

Relatório Técnico (50%)

Pad: Padrão (o trabalho contém todos os itens requeridos): 8,0

Ling: Linguagem (o texto está bem escrito, correto gramaticalmente e as ideias bem expressas): 8,0

Mod:Modelo (o modelo está bem descrito e sua instanciação feita adequadamente):7.0

Res: Resultados (os resultados foram devidamente descritos, inclusive utilizando gráficos): 9,0

Aval: Avaliação (os resultados foram adequadamente apresentados): 9,0

Implementação (50%)

Com: Comentários: (o código está devidamente documentado): 10,0

Cod: Codificação: (o código está bem escrito e roda adequadamente): 10,0

Result: Resultados: (os resultados foram produzidos adequadamente, inclusive utilizando gráficos): 10,0

Nota: 9,2

**Relatório Técnico
sobre Lógica Fuzzy**

*Thiago M. de Sousa Luana G. B. Martins
Ruan C. Rodrigues*

Technical Report - RT-INF_000-19 - Relatório Técnico
July - 2019 - Julho

The contents of this document are the sole responsibility of the authors.
O conteúdo do presente documento é de única responsabilidade dos autores.

Relatório Técnico sobre Lógica Fuzzy

Thiago M. de Sousa
thiagomontelesofc@gmail.com

Luana G. B. Martins
luanagbmartins@gmail.com

Ruan C. Rodrigues
ruanchaves93@gmail.com

Abstract. *This report describes what were the decisions, solving process and results of the proposed problem in the field of Artificial Intelligence. It was introduced the challenge to induce a fuzzy system that can classify the performance of the student during the course.*

Keywords: Technical Report, Fuzzy Logic, Fuzzy System.

Resumo. *Este relatório descreve quais foram as decisões, processo de resolução e resultados do problema proposto na matéria de Inteligência Artificial. Foi introduzido o desafio de induzir um Sistema Fuzzy que possa classificar o desempenho do aluno ao decorrer do curso.*

Palavras-Chave: Relatório Técnico, Lógica Fuzzy, Sistema Fuzzy.

1 Introdução

Este Relatório Técnico consiste na documentação de uma estratégia assumida para se obter um processo de tomada de decisões, buscando resolver o problema de se criar um modelo para obter classificações de desempenho período a período de novos alunos no Bacharelado em Ciência da Computação da Universidade Federal de Goiás, tendo em mãos apenas o histórico de disciplinas cursadas.

O processo de decisão se deu por meio da utilização de um Sistema *Fuzzy*, as quais por sua vez tiram proveito de uma estrutura lógica e informações probabilísticos que levam o processo à sua conclusão através de uma classificação .

O problema consiste em receber entradas relativas às médias, reprovações e trancamentos de cada aluno, para então criar um modelo que possa ser utilizado em sistema *Fuzzy*, com o objetivo final de gerar a classificação de desempenho de novos estudantes do curso.

No restante deste documento estão definidas a forma abordada na base de dados (Seção 2), descrição geral da solução proposta contendo nela a descrição geral do modelo utilizado e dos dados selecionados para o modelo (Seção 3), dos resultados obtidos (Seção 4), das propostas para como utilizar os resultados obtidos (Seção 5), conclusões finais (Seção 6) e referências.

2 Descrição da base de dados

Os dados consistem em um arquivo no formato csv (**Comma-separated values**) que é representado por uma matriz de 22361 linhas por 66 colunas, onde existe em cada coluna um determinado atributo referente a relação de um aluno com as disciplinas que cursou durante os anos e períodos.

	id	ano_nascimento_discente	idade_conclusao_ensino_medio	idade_ingresso_universidade	idade_colacao_grau	uf_naturalidade_discente
0	1	1989	17.0	19	26.0	GO
1	1	1989	17.0	19	26.0	GO
2	1	1989	17.0	19	26.0	GO
3	1	1989	17.0	19	26.0	GO
4	1	1989	17.0	19	26.0	GO

Os atributos nas colunas contém dados referentes ao aluno e sua passagem no curso. Dados como ano de nascimento, idade de ingresso à universidade e nota do Enem são exemplos de dados relacionados ao aluno anteriormente ao ingresso na faculdade. Já atributos como quantidade de reprovações, média global e ano de conclusão estão ligados ao aluno após a entrada na universidade.

