

Introdução à Inteligência Artificial - Projeto 3

RAYSA MASSON BENATTI

I. INTRODUÇÃO

O presente trabalho apresenta a descrição das conclusões e estratégias adotadas para a resolução do Projeto 3 da disciplina Introdução à Inteligência Artificial (MC906), ministrada no primeiro semestre de 2019 na Unicamp pela professora Esther Luna Colombini.

Conforme a especificação, um sistema *fuzzy* deveria ser construído para solucionar determinado problema e avaliá-lo segundo parâmetros distintos. O objetivo do sistema implementado é prever uma probabilidade de admissão de estudantes no sistema de pós-graduação estadunidense, dados determinados parâmetros, comparando os resultados obtidos com valores de referência.

A seção II do relatório define o problema. A seção III dedica-se à descrição do sistema *fuzzy* escolhido para buscar soluções e especificidades da implementação, bem como as variações sobre os parâmetros adotadas. A seção IV descreve os resultados observados. Finalmente, na seção V, tais resultados são discutidos e conclusões são apresentadas. As referências encontram-se ao final do trabalho.

II. O PROBLEMA

Todos os anos, milhões de estudantes candidatam-se a programas de pós-graduação em universidades nos Estados Unidos [1]. A participação de estudantes internacionais dentre os matriculados cresceu significativamente: entre 2000 e 2015, o aumento foi de 35% [2]; no outono de 2016, estrangeiros eram cerca de 19% do total de alunos inscritos em programas de mestrado ou doutorado no país [1].

Para ser aceito em um desses programas, os critérios podem ser bastante rigorosos, em especial no caso de universidades competitivas. Para aplicantes internacionais, há critérios extras, como comprovação de proficiência na língua inglesa.

A partir da avaliação de alguns parâmetros disponíveis, é possível obter um resultado razoável sobre a probabilidade de um estudante estrangeiro ser aprovado. É o que propõem Mohan S. Acharya, Asfia Armaan e Aneeta S. Antony, autores do *dataset Graduate admissions: predicting admissions from important parameters*, baseado no *UCLA Graduate Dataset* [3]. O trabalho utiliza modelos de regressão linear para obter tais probabilidades, cujos valores são os resultados de referência. No presente trabalho, o sistema *fuzzy* implementado

obtem valores a ser comparados com os resultados em questão.

Os parâmetros considerados na análise são sete, cada um representado em uma coluna do *dataset*: nota obtida no GRE (*Graduate Record Examination*), teste de conhecimentos gerais usado como critério de admissão; nota obtida no exame TOEFL de proficiência na língua inglesa; classificação da universidade; classificação da robustez da carta de motivação (*Statement of Purpose*); classificação da robustez da carta de recomendação (*Letter of Recommendation*); GPA (*Grade Point Average*), média das notas obtidas pelo estudante em seu curso de graduação; experiência de pesquisa.

Assim, o *dataset* tem nove colunas; aquelas que representam os parâmetros descritos acima são precedidas pela dos índices — números sequenciais de 1 a 400 (quantidade de entradas) — e seguidas pela dos resultados (probabilidade de admissão).

III. SOLUÇÃO COM UM SISTEMA FUZZY

A implementação do sistema *fuzzy* foi baseada em uma solução para o canônico problema da gorjeta, conforme descrito em [4] e [5]. Python é a linguagem adotada nessa abordagem, com auxílio das bibliotecas *Scikit-Fuzzy*, cuja documentação encontra-se em [6] e [7] e *NumPy*.

Para tal, foi feito *download* do *dataset* [3] como um arquivo CSV (*comma-separated values*), o qual é lido e armazenado na variável do tipo *DataFrame* *dataset*. Duas colunas foram adicionadas à estrutura: *Fuzzy chance of admit*, para armazenar os resultados obtidos pelo sistema, e *Difference*, para armazenar a diferença entre tais valores e os valores de referência, informação útil para compará-los e avaliar a qualidade da solução.

