

Algoritmo Híbrido Genético-Fuzzy aplicado em Câncer de Próstata

Murilo A. Tosatti¹, Júlio C. A. Galvão Filho¹, Felipe R. Forbeck¹, André B. Fonseca¹, João P. M. K. Katayama¹, Emmanuel D. Silva¹, Maria J. P. Castanho², Sandra M. G. S. Venske¹, Angelita M. De Ré¹, Fábio Hernandez¹

¹ Departamento de Ciência da Computação
Universidade Estadual do Centro-Oeste (UNICENTRO)
Caixa Postal 3.010 - 85.015-430 - Guarapuava - PR – Brasil

² Departamento de Matemática
Universidade Estadual do Centro-Oeste (UNICENTRO)
Caixa Postal 3.010 - 85.015-430 - Guarapuava - PR – Brasil

murilotosatti@gmail.com, juliogalvaofilho@yahoo.com.br,
hopper.frf@gmail.com, andrebritofonseca@gmail.com,
minoru.koka@gmail.com, emmanueldsilva@yahoo.com.br, zeza@unicentro.br,
sandravenske@gmail.com, amdre65@yahoo.com.br,
hernandes.fabio@gmail.com

Abstract. *Combining different Computational Intelligence techniques has obtained significant results in several areas. Genetic algorithms combined with fuzzy systems present advantages over their isolated utilization. In this paper a genetic-fuzzy system applied to the medical area is proposed. Two different approaches are implemented and compared. The proposed approach presents better results than the classic one.*

Resumo. *A combinação de diferentes técnicas de Inteligência Computacional vem obtendo resultados significativos em diversas áreas. Algoritmos genéticos combinados com sistemas fuzzy apresentam vantagens sobre a utilização destas técnicas isoladas. Neste trabalho é proposto um sistema genético-fuzzy aplicado à área médica. São implementadas e comparadas duas abordagens diferentes. A abordagem proposta apresenta melhores resultados que a clássica.*

1. Introdução

Sistemas genético-fuzzy [Cordon *et al* 1997] ajudam a explorar e questionar a combinação da computação evolucionária e a lógica fuzzy. A técnica de integrar algoritmos genéticos [Goldberg 1989] e lógica fuzzy [Dubois e Prade 1980, Yager e Zadeh 1994] visa melhorar o desempenho dos mecanismos envolvidos visto que, tomados individualmente, cada um apresenta vantagens e limitações que podem ser exploradas nessa combinação.

Existem importantes trabalhos publicados abordando os temas, com resultados interessantes especialmente em aplicações médicas. Em [Britto e Lopes 2001] é descrita uma abordagem de treinamento de um sistema fuzzy de apoio ao diagnóstico médico (câncer de mama e desordens do fígado), por meio de um algoritmo evolucionário baseado em exemplos. Em [Peña-Reyes e Sipper 1998, 1999] o diagnóstico de câncer também é alvo de utilização de sistemas genético-fuzzy. Já [Castanho *et al* 2008]

desenvolveu um sistema baseado em regras *fuzzy* para prever o estágio patológico do câncer de próstata, com dados existentes na literatura e auxílio de especialistas.

No caso específico do câncer de próstata, o grau de confinamento do câncer é fator determinante na análise de benefício do procedimento invasivo (cirurgia) para a remoção do tumor. Os riscos da cirurgia justificam-se apenas se o tratamento tiver razoável chance de sucesso, isto é, se o tumor estiver confinado no órgão. Caso contrário, outras terapias podem ser aplicadas.

Matemáticos e especialistas procuram refinar as regras para diagnóstico (usando sistemas especialistas, por exemplo) com base na literatura, experiências dos profissionais da área e observações de casos. São utilizados exames que retornam valores contínuos (como o PSA - *Prostate-Specific Antigen*) e exames sujeitos a variações perceptivas dos profissionais (como os exames de toque e o escore de *Gleason* da biópsia).

O objetivo deste trabalho é propor uma abordagem alternativa baseada em sistemas genético-*fuzzy* e compará-la com a abordagem clássica de *Pittsburgh* [Smith 1980], aplicando-as ao problema da determinação do estágio do câncer de próstata. O sistema proposto é baseado em regras *fuzzy* para auxiliar na determinação das características relevantes aos exames e seus graus de influência. Sua configuração, desde a formulação de funções de pertinência até a agregação de regras, é executada com o auxílio de um algoritmo genético.

