extremamente relevantes para áreas que envolvem sistemas para inteligência de negócios, em especial os sistemas de tomada de decisão.

O objetivo deste *survey* é apresentar as principais técnicas de tomada de decisão atualmente em uso e as suas respectivas vantagens e desvantagens, juntamente com as propostas mais recentes, com fortes ligações com a ciência cognitiva.

INTRODUÇÃO

Sistemas especialistas na área de tomada de decisões possuem um papel crucial em diversas organizações e áreas de negócios.

Basicamente, o processo de decisão compreende a análise de variáveis em um dado contexto e a obtenção de uma informação que sirva de resposta e, auxilie a tomada de decisões, o que seria impossível ou pouco prático se fosse feito apenas por seres humanos, seja devido à complexidade dos cenários, seja devido ao alto volume de dados envolvido.

Estes sistemas são desenvolvidos à partir de diversas técnicas, baseadas em conceitos estatísticos, redes neurais, algoritmos genéticos, técnicas de aprendizagem.

A relevância e aplicabilidade destes sistemas acaba por servir como uma aplicação prática de diversos conceitos e teorias que também são de interesse da ciência cognitiva, e com isso, estimulam o desenvolvimento e refinamento destas técnicas em ambas as áreas.

O objetivo deste survey é apresentar e descrever as principais técnicas atualmente em uso em sistemas no estado da arte que se encontra em uso no mercado, avaliando suas vantagens, desvantagens e aplicabilidade.

A seguir, serão apresentadas e descritas as novas propostas que ainda não são utilizadas no mercado, estudando as suas vantagens, o seu potencial de evolução, limitações e aplicabilidade.

As técnicas atualmente em uso em sistemas de tomada de decisão são :

- Redes Bayesianas
- Séries Temporais
  - Simples
  - Multivalorada
- Árvores de Decisão
- Redes Neurais
- Algoritmos Genéticos

As três primeiras técnicas utilizam fortemente conceitos estatisticos em sua construção, enquanto que as duas últimas técnicas são baseadas em conceitos computacionais, alguns deles já conhecidos na ciência congnitiva como as redes neurais e os algoritmos genéticos.

REDES BAYESIANAS

As redes bayesianas representam o conehcimento incerto e incompleto, através da teoria da Teoria da Probabilidade Bayesiana, definida pelo matemático Thomas Bayes.

Define-se conhecimento incerto como sendo aquele que apresenta deficiências em sua completude, ou seja, em algum momento os dados podem apresentar uma representação inexata, parcial ou aproximada de uma realidade. (BHATNAGAR, 1986).

Uma das principais características e potenciais de desta forma de representação é o enorme potencial de adaptação, sendo possível a partir de um novo conjunto de informações, baseadas em informações de fundo verdadeiras, poder produzir alterações nas dependências e nos conceitos.

De um modo geral, pode-se até afirmar que isto faz com que as probabilidades não sejam apenas acasos e sim eventos confirmados, que dão origem a conceitos novos a respeito do domínio em questão.

Um fato importante que deve ser observado é que o modelo é baseado em probabilidades anteriormente analizadas, enriquecidas com novas informações que possuem um certo grau de incerteza para representar este conhecimento.

Uma rede bayesiana é representada através de um grafo direcionado acíclico, no qual os nós representam variáveis de um domínio, e os arcos são a dependência condicional ou informativa entre as variáveis, sendo usadas probabilidades associadas entre os nós pais e mós filhos na rede para representar a força da dependência.(PEARL, 1988), como na Figura 1.

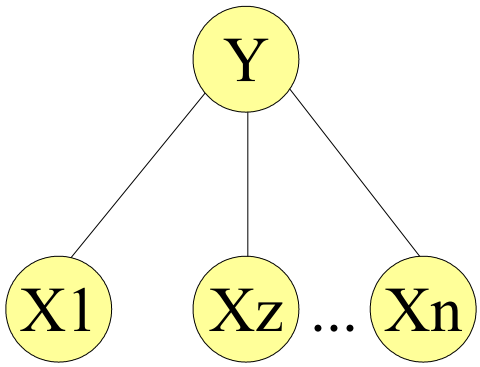

Figura 1 – Exemplo de rede Bayesiana

Uma rede bayesiana é usanda em sistemas de tomada de decisão onde existe uma clara situação de causa conseqüência, passível de ser modelada, como por exemplo nos sistemas de classificação de spams.