A. Variáveis de entrada e saída

Em sua configuração completa, o sistema tem sete **entradas**, cada uma das quais representando os sete parâmetros de admissão descritos na seção anterior. Elas são inseridas através do método *Antecedent* e são descritas conforme segue:

- *gre*: nota obtida no GRE, de 0 a 340, com passo 1;
- *uRating*: classificação da universidade, de 1 a 5, com passo 1;
- *gpa*: GPA obtido, de 0 a 10, com passo 0.01;

- `toefl`: nota obtida no TOEFL, de 0 a 120, com passo 1;
- `sop`: robustez da carta de motivação, de 1 a 5, com passo 0.5;
- `lor`: robustez da carta de recomendação, de 1 a 5, com passo 0.5;
- `rExp`: experiência de pesquisa, valor binário que assume 1 em caso de existência de 0 caso contrário.

No contexto deste trabalho, não é relevante discutir como cada um desses valores é obtido, e sim como eles influenciam o resultado, representado pela variável de **saída** `chance`, valor que varia de 0 a 1 com passo 0.01. O método `Consequent` é usado para registrá-la.

B. Funções de pertinência e modelo de inferência

Conforme descrito em [6] e [7], as funções da biblioteca `Scikit-Fuzzy` utilizam o modelo de inferência de Mamdani para cômputo dos valores a partir das regras e variáveis de entrada fornecidas ao sistema.

Para definir a pertinência, foram usadas funções triangulares para as variáveis `rExp`, `uRating`, `sop` e `lor`. No caso das três últimas, a construção da função é automática com o método `automf(3)`, que desenha, para cada uma delas, três curvas triangulares de pertinência: uma para valores baixos (*poor*), uma para médios (*average*) e uma para altos (*good*). A Figura 1 ilustra o comportamento da função para a variável `uRating`; os gráficos para `sop` e `lor` são similares.

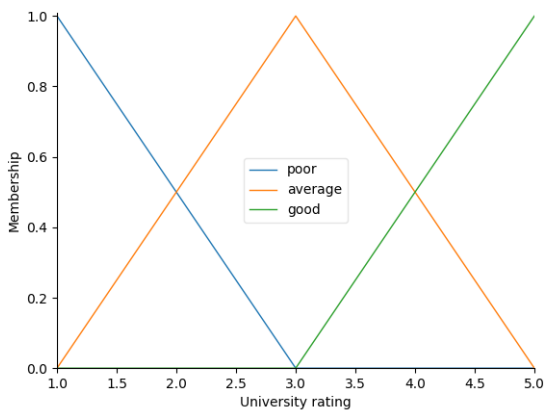


Figura 1. Função de pertinência de `uRating`.

Para `rExp`, a função foi customizada com o método `trimf`; por se tratar de valor binário, duas retas bastam para definir o comportamento desejado, conforme ilustrado pela Figura 2.

Para as demais variáveis de entrada — `gre`, `gpa`, `toefl` — e para a variável de saída `chance`, o código

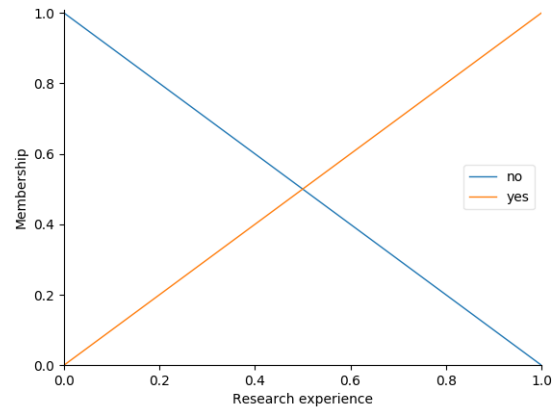


Figura 2. Função de pertinência de `rExp`.

define dois métodos, a ser usados alternativamente, que definem suas funções de pertinência: `tri()` constrói funções triangulares com `automf(3)`, de modo similar ao descrito acima, e `gauss()` constrói funções gaussianas customizadas, invocando `gaussmf`, também definindo os níveis de classificação *poor*, *average* e *good* para as variáveis. As Figuras 3 e 4 ilustram tais funções no caso da variável `gre`, a título de exemplo.