Este artigo está organizado como segue: a Seção 2 apresenta o sistema genético *fuzzy* desenvolvido; na Seção 3 alguns resultados principais são descritos; as conclusões e trabalhos futuros estão na Seção 4, seguidos das referências.

2. Metodologia

O sistema proposto é denominado híbrido por ser resultado da combinação de duas técnicas de inteligência artificial:

1. Um componente dinâmico para busca em um espaço de soluções (algoritmo genético).
2. Um componente estático, representando a relação e graus de influência dos fatores envolvidos (sistema *fuzzy*).

A Figura 1 retrata os elementos do sistema e suas relações. Os dados da base de casos são utilizados pelo Avaliador para determinar a eficiência das soluções encontradas.

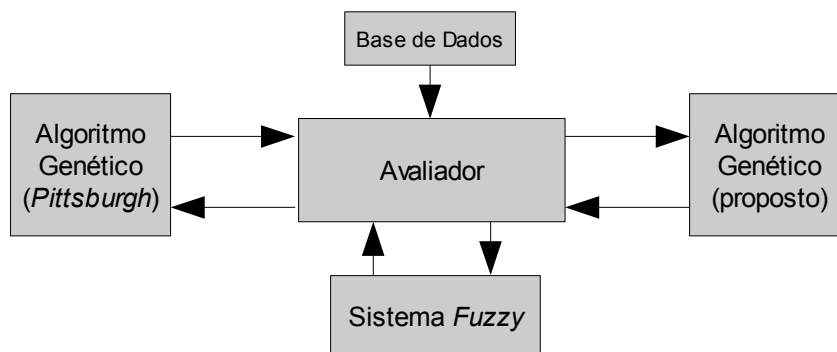


Figura 1. Arquitetura do sistema.

Para encontrar as configurações desejadas para o sistema *fuzzy*, foram implementadas (utilizando a linguagem Java) duas instâncias do algoritmo genético. A primeira, baseada na abordagem *Pittsburgh* [Smith 1980], possui o fluxo de execução padrão de um algoritmo genético e requer menor quantidade de recursos computacionais. A segunda é um algoritmo genético que utiliza a sobreposição de execuções como técnica de otimização dos indivíduos. Com isso a segunda ocupa mais espaço e maior poder de processamento.

O Sistema *Fuzzy* (Figura 1) é o componente que avalia as soluções encontradas pelo algoritmo genético. Utilizando um objeto, gerado pelo leitor da base de dados, as funções de pertinência e as regras, obtidas em um indivíduo do algoritmo genético, o sistema *fuzzy* obtém uma análise computacional sobre o confinamento do câncer do caso clínico em questão.

O avaliador utiliza os valores retornados pelo sistema *fuzzy* para analisar a frequência de acertos de um indivíduo (solução) do algoritmo genético sobre a base de casos. Com isso, pode-se quantificar a validade dessa solução em relação às outras e sua eficácia.

Cada componente do sistema é descrito pormenorizadamente nas subseções seguintes.

2.1. Base de Casos

Para configuração e validação do sistema, foi utilizada uma base de dados contendo informações sobre 285 pacientes que se submeteram à prostatectomia radical no Hospital das Clínicas da Faculdade de Medicina da UNICAMP.

Formatados em uma planilha eletrônica, os dados representam uma tabela com 285 pacientes e 57 colunas com informações de seus estados pré e pós-cirúrgicos. Entretanto, para a realização do estudo, foram utilizadas as informações sobre os três exames abordados (*Gleason*, PSA e toque) e o grau de confinamento observado após a análise do tecido removido (comprometimento dos tecidos adjacentes e margens cirúrgicas).

Com os resultados dos exames (*Gleason*, PSA e toque) é possível estabelecer o estado clínico do paciente e prever o grau de confinamento do tumor. Estes três exames combinados produzem melhores resultados do que analisados separadamente. Neste trabalho eles são combinados utilizando um sistema baseado em regras *fuzzy*.

O estágio do tumor, confinado ou não-confinado na próstata, só é conhecido após a análise pós-cirúrgica. Esses dados são utilizados para comparar os resultados obtidos com a realidade. O treinamento do sistema (formulação das funções e regras) é influenciado por essa comparação, buscando-se maximizar a quantidade de acertos.