3 Descrição da solução

Com o objetivo de permitir fazer uma classificação sobre o desempenho de alunos **bom base** no momento atual de integração do aluno, ~~foram considerados uma combinação~~ de regras com base na Lógica *Fuzzy* para induzir um Sistema *Fuzzy*.

3.1 Descrição do modelo utilizado

Nessa seção ~~será discutido os~~ princípios do funcionamento de uma Lógica *Fuzzy* baseado no Teorema *Fuzzy* e como foi aplicado como solução do problema proposto.

3.1.1 Lógica dos conjuntos Fuzzy

A necessidade de criar uma melhor representação das incertezas de experiências do mundo real proporcionou a criação de uma estratégia lógica que não representasse as situações somente em uma abordagem binária de verdadeiro ou falso ou até sim ou não. Por conta dessa necessidade, inicialmente foi introduzida em 1920 por Jan Lukasiewicz uma proposta de induzir conjuntos com graus de pertinência sendo 0, 1/2 e 1. Logo depois disso foi expandido para uma quantidade infinita entre 0 e 1. Anos depois foi a lógica *Fuzzy* foi formulada por Lofti A. Zadeh, combinando os conceitos de lógica clássica e os conjuntos de Lukasiewicz.

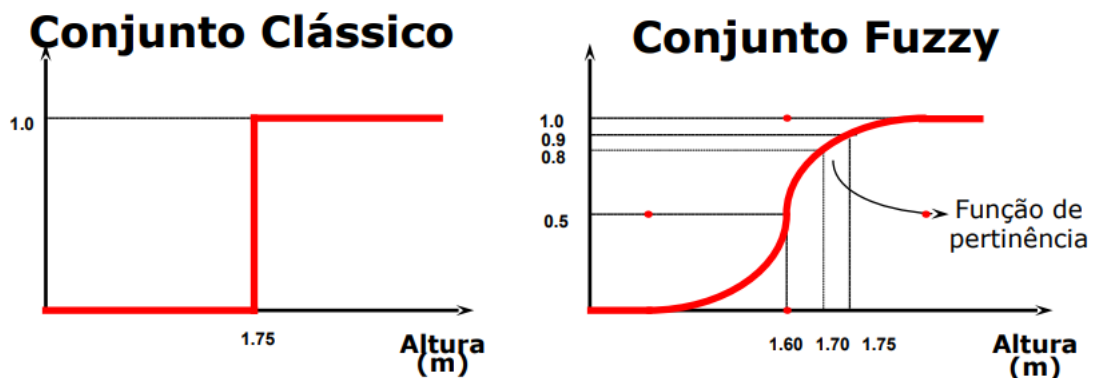
A Lógica *Fuzzy* é uma abordagem de multivalores capaz de capturar informações vagas, que em geral são descritas em uma linguagem natural, e convertê-las para um formato numérico, de fácil manipulação pelos sistemas computacionais atuais. A representação leva em conta não apenas o conceito, mas o contexto no qual ele está inserido. O objetivo da lógica *Fuzzy* é fazer com que as observações registradas pelo sistema se aproximem das observações humanas.

3.1.2 Classificação com lógica fuzzy

Para exemplificar no que consiste a classificação podemos pensar no seguinte caso: Como classificamos a altura de uma pessoa?

A lógica tradicional, por se basear em saídas binárias, teria uma classificação que informa somente se a pessoa é alta ou não é alta. Entretanto para o conjunto *fuzzy* é possível inserir variáveis linguísticas capazes de classificar em muito baixa, baixa, alta e muito alta, não estando também limitado a essas 4 denominações.

