

OPERACIONALIZAÇÃO DE UM MODELO DE BAIXO CUSTO DE PREVISÃO DE MARÉS PARA COSTA BRASILEIRA

MSC. BENJAMIN L. FRANKLIN

EPAGRI*, doutorando no programa de Engenharia e Gestão de Conhecimento EGC
- Universidade Federal de Santa Catarina
E-mail: belfra@ciram.com.br, belfra@uol.com.br

PHD. HAMILTON JUSTINO VIEIRA

EPAGRI*
E-mail: vieira@epagri.rct-sc.br

PHD. PATRÍCIA SUNYE

EPAGRI*
E-mail: sunye@epagri.rct-sc.br

MSC. RODRIGO PACHECO

EPAGRI*
E-mail: pacheco@epagri.rct-sc.br

**Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S.A.*
Avenida Admar Gonzaga 1347, Itacorubi CP 502 - Florianópolis Santa Catarina

Abstract— This work intends to show the operational results of a tide forecast model, based on an artificial neural network (ANN). The objective is a low cost implementation and maintenance and the emission of alerts for cellular phones in case of a forecast of extreme events. We will be discussing the technology used, the various forms of integration with systems legacies, the difficulties of operating in real time and also point out suggestions to improve the product in a reasonable period of time.

Keywords— Artificial Neural Network, Variability of the Level of the Sea, Meteorological Tide, Forecast of Time Series.

Resumo— Este trabalho pretende mostrar os resultados da operacionalização de um modelo de previsão de maré, baseado em redes neurais artificiais, que tem como objetivo um baixo custo de implementação e manutenção, além da emissão de alertas por celular em caso da previsão de eventos extremos. Comentaremos a tecnologia empregada, as formas de integração com sistemas legados, as dificuldades de operação em tempo real, além de apontar sugestões para melhoria do produto em médio prazo.

Palavras-chave— Redes Neurais Artificiais, Variabilidade do Nível do Mar, Maré Meteorológica, Previsão de Séries Temporais.

1 Introdução

A EPAGRI é uma instituição que tem como missão a pesquisa e desenvolvimento em recursos ambientais no Estado de Santa Catarina. Dentre as principais atividades do centro destacam-se o monitoramento, a previsão e o envio de alertas para diversos setores da sociedade, cuja atividade dependa de fatores climáticos. O centro conta hoje com um dos maiores bancos de dados de informações ambientais do Brasil.

A aplicação de técnicas de Inteligência Artificial (IA) permite a extração de conhecimento baseado no histórico das variáveis climáticas, ou seja, um conhecimento extraído de “baixo para cima” (EPSTEIN, 1996), sem a presença de um modelo teórico bem-estabelecido, ao contrário dos

“tradicional” modelos de previsão meteorológicos (BROOKS, 2004).

Na EPAGRI foram encontradas diversas possibilidades de aplicação de técnicas de IA, destacando a detecção de padrões entre variáveis climáticas, devido à extensa base de dados com mais de 50.000.000 de registros, contendo dados de séries temporais desde 1911. A exploração desta base de dados pode produzir serviços com alto valor agregado, como sistemas de previsão e alertas totalmente automatizados, funcionando 24 horas por dia, sete dias da semana.

Um exemplo destes serviços possíveis é a previsão de maré em tempo real. Este serviço pode auxiliar o embarque e desembarque de carga em portos de forma a melhorar a qualidade das decisões

e servir como diferencial competitivo a empresas do setor.

A produção deste tipo de sistema enfrenta, no entanto, dificuldades práticas que muitas vezes inviabilizam a operacionalização do modelo, tais como: instalação de instrumentos, qualidade de dados, rotinas de manutenção, preço de sensores, despesas de comunicação de dados, entre outras...