2.2. Componente Dinâmico: Funcionamento dos Algoritmos Genéticos

O sistema genético-*fuzzy* proposto foi implementado utilizando algoritmos genéticos baseados em duas abordagens, conforme descrito na Figura 1. Porém, apesar da diferença no fluxo de execução, os algoritmos possuem alguns componentes em comum: indivíduos, dinâmica da população, método de seleção e função de aptidão.

Neste projeto os cromossomos dos indivíduos são divididos em duas partes: funções de pertinência e regras. Três conjuntos de funções de pertinências são utilizados para caracterizar os exames em análise (PSA, toque e *Gleason*). Essas funções são da

forma trapezoidal e representadas por quatro pontos (vértices do trapézio). As regras possuem um antecedente e um conseqüente. O antecedente é formado por uma combinação de valores para os 3 exames, enquanto o conseqüente é o resultado da análise, se o câncer está confinado ou não. Um exemplo de estrutura de um indivíduo pode ser visualizado na Figura 2.

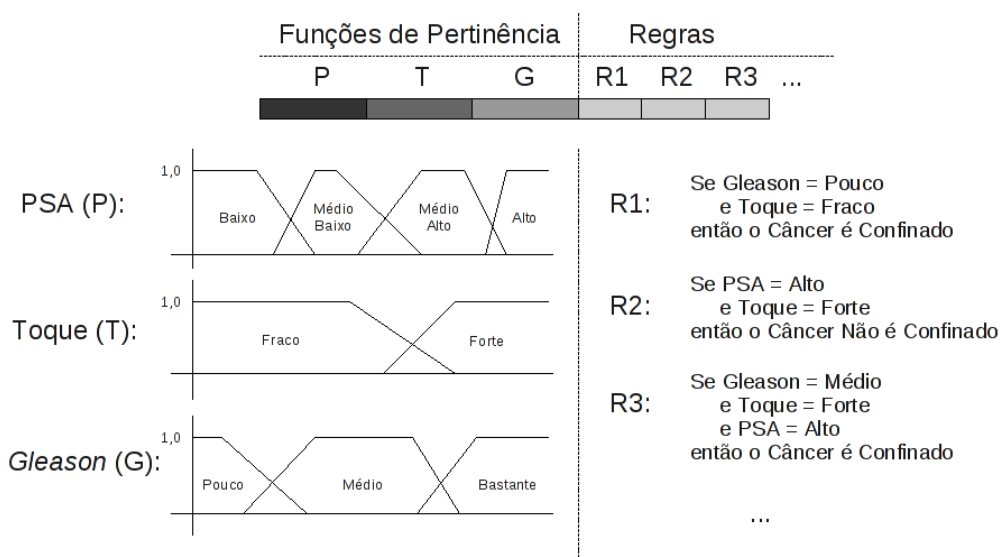


Figura 2. Representação do cromossomo.

A população é formada por uma quantidade fixa de indivíduos. Durante a recombinação são gerados novos indivíduos, aumentando o tamanho dessa população. Porém, ao final de cada geração, indivíduos sobressalentes são eliminados. Esse ajuste é realizado baseando-se no *fitness* (função de aptidão) dos indivíduos: os indivíduos com menores valores de *fitness* possuem maior probabilidade de serem eliminados.

Os operadores genéticos de mutação e recombinação são aplicados em uma parcela selecionada da população. A seleção para mutação ocorre de forma aleatória. Por outro lado, para a recombinação é utilizado o método da Roleta Viciada [Linden, 2006]. Com isso, os indivíduos com os maiores *fitness* possuem maiores chances de se reproduzirem, ao mesmo tempo em que qualquer indivíduo da população pode sofrer mutações, respeitando-se as quantidades especificadas pelas respectivas taxas.

A mutação é realizada em um ponto do cromossomo, podendo: remover, gerar ou alterar uma regra; ou alterar uma função de pertinência. A recombinação também é executada em um ponto, gerando dois novos indivíduos com partes dos cromossomos de ambos os pais. É importante observar que esses operadores possuem restrições na abordagem proposta, sendo que cada instância do algoritmo realiza ações em uma das duas partes do cromossomo: nas regras ou nas funções de pertinência.