Uma aplicação comum das redes bayesianas em sistemas de tomada de decisão é em sistemas de classificação de *spams*, onde as palavras possuem probabilidades particulares de estarem em um e-mail classificado como spam ou como um e-mal válido.

Inicialmente, o filtro não conhece estas probabilidades, e portanto deve ser treinado para que ele possa construí-las. No exemplo do filtro anti-spam, o usuário precisa indicar manualmente quando uma nova mensagem constitui ou não um spam.

Para cada palavra presente em cada mensagem aplicada como treinamento, o filtro irá ajustar automaticamente as probabilidades de cada palavra estar prsente em um e-mail legítimo ou em um e-mail de spam.

Esta possibilidade de treinamento para um domínio em questão mostra a principal vantagem de um sistema de tomada de decisão baseado em redes bayesianas, que é a de poder treinar o sistema em um problema onde pode-se estabelecer uma relação clara entre causa e consequência.

SÉRIES TEMPORAIS

Resumidamente, uma série temporal é um conjunto de observações ordenadas no tempo e que apresentam dependência entre os diversos instantes de tempo.

Pode ser usada quando se possui um histórico de valores de uma ou mais variáveis, que sempre deve ter um período de tempo associado.

Quando esta técnica é aplicada a uma única variável, é denominada de serie temporal simples; quando aplicada a mais de uma variável é denominada de série temporal multivalorada.

Um exemplo de uso de séries temporais em sistemas de tomada de decisão, é a situação onde uma distribuidora de energia elétrica quer decidir quando deve fazer novos investimentos em sua infra-estrutura em função do aumento de demanda por parte de seus consumidores.

Neste cenário, a companhia dimensiona que a sua rede para uma dada subestação deve ter uma demanda de carga de até um percentual do total possível pelos equipamentos para evitar desgaste precoce dos mesmos, tomando a decisão de expandi-los quando a demanda apresentar uma tendência de aumento que atinge ou ultrapassa o percentual estabelecido.

Assim, pode-se aplicar uma série temporal simples, onde a variável é o consumo total que está atrelada à períodos de 30 dias, de acordo com a tabela abaixo :

t	*Consumo Total*
01/2004	320MWh
...	...
12/2004	280MWh
01/2005	367MWh
02/2005	395MWh
03/2005	371MHh

Tabela 1 – Exemplo de dados para série temporal simples

Imaginando um outro cenário, onde além do consumo, a companhia quer saber o quanto deste consumo é decorrente de fraudes (ligações irregulares, medidores adulterados), para que possa tomar a decisão de aumentar a sua infra-estrutura apenas quando o aumento for decorrente dos consumidores em situação legal e ainda, quanto deste consumo está sendo pago, ou seja, qual o grau de adimplência dos consumidores.

Para o aumento de consumo por fraude, a companhia aplicará uma outra estratégia, eliminando estas ligações irregulares ou acionando legalmente os consumidores que adulteraram os medidores.

Caso o percentual de inadimplência for acima do tolerado pela companhia, ela pode tomar outras medidas, como negociação de dívidas ou cobrança legal.

Esta situação é completamente diferente da primeira, pois possui quatro variáveis co-relacionadas ao invés de apenas uma.

t	*Consumo Total*	*Consumo por fraude*	*Consumo Medido*	*Total das Faturas*
01/2004	320,00MWh	47,00MWh	273,00MWh	
02/2004	305,00MWh	47,23MWh	257.77MWh	
...	...	...	...	...
12/2004	280,00MWh	50,52MWh		
01/2005	367,00MWh	51,00MWh		
02/2005	395,00MWh	51,01MWh		
03/2005	371,00MHh	51,50MWh		

Tabela 2 – Exemplo de dados para série temporal multivalorada

Não pode ser usada adequadamente quando o histórico é incompleto, sendo impossível aplicar esta técnica para situações onde não existe este histórico ou o mesmo possui um intervalo de tempo muito curto para avaliação, ou ainda quando problema em questão não pode ser expresso como variáveis em função do tempo.