Com a intenção de viabilizar comercialmente um sistema de alertas totalmente automatizados, em tempo real - e com custos reduzidos - deu-se início a este programa, que tem o objetivo pragmático de viabilizar a instalação de sistemas de previsão e alertas, baseados em IA e, por isso, suficientemente genéricos, para que todo o litoral da costa brasileira possa ser beneficiado. Descreveremos, nas próximas seções, os desafios, os acertos e as dificuldades para a realização deste empreendimento.

2 Simplificando modelos e minimizando custos operacionais

A produção de modelos de previsão de maré esbarram na dificuldade de obter dados com qualidade e uma série temporal adequada. Tendo a intenção de produzir um modelo de fácil operacionalização, optamos por incluir, como entrada de dados, apenas as variáveis essenciais e que pudessem manter a qualidade operacional do modelo. Outras pesquisas, na criação de modelos utilizando RNAs para predição de maré, como (MITIDIERI et al, 2006 e SARAIVA, 2003) incluem variáveis como a pressão atmosférica, tensão zonal e meridional do vento, séries da maré meteorológica, valor do vento observado, entre outras, como forma de tornar seus resultados mais consistentes. A adição de novos sensores aumenta o preço de implementação do sistema, dificultando a disseminação do produto.

Optamos, nesta pesquisa, por reduzir o número de sensores a apenas um: um marégrafo para medição da maré observada, conjugado com o cálculo harmônico da maré astronômica (dados obtidos da DHN - Diretoria de Hidrografia e Navegação), ambos normalizados entre valores entre zero e um, e interpolados de quinze em quinze minutos com o método de Lagrange (ARCHER, 2006), para normatização de sua série histórica, visto que a tábua de maré astronômica oferecida pela DHN fornece apenas dados diários pontuais de máxima e mínima.

Lembrando que a maré astronômica é um componente da maré observada que pode ser calculado por constantes harmônicas para qualquer instante de tempo (FRANCO, 1981). A maré observada, no entanto, sofre influências meteorológicas locais que mudam os resultados dos cálculos astronômicos. A utilização de Redes Neurais Artificiais (RNAs) conjuga os efeitos da maré astronômica com seus efeitos meteorológicos, de

forma a produzir a previsão da maré observada (MITIDIERI et al, 2006).

Após o teste de diversas topologias de RNAs, concluímos a simplificação do modelo na decomposição de não apenas uma RNA, mas três delas, conforme o instante de tempo referente para 6, 12 e 24 horas - da seguinte forma: em todas as RNAs - perceptron de múltiplas camadas (LOESCH, 1996)- temos duas camadas escondidas de 12 neurônios e um neurônio de saída, correspondendo ao valor da maré no horário a ser calculado.

A Previsão para 6h tem o seguinte conjunto de entrada de dados

Maré observada da data atual
Maré astronômica da data atual
Maré observada da data atual menos 6 horas
Maré astronômica da data atual menos 6 horas
Maré observada da data atual menos 12 horas
Maré astronômica da data atual menos 12 horas
Maré observada da data atual menos 18 horas
Maré astronômica da data atual menos 18 horas
Maré observada da data atual menos 24 horas
Maré astronômica da data atual menos 24 horas
Maré astronômica da data atual **mais 6 horas**

A Previsão para 12h tem o seguinte conjunto de entrada de dados

Maré observada da data atual
Maré astronômica da data atual
Maré observada da data atual menos 6 horas
Maré astronômica da data atual menos 6 horas
Maré observada da data atual menos 12 horas
Maré astronômica da data atual menos 12 horas
Maré observada da data atual menos 18 horas
Maré astronômica da data atual menos 18 horas
Maré observada da data atual menos 24 horas
Maré astronômica da data atual menos 24 horas
Maré astronômica da data atual **mais 12 horas**

A Previsão para 24h tem o seguinte conjunto de entrada de dados

Maré observada da data atual
Maré astronômica da data atual
Maré observada da data atual menos 6 horas
Maré astronômica da data atual menos 6 horas
Maré observada da data atual menos 12 horas
Maré astronômica da data atual menos 12 horas
Maré observada da data atual menos 18 horas
Maré astronômica da data atual menos 18 horas
Maré observada da data atual menos 24 horas
Maré astronômica da data atual menos 24 horas
Maré astronômica da data atual **mais 24 horas**

