



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIENTÍFICA E TECNOLÓGICA - 2019

APRENDIZAGEM GENÉTICA DE SISTEMAS BASEADOS EM REGRAS FUZZY PARA PROBLEMAS DE REGRESSÃO

Allen Hichard Marques dos Santos¹ e Matheus Giovanni Pires²

1. Bolsista PIBIC/FAPESB, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: allenhichard21@gmail.com
2. Orientador, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, e-mail: mgpires@ecomp.uefs.br

PALAVRAS-CHAVE: *aprendizado genético de sistemas fuzzy, algoritmos genéticos multiobjetivos, problemas de regressão.*

INTRODUÇÃO

Um Sistema Baseado em Regras Fuzzy (SBRF) tem como objetivo modelar o modo aproximado de raciocínio, permitindo o desenvolvimento de sistemas que imitem a habilidade humana de tomar decisões racionais em um ambiente de incerteza e imprecisão (Mendel, 1995). A aquisição de conhecimento da maioria destes sistemas é realizada a partir de amostras ou exemplos do problema, ou seja, a partir de dados numéricos que representam exemplos do problema.

As formas mais bem sucedidas de extração automática de conhecimento a partir de dados numéricos para a construção de SBRF são as que combinam metodologias de aprendizado de máquina com conceitos de Sistemas Fuzzy (SF). Em particular, um alto número de publicações tem explorado o uso de Algoritmos Genéticos (AG) para a construção automática de SF durante os últimos anos. Esta tarefa pode ser considerada um processo de otimização ou busca em um espaço de soluções em potencial (Herrera, 2008), e os AG são bem conhecidos e amplamente usados como técnica de busca, por sua habilidade de explorar espaços complexos irregulares (Mitchell, 1996).

A construção da Base de Conhecimento (BC) de um SBRF é um problema desafiador para a comunidade científica, em função da necessidade de considerar o balanceamento entre precisão e interpretabilidade. Aumentar a precisão de um SBRF significa aumentar a taxa de acerto (ou acurácia), e aumentar a interpretabilidade consiste em otimizar dois aspectos: a complexidade e a semântica. A complexidade está relacionada com a quantidade de regras, com a quantidade de condições das regras, entre outros. Já a semântica está relacionada com a preservação associada às funções de pertinência (como por exemplo, a cobertura e a distinguibilidade das funções de pertinência) e com regras (por exemplo, consistência).

No final da década de 90 e início da década de 2000, surgiram às primeiras propostas baseadas em otimização multiobjetivo para aprendizado de SBRF (Ishibuchi *et al*, 1997; Voget, 1998; Ishibuchi, 2001). A ideia básica da otimização multiobjetivo é trabalhar com um conjunto de modelos de SBRF com diferentes balanceamentos entre

precisão e interpretabilidade, ao invés de apenas um modelo de sistema. Sendo assim, o especialista ou o processo automático de decisão pode selecionar, de acordo com os requisitos do problema, o modelo mais adequado. Uma das formas muito utilizadas atualmente para a construção automática de SBRF é por meio de Algoritmos Evolutivos Multiobjetivos (AEMO) (Deb *et al*, 2002). A utilização de AEMO pode ser aplicada em diferentes componentes de um SBRF, como por exemplo, no aprendizado da Base de Regras (BR) e/ou no ajuste das funções de pertinências.

Os Sistemas Baseados em Regras Fuzzy têm sido amplamente usados para a resolução de diversos tipos de problemas, dentre eles os de regressão. Um problema de regressão consiste em estimar uma função a partir de um conjunto de exemplos com algum ou nenhum conhecimento da forma da função. A tarefa de regressão é conceitualmente similar à de classificação. A principal diferença é que o atributo a ser predito é contínuo em vez de discreto.

Portanto, o objetivo central deste trabalho consiste no desenvolvimento de um sistema capaz de realizar o aprendizado da BC de um SBRF, por meio de AEMO, para a resolução de problemas de regressão, mantendo o balanceamento entre a interpretabilidade e a precisão do sistema.

METODOLOGIA

Para a modelagem dos atributos da base de dados, que contém as informações do problema de regressão, foi utilizado três conjuntos fuzzy distribuídos de maneira uniforme, onde os conjuntos laterais possuem funções de pertinências trapezoidais e o conjunto central triangular. O AEMO utilizado no aprendizado da BC do sistema fuzzy foi o *Non-Dominated Sorting Genetic-II* (NSGA-II), com o objetivo de construir uma base de regras e ajustar as funções de pertinência simultaneamente, usufruindo ao máximo da sinergia que existe entre a BR e BD.

A população inicial do AEMO é gerada de maneira aleatória com exceção de um único indivíduo, denominado de semente. O objetivo de usar esta semente é otimizar o tempo de convergência do algoritmo, em outras palavras, minimizar os recursos computacionais, tais como, memória e tempo de processamento. A semente é gerada através do algoritmo de Wang e Mendel (Wang e Mendel, 1992).

A taxa de cruzamento foi definida em 95% e a mutação em 5%, além disso, o sistema possui um tamanho populacional de 100 indivíduos. Não há uma quantidade de gerações pré-definida, sendo assim, o critério de parada do AGMO ocorre quando o melhor indivíduo da população não evoluir por 100 gerações.

A acurácia do sistema é calculada pelo Erro Quadrático Médio (EQM) e a interpretabilidade é medida por $(1 - \text{quantidade de regras gerada pela otimização genética} / \text{número máximo de regras permitido})$. Este número máximo corresponde à quantidade de exemplos da base de dados. Isto porque, o algoritmo de Wang e Mendel gera uma regra para cada exemplo da base de dados.