ÁRVORES DE DECISÃO

As árvores de decisão são compostas de nós que representam os atributos, arcos provenientes destes nós e que podem assumir os valores possíveis para estes atributos, e de nós folha representando as as diferentes classes de um conjunto de treinamento [3] especificado.

Pode-se definir a funcionalidade principal de uma

árvore de decisão é efetuar recursivamente o particionamento de um dado conjunto de treinamento até que cada subconjunto resultante contenha casos de apenas uma classe.

Este objetivo é conseguido através da análise e comparação de distribuição de classes durante a construção; quando completam os dados rpresentes estão organizados completamente, servindo para classificar novos casos [3, 4].

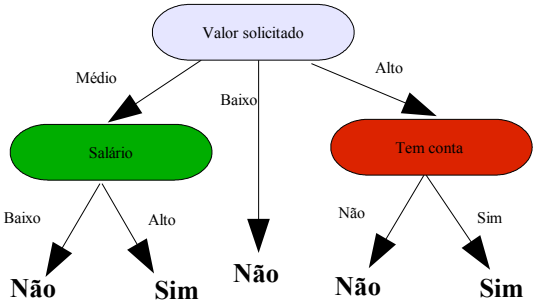

Figura 2 – Exemplo simples de uma árvore de decisão

Na figura 2 é apresentado um exemplo simples de uma árvore de decisão, que no escopo apresentado trata da análise de condições para decidir se um empréstimo deve ou não ser concedido à uma pessoa, classificando as probabilidades do valor solicitado ser baixo, médio ou alto.

Alguns objetos exibidos são exemplos positivos de uma classe "***sim***", que indica que os requisitos do banco para concessão de empréstimo para a pessoa foram atingidos satisfatoriamente.

A classe **"não"** indica que estes requisitos não foram atingidos e portanto o empréstimo não deve ser concedido.

Neste exemplo a classificação é a modelagm de uma estrutura de árvore que pode ser usada para classificar de forma correta todos os objetos do conjunto[4].

Existem diversos algoritmos de classificação que podem ser aplicados nas árvores de decisão, não existindo uma forma de determinar qual é o melhor, ou seja, um dado algoritmo pode ter um melhor desempenho em uma situação e por outro lado, apresentar um desempenho inferior quando aplicado à outras situações.

Um dos primeiros algoritmos de árvores de decisão que foi elaborado é o ID3, que é baseado em sistemas de inferência e aprendizagem.

Hoje em dia existem algoritmos mais conhecidos, como o C4.5, CART (*Classification and Regression Trees*), CHAID (*Chi Square Automatic Interaction Detection*), por exemplo.

Uma vez construída, é importante efetuar a avaliação de uma árvore de decisão através da utilização de dados que não foram aplicados no processo de treinamento.

Isto permite estimar a forma pela qual a árvore aplica a generalização dos dados e como a mesma se adapta em novas situações, possibilitando ainda conseguir uma estimativa da proporção de erros e acertos na construção da árvore. [4].

Consegue-se derivar regras à partir de uma árvore de decisão, considerando o trajeto do nó raíz até uma folha da árvore.

Como uma árvore de decisão tende a ser grande, dependendo das aplicações, elas são substituídas por regras que podem ser mais facilmente modularizadas. Uma regra pode ser entendida sem a necessidade de referenciar outras regras [5].

Considerando a árvore de decisão daa figura 1, são obtidos dois exemplos de regras:

Se "Valor solicitado" **for igual a** "médio" e o "Salário" **for igual a** "baixo", **então** a classe é "não".

Se "Valor solicitado" **for igual a** "médio" e o "Salário" **for igual a** "alto", **então** a classe é "sim".

Árvores de decisão são adequadas em problemas que podem ser decompostos em regras diretas. Alguns cenários de tomada de decisão apresentam, além de uma grande complexidade, muitas variáveis difíceis ou impossíveis de serem identificadas.