É importante lembrar que nenhum tratamento foi dado aos dados de maré observada, além da sua normatização e interpolação, simplificando o processo de aquisição de dados, visando à criação de previsões horárias. A

interpolação tem um papel importante no sentido de criar tolerância às falhas eventuais no sistema de transmissão ou captação de dados, as equações de interpolação de Lagrange funcionam adequadamente para falhas menores que seis horas na série histórica.

3 Treinamento das RNAs por computação evolutiva distribuída

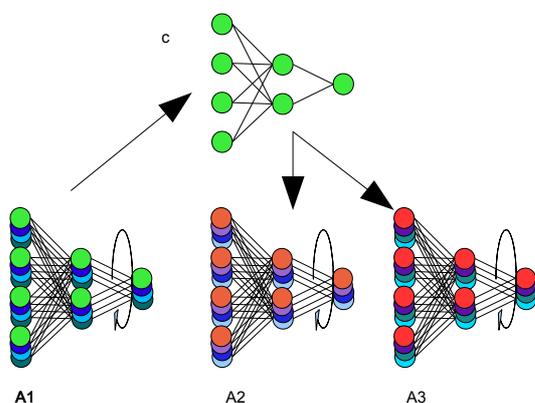


Figura 1. Evolução distribuída de RNAs.

O treinamento da rede foi realizado por algoritmos genéticos paralelos (PIT, 2004). Esta abordagem deriva da computação evolutiva que intenta *aplicar o processo de evolução natural como um paradigma de solução de problemas, a partir de sua implementação em computador.* (ZUBEN, 2003). Esta abordagem permite a otimização do treinamento de uma RNA ao distribuir o processamento em diferentes sistemas computacionais, de forma paralela e colaborativa (ALBUQUERQUE et al, 2004).

Tendo esta abordagem em mente, criaram-se agentes de software, no sentido de criar programas autônomos com sociabilidade, como proposto por (FRANKLIN, 1996) onde cada um contém uma população inicial de n RNAs, como ilustrado na fig. 1, onde mostramos três agentes de software A1, A2 e A3 - este número inicial n é definido pelo usuário administrador. Cada RNA contida no agente tem uma matriz inicial de pesos gerada aleatoriamente. O processo de treinamento segue como descrito, para todos os agentes, que funcionam paralelamente em dispositivos diferentes:

1. Estabelecida a população inicial de RNAs, testa-se o desempenho de cada uma delas sobre o problema apresentado, ou seja, o relacionamento das informações de entrada com as informações de saída. As RNAs que produzirem uma resposta mais adequada serão selecionadas para a próxima fase do processo. Esta seleção é baseada em uma porcentagem P sobre a população inicial, definida pelo usuário, que selecionará as melhores redes - as que forem melhor pontuadas no teste de ajuste, como é comum na maioria dos algoritmos genéticos

(GOLDBERG, 1998) - as restantes são apagadas da memória.

2. Escolhemos sobre a população selecionada, aleatoriamente, um casal de RNAs e trocaremos, também de forma aleatória, seus pesos de matriz de forma a manter uma troca de 50% do material “genético” dos pais sobre a nova rede “filha”. Repete-se o processo até que a população inicial n seja recuperada. Obs.: caso o casal escolhido tenha a mesma nota de ajuste, então é gerada uma mutação na matriz de pesos, ou seja, escolhe-se aleatoriamente um novo peso que substitui alguns pontos na matriz, “revigorando” a RNA “filha”.
3. A RNA resultante do processo é então comparada com uma RNA de controle, como ilustrado na fig. 1. C. Caso o resultado obtido pelo agente tenha melhor desempenho que o agente de controle C, então as configurações do agente – leia-se, seus pesos sinápticos – são transferidos para o agente de controle.
4. Caso o resultado da rodada evolutiva do agente seja pior que o resultado da RNA de controle, então uma das RNAs do agente recebe as configurações da RNA de controle, revitalizando assim o agente como um todo, conforme ilustrado na figura, para agentes 2 e 3.
5. Repete-se, desta forma, o algoritmo, reiniciando um processo de seleção sobre a nova população durante um número G de gerações definidas pelo usuário administrador. Este processo, gradualmente, converge as RNAs para um padrão estável de resposta, que será gravado em um banco de dados e reutilizado depois, quando solicitado.