A = Conjunto de pessoas altas



Um conjunto *fuzzy* A definido no universo X é caracterizado por uma função de pertinência U_A , a qual mapeia elementos de X para um intervalo $[0,1]$. Desta forma, a função de pertinência associa a cada elemento y pertencente a X um número real entre $[0,1]$, que representa o grau de pertinência do elemento y ao conjunto A .

Definição formal: Um conjunto *fuzzy* A em X é expresso como um conjunto de pares ordenados:

$$A = \{(x, u_A(x)) | x \in X\} \quad (-1)$$

Sendo A o conjunto *fuzzy*, $u_A(x)$ função de pertinência e X o universo. A função de pertinência reflete o conhecimento que se tem em relação a intensidade com que o objeto pertence ao conjunto *fuzzy*.

3.2 Raciocínio fuzzy

O raciocínio para se aplicar a estratégia fuzzy consiste em 5 etapas:

1. Transformação de variáveis do problema: No primeiro passo, cada valor de entrada é associado a uma função de pertinência que resulta em um valor entre 0 a 1 a qual é definido o grau de pertinência do conjunto.
2. Aplicando os operadores: O segundo passo é aplicar os operadores fuzzy de maneira semelhante a lógica clássica, os quais para a lógica fuzzy são AND e OR. A partir da aplicação dos operadores são definidos o grau máximo e mínimo de pertinência do conjunto.
3. Aplicação dos operadores de implicação: Para que possa ser definido o peso no resultado final, a terceira etapa consiste na criação da hipótese de implicação.

4. Combinar todas as saídas fuzzy : No quarto passo é feita a combinação das saídas em um único conjunto fuzzy.
5. Transformar os resultados fuzzy em resultados nítidos: Esse processo é conhecido como defuzzificação, a qual consiste em retornar o resultado final na forma numérica dentro da faixa estimulada pela lógica fuzzy.

Frase longa e confusa.

3.3 Definição final do modelo

— Afim de obter um modelo que seja capaz de trabalhar com o problema proposto, foi utilizado, inicialmente, a criação de regras a serem aplicadas com base em observações dos dados selecionados, para obter uma melhor delimitação no desempenho do aluno até a integralização corrente. Foi definido como variáveis linguísticas as abstrações dos dados o conceito de um desempenho mau, um desempenho médio e um desempenho alto.

Para função de pertinência foi selecionado o modelo Triangular que consiste na seguinte função:

$$\text{trimf}(x : a, b, c) = \max(\min(\frac{x-a}{b-a}, \frac{c-x}{c-b}), 0) \quad (-2)$$

que conjunto?

Onde x é a variável pertencente ao universo e (a,b,c) um vetor de pertencente ao conjunto. A função de pertinência resultará em um valor que consiste no seguinte intervalo:

$$U_A(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{x-a}{b-a}, & \text{se } a < x \leq b \\ \frac{x-c}{b-c}, & \text{se } b \leq x < c \\ 0, & \text{se } x \geq c \end{cases}$$

A função sempre resultará entre o intervalo de 0 a 1 e seguirá a descrição exibido acima. Na etapa de defuzzificação a qual as regiões resultantes são convertidas em valores para a variável de saída espero, foi utilizado a técnica Centróide que consiste em:

- Caso discreto:

$$G(B) = \frac{\sum_{i=0}^n u_i \varphi_b(u_i)}{\sum_{i=0}^n \varphi_b(u_i)} \quad (-3)$$

frase incompleta

- Caso contínuo:

$$G(B) = \frac{\int_{\mathbb{R}} u \varphi_B(u) du}{\int_{\mathbb{R}} \varphi_B(u) du} \quad (-4)$$

O centróide ou centro de gravidade dá a média das áreas de todas as figuras que representam os graus de pertinência de um subconjunto fuzzy, sendo utilizado em caso discreto quando os dados são referentes a valores finitos ou quanto a quantidade é enumerável, já no caso que os valores podem assumir qualquer valor em um determinado intervalo contínuo.