REDES NEURAIS

Em 1943 ocorreu o início dos estudos que levaram à criação das redes neurais por McCulloch e Pitts, que proporam uma máquina que fosse construída com inspiração no cérebro humano.

Basicamente uma rede neural é um grupo interconectado de neurônios artificiais, que são estruturas que simulam os neurônios biológicos.

Cada neurônio representanda funções de uma ou mais dimensões que recebem entradas e efetuam uma somatória nestes valores para gerar uma saída, como ilustração na figura 3.

De um modo geral, a somatória de cada nó possui um peso específico, e é transferida através de uma função não linear conhecida como função de ativação ou função de transferência.

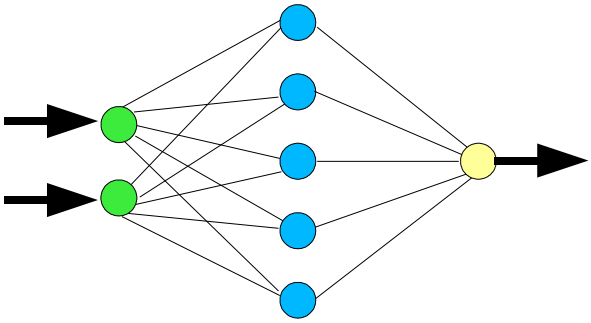

Figura 3 – Exemplo de estrutura de uma rede neural

Uma vez construída, uma rede neural passa por um processo de treinamento, onde a mesma é submetida à estimulos para adquirir e classificar as informações.

Existem basicamente três paradigmas de aprendizado; supervisionado, não supervisionado e aprendizado por reforço, que podem ser aplicados a praticamente qualquer arquitetura de redes neurais.

O paradigma do aprendizado supervisionado, a rede é submetida a um conjunto de dados de exemplo e o objetivo é encontrar uma função na classe de funções permitidas que atenda aos exemplos.

O objetivo do paradigma de aprendizado supervisionado é inferir o mapeamento implicado pelos

dados fornecidos, enquanto que a função de custo se relaciona com a não combinação entre o mapeamento e os dados.

No paradigma do aprendizado não supervisionado, a rede neural recebe os dados, a função de custo e a saída esperada da rede.

O treinamento por reforço é um paradigma de aprendizado onde conjunto de dados não é explicitamente fornecido, e sim produzido através de interações da rede com o seu ambiente.

Em um dado ponto no tempo, o agente realiza uma ação e o ambiente produz a observação e o custo de acordo com sua dinâmica, que pode ser desconhecida.

Neste treinamento, o objetivo é a rede neural descobrir a política de seleção de ações que minimizem o custo acumulado.

As redes neurais são mais adequadas a domínios onde os problemas a serem resolvidos não são lineares e portanto não são simples para serem modeladas em um sistema especialista de tomada de decisão.

ALGORITMOS GENÉTICOS

Os algoritmos genéticos são assim denominados por pertencerem à uma classe particular de algoritmos que implementam conceitos da biologia como herança, mutação, seleção e recombinação.

Normalmente estes algoritmos são implementados como uma simulação computacional onde uma dada população de cromossomos de soluções canditadas para um problema evolui até as melhores soluções.

O processo evolucionário de um algoritmo genético inicia a partir de uma população completamente aleatória de indivíduos, que vão passando pelo processo evolutivo produzindo novas gerações de indivíduos.

A cada nova geração, a energia da população como um todo é avaliada através da seleção estocástica de indivíduos baseada na energia dos mesmos. Estes indivíduos são modificados através de mutação ou recombinação e com isso geram uma nova população que é usada para a próxima iteração do algoritmo.

No domínio dos problemas relacionados a tomada de decisão, os algoritmos genéticos são bem adequados para suporte as fases de projeto e escolha, no processo de tomada de decisão.

No processo de solução de um problema objetivo, o algoritmo evolui soluções até que: não haja nenhuma melhoria extra possa ser conseguida, um número predeterminado de gerações tenha evoluído ou quando o tempo de processamento determinado tenha se completado.