O treinamento foi realizado sobre três meses de dados horários (outubro, novembro e dezembro de 2005), sendo este o período necessário para operacionalização do modelo a partir da instalação do marégrafo.

3 Resultados Obtidos

Este sistema encontra-se atualmente em operação em dois portos, de Itajaí, em Santa Catarina, e de Vitória, no Espírito Santo, como mostram as figuras 2 e 3.

Ambos mantêm uma performance muito atraente, visto o erro médio absoluto e sua comparação com a previsão DHN, para os meses de setembro, outubro e novembro de 2006, conforme os resultados indicados na tabela 1:

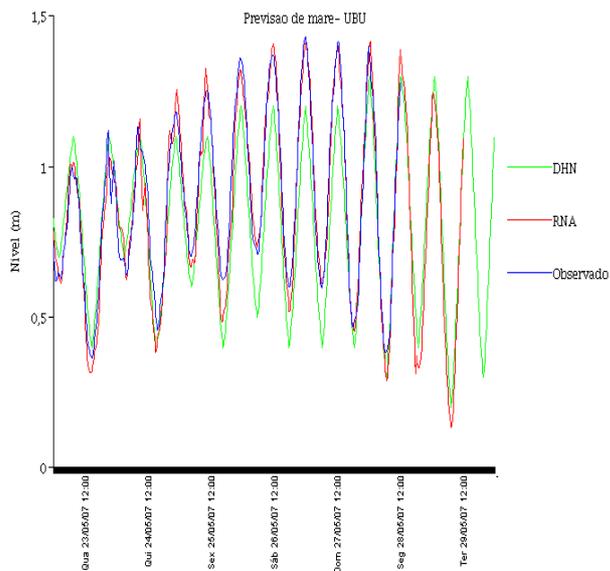


Figura 2. Porto de Ubu, Espírito Santo.

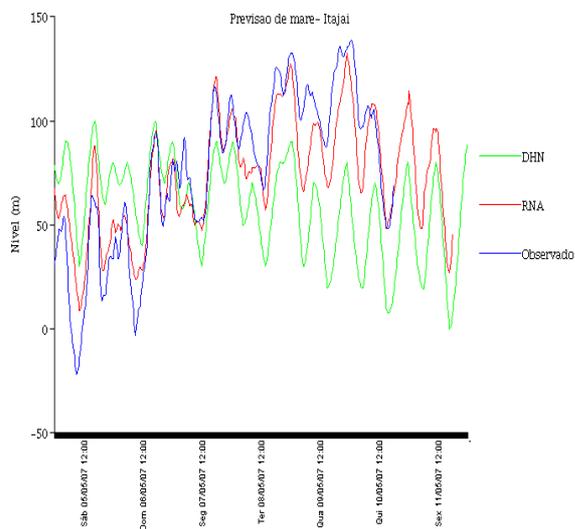


Figura 3. Porto de Itajai, Santa Catarina.

	Erro DHN (cm)	Erro RNA 06H (cm)	Erro RNA 12H (cm)	Erro RNA 24h (cm)
Itajai	30,65	13,31	18,66	28,68
Vitória	12,24	4,87	5,60	8,60

Tabela 1. Resultados comparativos de erro médio absoluto entre maré observada e previsões das RNAs e DHN nos meses de setembro, outubro e novembro de 2006.