A responsabilidade do sistema *fuzzy* está relacionada à mensuração do *fitness* dos indivíduos do algoritmo genético. O sistema é configurado utilizando-se as funções de pertinência e as regras definidas por um indivíduo, e, sua aptidão é calculada de acordo com a porcentagem de acertos em relação à base de dados. A quantidade de resultados corretos é dividida pelo total de testes obtendo-se um valor entre 0 e 1.

O primeiro algoritmo genético implementado neste trabalho seguiu a abordagem *Pittsburgh* [Smith 1980]. Nessa abordagem, os indivíduos representam uma solução completa para o problema, contendo representações para as funções de pertinência e as

regras em seus genes. Os operadores genéticos podem atuar em qualquer posição do cromossomo, respeitando os limites de valores de cada gene.

A abordagem proposta foi implementada seguindo o esquema da Figura 3, sendo dividida em duas instâncias, denominadas *Algoritmo Externo* e *Algoritmo Interno*.

O Algoritmo Externo é o responsável pela evolução das funções de pertinência, sendo que seu funcionamento segue o padrão estabelecido pelos algoritmos genéticos. Entretanto, nessa instância, não são efetuadas alterações nas regras. Quando um indivíduo da população sofre a ação de um operador genético ocorre um desvio de fluxo, sendo executada a próxima instância (Algoritmo Interno). Após a execução dessa segunda instância, um indivíduo otimizado é obtido e inserido na população.

No caso do Algoritmo Interno, essa instância proporciona a otimização das regras do indivíduo, de acordo com as funções de pertinência. Agindo sobre um indivíduo da instância externa, esse algoritmo gera uma nova população através da replicação daquele. Com o intuito de perturbar a região de busca, as regras de 50% dessa população são alteradas por valores aleatórios. O melhor indivíduo encontrado através das gerações é retornado para a instância externa.

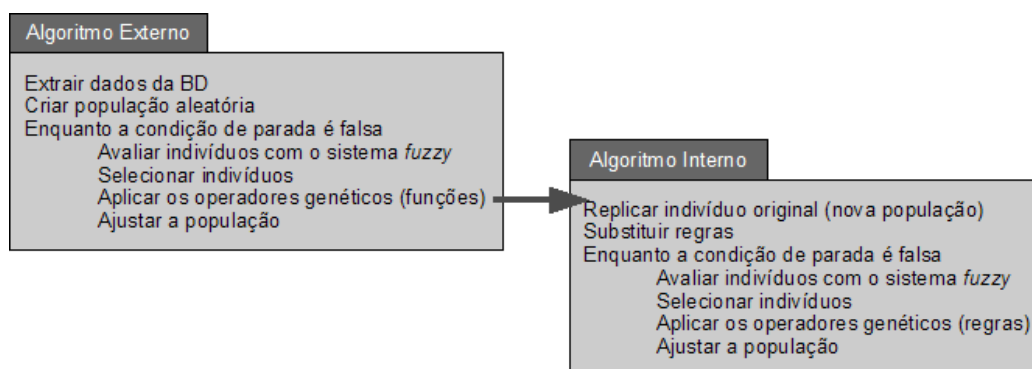


Figura 3. Funcionamento do algoritmo genético (abordagem 2).

2.3. Componente Estático: Estrutura do Sistema *Fuzzy*

O sistema *fuzzy* responde à necessidade de avaliação das soluções encontradas pelo algoritmo genético. Cada indivíduo do algoritmo genético possui as configurações necessárias ao funcionamento desse sistema.

As funções de pertinência e as regras obtidas em um indivíduo do algoritmo genético são utilizadas para compor os bancos de conhecimentos e de regras do sistema *fuzzy*. Estes bancos compõem o motor de consulta complementar ao sistema de decisão.

As três características analisadas (correspondentes aos três exames) são variáveis lingüísticas representadas por funções de pertinência, sendo que o *Gleason* possui três termos, o PSA contém quatro e o exame do toque, dois termos. Estas funções são usadas para obter a pertinência dos casos clínicos em cada variável.

O processo de inferência é realizado pelo método proposto por Sugeno [Sugeno e Kang 1986]. Dessa maneira, as regras ativadas, combinadas com os respectivos valores de pertinência, são agregadas e produzem resultados contidos no intervalo entre 0 e 1, em que valores próximos de zero indicam câncer confinado e, quanto mais próximo de um, maior a extensão do tumor extra-prostático. □

A Figura 4 representa o funcionamento do Sistema *Fuzzy*.