A solução mais adequada será aquela na geração final seja a única a possuir a função que resolva da forma mais adequada o objetivo.

Quando aplicado à problemas com objetivos únicos, o usuário de um sistema baseado em algoritmos genéticos é auxiliado na fase de escolha do processo de decisão.

Por outro lado, na resolução de problemas com múltiplos objetivos, um sistema baseado em algoritmos genéticos fornece várias soluções satisfatórias para os objetivos, e permite ao usuário selecionar a melhor alternativa, portanto ele auxilia a fase de projeto do processo de decisão.

Uma das grandes vantagens no uso de algoritmos genéticos é a sua capacidade de não apenas examinar alternativas de diversas soluções como também a produção de novas soluções para o problema.

Como exemplo de um sistema baseado em algoritmos genéticos, podemos considerar um sistema de trading – compra e venda de ações, onde a energia em questão é o lucro obtido na operação.

Consegue-se lucro quando se efetua compra de uma ação que está com baixo valor e a sua respectiva venda quando valorizada.

A dificuldade de tomada de decisão neste cenário é determinar quando uma ação atinge um valor baixo , como em uma depressão, e quando a mesma está em seu valor mais alto, ou seja, seu pico.

Se o algoritmo for modelado de forma genérica, sem especificar restrições de tempo, valores, indicadores e conceitos, o custo computacional desta solução pode ser muito alto, podendo até ser proibitivo, visto que não é possível conseguir a solução ótima desta tarefa através de todas as possibilidades.

A vantagem do uso de redes neurais é que elas podem encontrar novas soluções ou combinações de soluções através de seu processamento.

A desvantagem é quando o problema possui poucas restrições, ou não se sabe com exatidão quais restrições aplicar, o que pode comprometer ou até inviabilizar a aplicação desta técnica em um problema em particular.

PROPOSTAS EM SISTEMAS DE TOMADA DE DECISÃO

De um modo geral, as propostas inovadoras em sistemas de tomada de decisão são baseadas em modelos híbridos.

Um exemplo de modelo construído de acordo com estas técnicas foi o modelo de agentes Clarion, proposto por Ron Sun e Todd Peterson (Sun, Peterson 1998) para aprendizado seqüencial.

O Clarion era um modelo conexionista híbrido, que possuía uma arquitetura que unificava os aprendizados através de redes neurais, o aprendizado simbólico e o aprendizado de reforço de modo integrado.

O objetivo desta arquitetura era poder construir um modelo que pudesse tratar com aprendizado concorrente,

durante o contato do agente com seu ambiente sem a necessidade de conjuntos de dados pré-construídos ou conceitos anteriormente informados.

Outro objetivo do Clarion era que o modelo permitisse não apenas o aprendizado de tarefas simples, de baixo nível, como também o aprendizado declarativo de alto nível.

Apesar da idéia do Clarion era orientar a tomada de decisão de um agente em um ambiente para sua movimentação adequada, o conceito de aprendizado seqüencial é extremamente importante para o processo de tomada de decisão em outras áreas.

Os problemas relacionados a seqüências podem ser simplificados em quatro categorias básicas: predição, geração, reconhecimento e decisão.

Predição de seqüências é o processo de determinar os próximos elementos de uma seqüência à partir dos seus precedentes.

Geração de seqüências é o processo de gerar elementos de uma seqüência de acordo com uma regra que explicite a ordem natural dos mesmos.

Reconhecimento de seqüências é o processo de determinação se uma dada seqüência é legítima de acordo com algum critério.

Decisão envolvendo seqüências é o procedimento de selecionar um conjunto de ações na ordem (seqüência) correta para a completude de um objetivo.

Os três primeiros problemas, predição, geração e reconhecimento de seqüências normalmente se baseia em modelos legítimos de seqüências para seu desenvolvimento através de algum método de treinamento com exemplos de dados.

Para isso, os modelos podem ser cadeias estatísticas, redes neurais ou um outro modelo adequado.