4 Operação

O sistema faz uma conexão com a base de dados da EPAGRI, para obter os dados necessários de entrada de uma RNA, os dados de saída e a configuração básica de um Perceptron de Múltiplas Camadas, através de comandos em SQL.

A plataforma JAVA serviu de base tecnológica para a implementação do projeto e integração com o sistema legado da EPAGRI, como a

base de dados ORACLE e os sistemas de envio de mensagens de SMS e os servidores de E-mail, WEB, FTP e outros serviços.

4 Reduzindo o custo de transmissão de dados

A quase totalidade das estações automáticas que compõem a rede da EPAGRI possui um sistema telemétrico. Hoje existe o domínio tecnológico de envio de dados através dos seguintes sistemas: telefonia fixa, telefonia celular GSM (torpedos), telefonia celular GPRS (e-mails), Internet, satélite (orbcomm e inmarsat) e rádio UHF.

A tecnologia GPRS (General Packet Radio Service) é um serviço de segunda geração de telefonia móvel e permite a conexão com a Internet em tempo real. A utilização dos protocolos da Internet permite a transmissão de dados em multimídia, além de uma comunicação síncrona de informações a uma velocidade de 171,2 kbps (WIKIPEDIA, 2006). Dependendo da política de preços da operadora, a transmissão por GPRS pode ser cobrada por bytes transmitidos, ou por pacotes de kbytes, chegando a R\$ 0,002 por caractere transmitido.

Com o avanço da cobertura celular no Brasil, este sistema tem se mostrado mais rápido, mais eficiente e barato que os demais. Hoje o custo de transmissão de dados via GPRS (dados horários) gira em torno de R\$ 10,00, contra R\$ 300,00 da telefonia fixa e do satélite Orbcomm (este último somente para estações hidrológicas), e dos R\$ 290,00 do celular GSM com SMS.

Além de diminuir o custo de transmissão, a utilização de um módulo de celular (Siemens - TC65) com a tecnologia JAVA embarcada, amplia a capacidade de programação das estações, possibilitando a verificação da transmissão de dados e a repetição seguida, em caso de falhas, ou o envio por via alternativa como SMS - melhorando o índice de transmissões bem sucedidas e, por conseguinte, a qualidade das séries históricas da base de dados, componente vital para a previsão qualificada das RNAs e o sucesso de sua operação.

Os dados são transmitidos segundo a seguinte dinâmica:

1. As informações geradas pelo datalogger (figura 4.1) são enviadas para o módulo TC 65 (fig. 4.3).
2. O módulo TC65 armazena as informações em sua memória principal (fig.4.2), marcando as informações que ainda não foram enviadas corretamente.
3. O módulo TC65 seleciona periodicamente, em sua memória, as informações que não foram enviadas, abre uma conexão com a Internet via protocolo HTTP (fig. 4.4) e

envia as informações para os servidores da EPAGRI (fig. 4.5).

4. Os servidores da EPAGRI enviam um sinal de conclusão, caso a operação de transmissão seja bem sucedida, encerrando o ciclo de transmissão no TC65 que retira a marca de envio da informação armazenada.
5. Caso o sinal de conclusão não seja recebido, o TC65 continua a transmitir a informação armazenada durante um novo ciclo de transmissão.

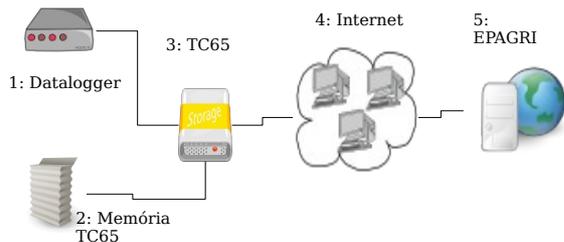


Figura 4. Transmissão de dados para previsão em tempo real.