RESULTADOS E/OU DISCUSSÃO

Nos experimentos foram utilizadas 15 bases de dados estratificadas em cinco *fold*s. Além disso, cada *fold* foi testado três vezes, logo, para cada base de dados foram efetuadas 15 execuções. Nas Tabelas 1 e 2, há colunas identificadas com **AVG** e **DIV**,

onde **AVG** indica que os resultados são a média das 15 execuções realizadas para cada base de dados e **DIV** representa o desvio padrão dos 15 resultados obtidos. A sigla **EQM** significa Erro Quadrático Médio e **INTER** corresponde à interpretabilidade.

A Tabela 1 contém os resultados do SBRF sem o aprendizado genético, ou seja, utilizando somente o algoritmo de Wang e Mendel para a construção da BR e adotando a distribuição uniforme dos conjuntos fuzzy.

Tabela 1. Resultados do SBRF sem aprendizado genético.

Base de dados	Treinamento		Teste	
	EQM-AVG	INTER-AVG	EQM-AVG	EQM-DIV
anacalt	0,4446	0,9777	0,4520	0,0083
autompg6	41,0936	0,8584	40,6813	6,6165
autompg8	30,9435	0,7710	32,7513	7,2754
concrete	189,0486	0,8369	195,4751	20,6550
dee	0,5954	0,7336	0,6149	0,0566
delta-ail	0,0000	0,9886	0,0000	0,0000
diabetes	0,3276	0,7845	1,1174	1,2468
forestFires	64519,6959	0,4154	66500,4920	16073,8811
friedman	16,4765	0,7752	17,1276	2,0159
laser	1704,4126	0,9778	1729,2362	156,2626
machine	54122,8379	0,8852	62978,6916	10280,0831
plastic	236,6667	0,9947	236,6667	6,2685
quake	0,0892	0,9891	0,0895	0,0160
stock	27,1964	0,8403	27,0736	0,8027
wankara	162,6298	0,8731	164,3374	7,6688
Média geral	8070,164	0,847	8794,987	1770,857

A Tabela 2 contém os resultados do SBRF com o aprendizado genético, ou seja, utilizando o algoritmo de Wang e Mendel para a geração da semente da BR, e adotando o NSGA-II para construção das regras e otimização das funções de pertinência simultaneamente.

Tabela 2. Resultados do SBRF com aprendizado genético.

Base de dados	Treinamento		Teste		GERAÇÃO
	EQM-AVG	INTER-AVG	EQM-AVG	EQM-DIV	
anacalt	0,4431	0,9852	0,3991	0,0003	146
autompg6	14,6418	0,9687	19,4649	4,0717	2196
autompg8	21,0410	0,9040	29,9027	11,5865	264
concrete	149,4358	0,9660	182,7684	16,5240	812
dee	0,2926	0,9623	0,5691	0,1903	1032
delta-ail	0,0000	0,9995	0,0000	0,0000	304
diabetes	0,3016	0,8721	1,9023	1,3766	939
forestFires	4342,2500	0,9927	947,5457	225,3570	138
friedman	15,4986	0,9000	17,6595	13,7021	110
laser	770,0814	0,9962	809,9165	117,3509	2801
machine	9603,7400	0,9700	18348,3503	11249,3397	3157

plastic	236,6667	0,9977	236,6667	6,2685	101
quake	0,0371	0,9963	0,0914	0,0446	110
stock	19,4913	0,9403	35,3369	13,0416	218
wankara	51,1340	0,9891	50,4329	6,9019	244
Média geral	1015,0037	0,9627	1378,7338	777,7170	838

Comparando os resultados da Tabela 1 com a Tabela 2, encontra-se um EQM aproximadamente oito vezes menor na fase de treinamento e seis vezes na fase de testes, quando aplicado o NSGA-II. Além disso, o índice de interpretabilidade do sistema foi melhor, 96,3% contra 84,7%.

CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo avaliar o desempenho de um SBRF com e sem aprendizado genético da BC para a resolução de problemas de regressão, com o objetivo de manter o balanceamento entre a precisão e interpretabilidade do sistema.

Os resultados obtidos das duas abordagens foram comparados e constatou-se que o aprendizado genético gerou um sistema mais preciso (com menor EQM) e mais interpretável, evidenciando que o aprendizado genético é recomendável, pois é possível obter um sistema fuzzy preciso e interpretável ao mesmo tempo.

REFERÊNCIAS

- Mendel, J. M. 1995. Fuzzy Logic Systems for Engineering: A Tutorial. IEEE Special Issue on Fuzzy Logic, vol.83, no.3, pp.345-377.
- Herrera, F. 2008. Genetic fuzzy systems: taxonomy, current research trends and prospects. Evolutionary Intelligence, vol.1, n.1, pp.27-46.
- Mitchell, M. 1996. An Introduction to Genetic Algorithms, The MIT Press.
- Ishibuchi, H., Murata, T. e Türksen, I. B. 1997. Single-Objective and Two-Objective Genetic Algorithms for Selecting Linguistic Rules for Pattern Classification Problems. Fuzzy Sets and Systems, vol.89, n.2, pp.135-150.
- Voget, S. e Kolonko, M. 1998. Multidimensional Optimization with a Fuzzy Genetic Algorithm. Journal of Heuristics, vol.4, n.3, pp.221-244.
- Ishibuchi, H., Nakashima, T. e Murata, T. 2001. Three-Objective Genetics-Based Machine Learning for Linguistic Rule Extraction. Information Sciences, vol.136, n.1-4, pp.109-133.
- Deb, K., A. Pratap, et al. 2002. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2): 182-197.
- Wang, L. X., Mendel, J. M. 1992. Generating Fuzzy Rules by Learning from Examples. IEEE Transaction on Systems, Man and Cybernetics, vol.22, pp.1414-1427.