De um modo geral, as técnicas de treinamento conseguem extrair tendências comuns ao conjunto de dados de exemplos.

A tomada de decisão seqüêncial possui três variações, podendo ser orientada a um objetivo, orientada a trajetória ou orientada à reforço.

A typical task faced by organizations is classification decision making. In a classification task, agents gather information about

problems, classify them, and then make further decisions based on the classification. For instance, an organization may classify

a company as financially promising or unpromising, and on that basis decide to approve or not approve a loan. In this case, the

task is to determine whether a blip on a screen is a hostile aircraft, a flock of geese, or a civilian aircraft (Carley et al 1998).

Hence, this is a ternary choice task. It has been used before in studying organizational design (e.g., in Carley and Prietula

1992; Ye and Carley 1995; Carley and Lin 1995).

In each case, there is a single object in the airspace. The object has nine different attributes, each of which can take on

one of three possible values (e.g., its speed can be low, medium, or high). An organization must determine the status of an

observed object: whether it is friendly, neutral or hostile. There are a total of 19,683 possible objects, and 100 problems are

chosen randomly (without replacement) from this set.

The true status of an object is determinable by adding up all nine characteristic values. If the sum is less than 17, then it is

friendly; if the sum is greater than 19, it is hostile; otherwise, it is neutral.

No one single agent has access to all the information necessary to make a choice. Decisions are made by integrating

separate decisions made by different agents, each of which is based on a different set of information. Each organization is

assumed to have sufficient personnel to observe all the necessary information (in a distributed way).

In terms of organizational structures, there are two archetypal structures of interest: (1) teams, in which decision makers

act autonomously, individual decisions are treated as votes, and the organization decision is the majority decision; and (2)

hierarchies, which are characterized by agents organized in a chain of command, such that information is passed from

subordinates to superiors, and the decision of a superior is based solely on the recommendations of his subordinates (Carley

1992). In this task, only a two-level hierarchy with nine subordinates and one superior is considered.

In addition, organizations are distinguished by the structure of information accessible by each agent. There are two

varieties of information access: (1) distributed access, in which each agent sees a different subset of three attributes (no two

agents see the same set of three attributes), and (2) blocked access, in which three agents see exactly the same attributes. In

both cases, each attribute is accessible to three agents.

Several simulation models were considered in the study (Carley et al 1998). Among them, CORP-ELM produces the

most probable classification based on an agent's own experience, CORP-P-ELM stochastically produces a classification in

accordance with the estimate of the probability of each classification based on the agent's own experience, CORP-SOP

follows organizationally prescribed standard operating procedure (which involves summing up the values of the attributes

available to an agent) and thus is not adaptive, and Radar-Soar is a (somewhat) cognitive model built in SOAR, which is based

on search through problem spaces (Rosenbloom et al 1991).

CONCLUSÃO

Existem diversas técnicas para construção de sistemas de auxílio à tomada de decisão, não existindo uma que seja a solução para todas as situações pssíveis.

Redes bayesianas são adequadas para problemas que naturalmente possuem um certo grau de incerteze e existe a possibilidade de realimentá-la com informações para aumentar a sua acurácia, como em um filtro anti-spam.

Para situações onde problema a ser decidido possui um histórico de dados

REFERÊNCIAS

BHATNAGAR, R. K.; KANAL, L. N. Handling Uncertain Information: A Review of Numeric and Non-numeric Methods In: KANAL, L.N.; LEMMER, J.F. (Editors). Uncertainty in Artificial Intelligence. Amsterdam: Elsevier Science. 1986

PEARL, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Inc., 1988.

[3] HOLSHEIMER, M. and SIEBES, A. Data Mining: the search for knowledge in databases 1999.

[4] BRAZDIL, P. Construção de Modelos de Decisão a partir de Dados. Disponível por WWW em: http://www.ncc.up.pt/~pbrazdil/Ensino/ML/DecTrees.html,

[5] INGARGIOLA, Giorgio. Building Classification Models: ID3 and C4.5. Disponível por WWW em: http://www.cis.temple.edu/~ingargio/cis587/ readings/id3-c45.html, 1996.