5 Emitindo alertas para celulares

Os usuários do sistema de monitoramento encontram-se muitas vezes em campo, durante o exercício de suas atividades, em locais de difícil acesso como áreas portuárias ou mesmo em embarcações, onde o acesso a Internet não está disponível.

SMS (Short message service), denomina o serviço de envio de mensagens de texto de até 160 caracteres de um aparelho móvel para outro. Estas mensagens não têm garantia de entrega e podem sofrer atrasos devido ao tráfego de mensagens nos servidores das operadoras. Apesar disso o serviço é amplamente difundido e está disponível em praticamente todos os celulares do mercado, tornando-se uma alternativa atraente para envio de alertas, a um custo de cerca de R\$ 0,40 por mensagem enviada.

O servidor de SMS é um sistema integrado que está vinculado à base de dados da EPAGRI, possibilitando que diversos projetos possam enviar mensagens de acordo com suas necessidades. Por ser uma ferramenta genérica o servidor está restrito as seguintes funções básicas:

1. Consulta periodicamente uma tabela no servidor Oracle contendo todas as mensagens;
2. Seleciona somente as mensagens que não foram enviadas;
3. Envia as mensagens selecionadas, uma por vez, a um módulo de celular através de um

comando AT (padrão de comunicação entre modems);

4. Caso o envio seja bem sucedido, muda o status da mensagem.
5. Caso não seja bem sucedido, então muda o status para um código de erro.
6. Envia mensagem para usuário

Com o servidor de SMS podemos enviar alertas diretamente a um celular, bastando para isso inserir a mensagem na *tabela de mensagens* na base de dados da EPAGRI. Cada linha na tabela de mensagens corresponde a uma mensagem a ser enviada.

O servidor de SMS foi planejado para enviar alertas para um pequeno número de usuários. Sua principal vantagem é a integração com a base de dados da EPAGRI, possibilitando uma ampla personalização destas mensagens, outra vantagem é a independência em relação à operadora de telefonia celular.

O servidor de SMS não é recomendável para o envio de alertas para um grande número de usuários, uma vez que manda uma mensagem por vez e isso pode prejudicar o envio de alertas para os últimos usuários de uma lista extensa. Recomendamos que os alertas utilizem listas em torno de 50 usuários, para que todos possam receber os avisos em menos de uma hora, entre o evento ocorrido e o recebimento do alerta.

O sistema tem uma capacidade de envio de uma mensagem SMS a cada 20 segundos, continuamente, vinte quatro horas por dia sete dias por semana. Esta capacidade limite de $(1 \times 3) * 60 * 24 = 4320$ mensagens por dia ainda não foi testada. Vale lembrar que esta capacidade pode ser ampliada usando outro servidor de SMS.

Devemos lembrar também que a tecnologia SMS não funciona em tempo real, podendo haver atraso na entrega, conforme o tráfego entre as operadoras de telefonia celular. Os melhores resultados estão na entrega de mensagens da mesma operadora, no caso de nosso servidor, a TIM, e tem um tempo de entrega em torno de 1 minuto.

Os piores resultados ocorrem entre diferentes operadoras, em dias específicos (exceções) e podem demorar até 12 horas para realização da entrega. Estes desvios não são a regra para a troca de mensagens entre operadoras, que tem realizado o procedimento em torno de 5 minutos em média. Testamos com sucesso a troca de mensagens entre as seguintes operadoras: TIM, CLARO, VIVO e BRASILTELECOM. É importante ressaltar que não encontramos nenhuma estatística oficial sobre o tempo de envio de mensagens SMS vinda das operadoras, nossas conclusões são empíricas e não podem ser garantidas oficialmente.

3 Conclusão

A criação de um modelo simplificado de previsão e alerta de marés, juntamente com uma solução de baixo custo de coleta e transmissão de dados, pode ajudar a viabilizar a operacionalização de um sistema de previsão e alertas baseado em RNAs, que possa estar disponível a uma ampla fatia da sociedade.