3.4 Descrição dos dados selecionados

Os dados consistem na seleção de 9 atributos relacionado a passagem do aluno no curso, esses atributos consistem em:

- Média em disciplinas de programação

- Média em disciplinas básicas de matemática
- Média em disciplinas do núcleo específico
- Média em disciplinas do núcleo comum
- Quantidade de reprovações em disciplinas de programação
- Quantidade de reprovações em disciplinas básicas de matemática
- Quantidade de reprovações em disciplinas do núcleo específico
- Quantidade de reprovações em disciplinas do núcleo comum
- Quantidade de trancamentos

4 Resultados obtidos

Nessa seção, serão apresentados os resultados obtidos após implementação do modelo e o seu treinamento com os dados obtidos na fase de tratamento de dados.

4.1 Modelo gerado

como isso foi feito?

A partir dos atributos de entrada (médias, reprovações e trancamentos) tivemos como objetivo chegar, para cada aluno, a uma classificação entre 0 e 10, sendo 0 um aluno muito ruim, 5 o aluno mediano, e 10 a classificação do aluno muito bom.

Percebemos que a divisão em períodos seria inadequada, pois, dentro do universo de alunos, muitos bons alunos não cumprem com as disciplinas no período proposto, principalmente aqueles que ingressaram antes do ano de 2017. O curso permite tal flexibilidade na escolha das disciplinas a serem cursadas que não é possível penalizar um aluno simplesmente porque resolveu adiantar ou atrasar uma disciplina.

Portanto, resolvemos trabalhar com o conceito de **taxas de integralização**. A taxa de integralização é a porcentagem de matérias que já foi concluída pelo aluno. Sendo assim, dividimos o intervalo que vai de 0 a 100% em dez faixas, e elaboramos um modelo que torna possível comparar e acompanhar o desempenho de cada aluno progressivamente em cada uma destas dez faixas.

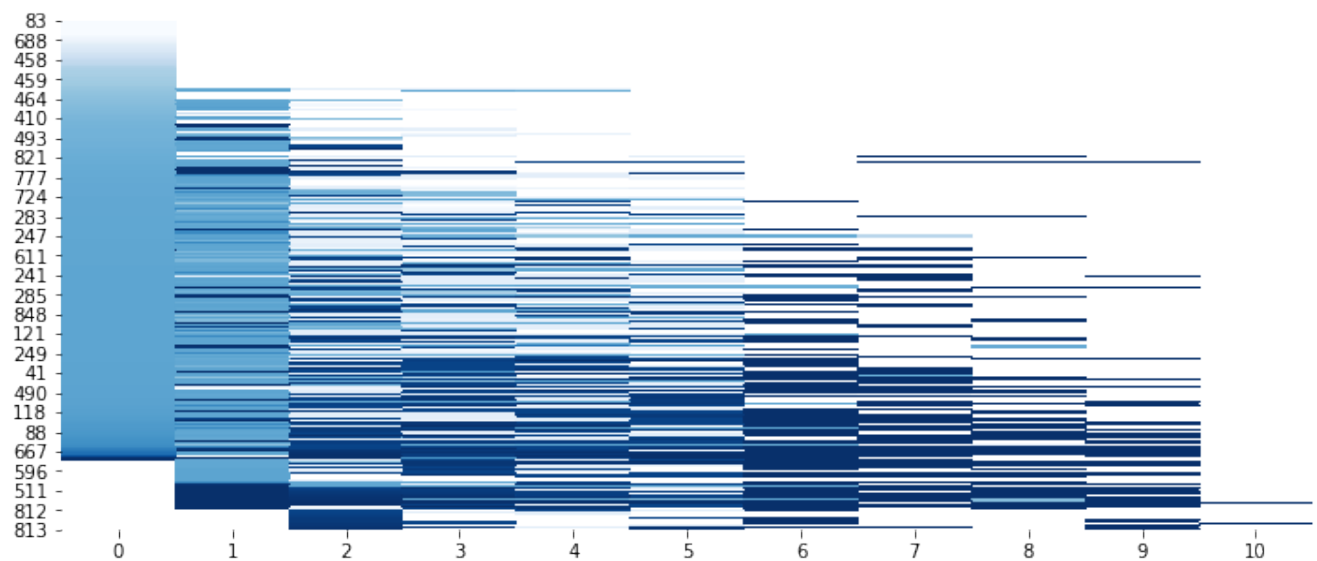
As entradas foram pré-processadas separadamente para cada faixa de integralização. Para a definição das regras, realizamos uma regressão linear dos valores de entrada dois a dois. Observando os pares de entradas que apresentaram um menor erro quadrático médio, formulamos regras a partir de nosso conhecimento empírico do problema usando estes pares de entradas.