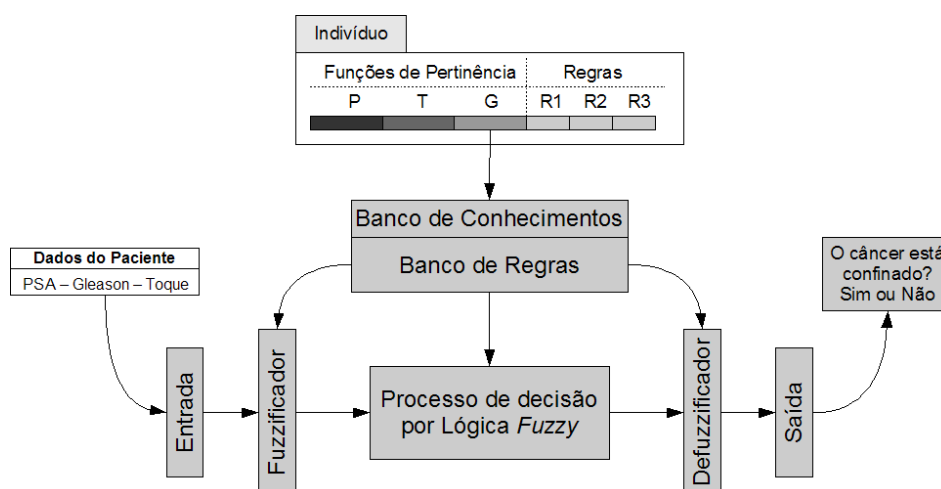


Figura 4. Estrutura e fluxo do sistema *fuzzy*.

3. Resultados Obtidos

Os resultados obtidos com a execução do sistema genético-*fuzzy* implementado são sumarizados nesta seção.

Para cada configuração de mutação e cruzamento foram realizadas cinco execuções do algoritmo, com tamanho da população fixa em 100 indivíduos e 50 gerações. A Tabela 1 apresenta os resultados referentes à abordagem *Pittsburgh*.

Tabela 1. Resultados obtidos com a abordagem *Pittsburgh*.

Configurações	Operadores		<i>Fitness</i> do melhor Indivíduo (%)	Média dos melhores <i>fitness</i> (%)	Média geral dos <i>fitness</i> (%)	
	Mutação	Cruzamento			Primeira	Última
1	1%	10%	63,50	63,26	42,16	47,67
2	1%	25%	64,21	62,80	40,39	48,56
3	1%	50%	64,21	63,50	40,84	52,89
4	5%	10%	62,45	62,45	40,94	44,33
5	5%	25%	63,85	63,26	40,20	47,03
6	5%	50%	63,85	63,50	39,44	50,26
7	10%	10%	63,50	62,80	39,72	45,89
8	10%	25%	63,50	62,91	40,51	48,68
9	10%	50%	64,56	64,20	40,50	48,13

A Tabela 2 apresenta os resultados referentes à abordagem proposta. Os parâmetros definidos para o Algoritmo Interno foram: 20, 40, 30% e 10% para a quantidade de gerações, tamanho da população, taxa de cruzamento e taxa de mutação, respectivamente.

A análise dos resultados permite observar que a taxa de cruzamento de 50% tem as melhores médias de *fitness* para os indivíduos mais aptos, verificando sua relevância.

Tabela 2. Resultados obtidos com a abordagem proposta.

Configurações	Operadores		<i>Fitness</i> do melhor Indivíduo (%)	Média dos melhores <i>fitness</i> (%)	Média geral dos <i>fitness</i> (%)	
	Mutação	Cruzamento			Primeira	Última
1	1%	10%	66,66	66,42	41,24	60,68
2	1%	25%	67,01	66,89	40,63	63,41
3	1%	50%	67,71	67,59	40,07	62,96
4	5%	10%	67,36	67,12	41,10	60,26
5	5%	25%	67,36	67,01	41,11	61,15
6	5%	50%	68,77	68,28	41,26	62,33
7	10%	10%	67,36	67,24	41,01	59,70
8	10%	25%	67,71	67,36	39,82	60,36
9	10%	50%	68,42	68,30	42,43	61,54

Analisando as Tabelas 1 e 2, verifica-se que as médias de *fitness* dos melhores indivíduos obtidos com a abordagem *Pittsburgh* apresentam valores em torno de 64%, enquanto na abordagem proposta as médias estão em torno de 68%.