A utilização de RNAs na predição de marés tem se comprovado uma forma robusta e flexível para lidar com condições adversas de operacionalização de modelos, com alto grau de adaptabilidade a características locais de diferentes regiões, como visto nos dois casos em que o modelo foi aplicado.

Novas pesquisas comparando a performance com outros modelos já desenvolvidos poderiam ajudar a elaborar um cálculo de custo/benefício na utilização de diversos sistemas, que poderiam incluir novas variáveis e outros métodos de previsão.

Esperamos brevemente implementar o modelo em outros locais da costa brasileira, além de formalizar um modelo de negócios capaz de formatar uma estrutura de serviços na difusão das previsões, de forma a ampliar a lista de beneficiados.

Agradecimentos

Este projeto não seria possível sem o trabalho de todas as pessoas na EPAGRI, cujo empenho, em frentes tão diversas, mostra continuamente a força de realização de sonhos feitos em equipe.

Referências Bibliográficas

- Albuquerque Ana C., Melo Jorge, Neto Adrião (2004) Parallel genetic algorithm with different evolution behavior for multilayer perceptrons design and learning *Learning and Nonlinear Models – Revista da Sociedade Brasileira de Redes Neurais*, Vol. 2, No. 2, pp. 73-83,
- Archer, Branden and Weisstein, Eric W. Lagrange Interpolating Polynomial. (1996) From *MathWorld--A Wolfram Web Resource*. <http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html> [16/03/06]
- Barreto, J.L. (2001) Introdução às redes neurais artificiais. *Escola Regional de Informática da SBC Regional Sul*, Santa Maria, Florianópolis, Maringá, p.41-71, 5-10
- Brooks, R. (2004) Intelligence without representation. *Artificial Intelligence*. (on-line) <http://www.ai.mit.edu/people/brooks/papers/representation.pdf>. [16/03/04]
- Cilliers, P. (1998) Complexity & Postmodernism: understanding complex systems. *New York: Routledge*.
- Epstein, J. M. e Axtell, R. (1996) Growing artificial societies: social science from the bottom up. MIT Press, Cambridge, MA.
- Franco, A. S. (1981) Tides: fundamentals analysis and prediction. São Paulo: *Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT*.
- Franklin, S. e Graesser, A. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*, Springer-Verlag, 1996.
- Goldberg, D.E. (1998) Genetic Algorithms. USA: *Addison-Wesley Longman Inc*.
- Loesch, C. e SARI, S. T. (1996) Redes Neurais Artificiais: fundamentos e modelos. Blumenau : *Editora da FURB*.
- Mitidieri M. F. et al. (2006) Modelagem da maré meteorológica utilizando redes neurais artificiais: uma aplicação para a baía de paranaguá – PR. *Revista Brasileira de Meteorologia Vol. 21*, n.2. Agosto.
- Pit, J.L. (1995) Parallel Genetic Algorithm: *Leiden University*, (on-line) <http://www.geocities.com/igoryepes/AGParalelos.zip> [03/03/04]
- Russell, S. e Norving, P. (1995) Artificial intelligence: a modern approach. *New Jersey: Prentice-Hall*.
- Saraiva, R.A., Sperb R.M e Truccolo E. C. (2003) Aplicação de uma rede neural temporal em um estudo de caso para previsão do nível do mar. *I Workshop de Tec. da Inf. aplicada ao Meio Ambiente - CBComp*.
- Valença, M. (2003) Aplicação de redes neurais na área de recursos hídricos. *XII Simpósio Brasileiro de Recursos Hídricos*.
- Wikipedia (2006) General Packet Radio Service http://pt.wikipedia.org/wiki/General_Packet_Radio_Service [23.10.06]
- Zuben, Fenando V. (2003) Computação evolutiva: uma abordagem pragmática. *UNICAMP*, 1998. Disponível em <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/tutorial/tutorialEC.pdf>. [17.10.03]