Outro auxílio para a elaboração de regras foi a observação da média necessária para a aprovação (6.0) com cada um dos atributos relativos a média. Isso possibilitou a elaboração de regras simples do tipo "Se a média dos alunos neste atributo e nesta faixa de integralização está acima de seis, e o aluno está acima da média dos alunos, então ele é um bom aluno".

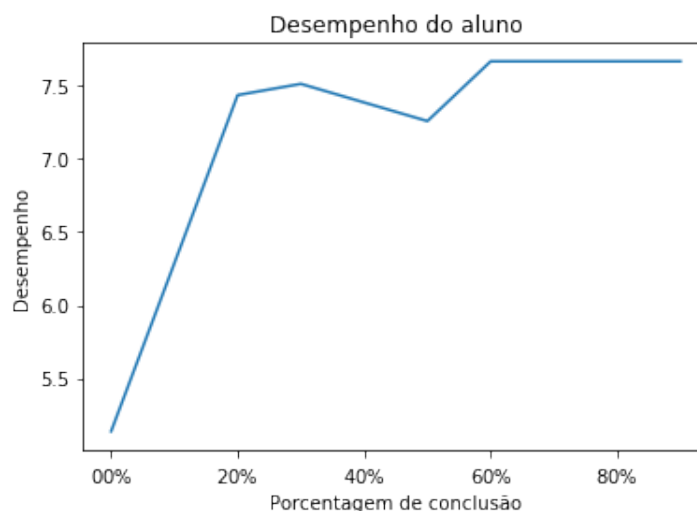
Gráfico das funções de pertinência???

5 Como a solução proposta pôde resolver o problema

A visualização gráfica dos resultados mostra que o nosso modelo foi capaz de categorizar adequadamente os alunos de acordo com o seu desempenho. O *heatmap* abaixo mostra cada linha como um aluno, e cada coluna é uma faixa da taxa de integralização (ou seja, a coluna 6 significa o desempenho dos alunos em cada linha quando começaram o semestre com taxa de integralização próxima de 60%). Uma tonalidade mais escura indica um bom valor de classificação, e uma tonalidade mais clara indica um valor pior. A tonalidade totalmente branca indica um valor ausente, e caso continue por várias colunas em uma mesma linha, pode indicar a desistência do aluno correspondente àquela linha.



Também foi possível observar o desempenho de um aluno individualmente. O gráfico abaixo corresponde a uma linha do *heatmap* acima, e mostra o desempenho de um único aluno. Ao invés de cores, usamos o eixo y para sinalizar variações no desempenho de um aluno em relação à uma classificação que varia de 0 a 10.



Como foi feita a fuzzificação e a defuzzificação?

Quais as regras utilizadas?

6 Conclusões

Percebemos que os modelos *fuzzy* permitem assimilar conhecimentos de especialistas em um sistema e, a partir de regras simples, ser capaz de gerar classificações de alta eficiência. Embora a alta dependência do conhecimento especializado do domínio possa ser uma inconveniência do modelo, ela é contrabalanceada por sua principal vantagem: a facilidade e eficácia na assimilação deste conhecimento para que ele seja aplicado à resolução de problemas de classificação.