Da mesma forma, a diferença da média geral dos *fitness* dos indivíduos entre as primeiras e as últimas gerações, observada na abordagem proposta, alcança valores próximos a 21%, enquanto na abordagem *Pittsburgh* esse valor não chega a 13%.

Logo, os resultados mostram que a abordagem proposta apresenta resultados mais significativos que a clássica.

4. Conclusões e Trabalhos Futuros

Sistemas genético-*fuzzy* têm aplicações para ajudar a solucionar problemas nas mais diversas áreas. As aplicações na área da Medicina vêm apontando pontos específicos de aplicações que podem ajudar no diagnóstico médico de doenças. No caso do câncer de próstata, a definição do seu estágio é fator determinante na análise de benefício da realização de cirurgia para a remoção do tumor.

Este trabalho apresentou um sistema genético-*fuzzy*, aplicado ao diagnóstico de estágio de câncer de próstata para auxiliar o profissional da área a aumentar sua acurácia. A configuração do sistema *fuzzy* é executada com o auxílio de um algoritmo genético, sendo implementadas duas abordagens (proposta e *Pittsburgh*) e comparados seus desempenhos.

Apesar das duas abordagens apresentarem evolução dos indivíduos, a proposta apresentou resultados mais significativos, haja vista que a diferença entre as médias de *fitness* das primeiras e das últimas gerações atingiu faixas maiores de valores. Além disso, os melhores indivíduos encontrados possuem valores de *fitness* maiores do que os da abordagem clássica.

Por outro lado, a vantagem da abordagem *Pittsburgh* é a necessidade de menos recursos computacionais.

Como trabalhos futuros, pretende-se comparar os resultados obtidos com o sistema genético-*fuzzy* abordando outras técnicas de Inteligência Computacional, tais

como redes neurais e algoritmos SVM (*Support Vector Machine*).

Agradecimento

Ao Dr. Athanase Billis, do Hospital das Clínicas da UNICAMP, pelo fornecimento dos dados.

Referências

- Britto, J., Lopes, H. S. “Mineração de regras *fuzzy* em bancos de dados por algoritmos genéticos”, In: CBRN - Congresso Brasileiro de Redes Neurais, 2001, Rio de Janeiro. Anais do IV Congresso Brasileiro de Redes Neurais, 2001. v. único. p. 241-246.
- Castanho, M. J. P. ; Barros, L. C. ; Yamakami, A. ; Vendite, L. L. “*Fuzzy* expert system: An example in prostate cancer”, *Applied Mathematics and Computation*, v. 202, p. 78-85, 2008.
- Cordon O, Herrera F, Lozano M. “On the combination of *fuzzy* logic and evolutionary computation: a short review and bibliography”, In: Pedrycz W, editor. *Fuzzy Evolutionary Computation*. Kluwer, 1997:33–56.
- Dubois, D., Prade, H. “*Fuzzy Sets and Fuzzy Systems: Theory and Applications*”, Academic Press, 1980.
- Goldberg, D. E. “Genetic Algorithms in Search, Optimization, and Machine Learning”, EUA: Addison-Wesley, 1989.
- Linden, R. “Algoritmos Genéticos: Uma importante ferramenta da Inteligência Computacional”. Rio de Janeiro. Brasport. 2006.
- Peña-Reyes, C. A., Sipper, M. “Evolving *fuzzy* rules for breast cancer diagnosis”, *Proceedings of 1998 International Symposium on Nonlinear Theory and Applications (NOLTA’98)*, v. 2, pages. 369-372, Lausanne, 1998.
- Peña-Reyes, C. A., Sipper, M. “A *fuzzy*-genetic approach to breast cancer diagnosis”, *Artificial intelligence in Medicine*, v. 17, v. 2, 1999.
- Smith, S. “A Learning System Genetic Adaptive Algorithms”, Ph.D. Dissertation, University of Pittsburgh, Pittsburgh, PA, 1980.
- Sugeno, M., Kang, G.T. “*Fuzzy* Modelling and Control of Multilayer Incinerator”. *Fuzzy Sets and Systems*, Vol. 18: 329-346, 1986.
- Yager, R. R., Zadeh, L. A. “*Fuzzy Sets, Neural Networks, and Soft Computing*”, Van Nostrand Reinhold, New York, 1994.