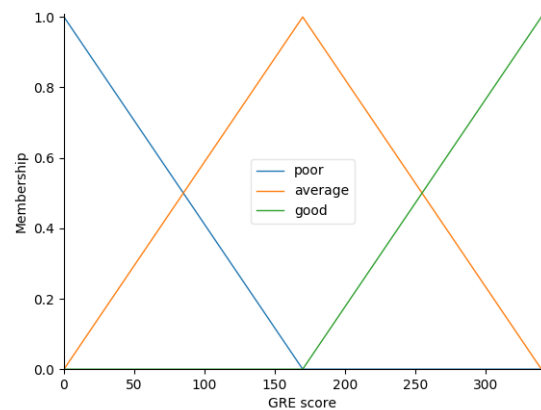


Figura 3. Função triangular de pertinência para os valores de `GRE`.

Assim, ao invocar o método `tri()`, o sistema utiliza funções de pertinência triangulares para essas variáveis; caso se queira, em vez disso, utilizar funções gaussianas, deve-se invocar `gauss()`. Isso permite a execução em diferentes configurações, de modo que se comparem os resultados, conforme será detalhado em subseção adiante.

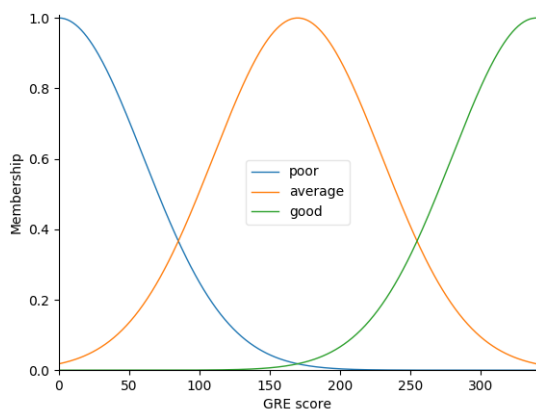


Figura 4. Função gaussiana de pertinência para os valores de GRE.

C. Sistema de regras

Na configuração completa, adota-se um sistema de quatro regras:

- **Regra 1:** se (gre é bom) e (gpa é bom) então (chance é boa);
- **Regra 2:** se (gre é ruim) ou (gpa é ruim) ou (não há rExp) então (chance é ruim);
- **Regra 3:** se (uRating é ruim) então (chance é boa);
- **Regra 4:** se (toefl é ruim) ou (sop é ruim) ou (lor é ruim) então (chance é ruim).

A escolha por esse conjunto de regras segue a intuição de que a nota obtida no GRE e o valor do GPA do aluno seriam os fatores mais determinantes em uma seleção dessa natureza. Assim, caso tais valores sejam altos, as chances de admissão aumentam consideravelmente; havendo mau desempenho em qualquer uma dessas métricas, há penalização. A experiência de pesquisa não é, nesse sistema, compreendida como determinante para aprovação, mas sua ausência também prejudica o estudante.

A classificação da universidade é introduzida como um "fator-bônus" para o aluno ao aplicar para instituições de menor prestígio acadêmico. Interpreta-se que, nesse caso, a competitividade será menor e isso por si só impulsionará a probabilidade de admissão.

Por fim, toefl, sop e lor são compreendidos como parâmetros mais eliminatórios que classificatórios: nessa perspectiva, comprovar boa proficiência em inglês e apresentar boas cartas de motivação e recomendação não diferenciam significativamente um candidato, servindo primordialmente como um filtro na seleção. O operador **ou** foi escolhido nesse caso pois um valor baixo em qualquer deles é prejudicial, de modo similar ao que ocorre na regra 2; assim, faz sentido agrupá-los em uma mesma regra.

D. Validação

A validação dos resultados obtidos por esse sistema fuzzy é feita ao compará-los com os valores de referência do *dataset* original. Para tanto, computa-se a diferença entre cada um desses valores.

O valor da menor e da maior diferença são especialmente armazenados nas variáveis `min_difference` e `max_difference`, as quais são utilizadas, em seguida, para recuperar do *dataset* as linhas que apresentam esses valores. Tais linhas são armazenadas nos *dataframes* `bestResults` e `worstResults`, respectivamente, o que permite analisar as características das entradas que apresentaram esses resultados de destaque.

Além disso, calcula-se o valor do erro médio absoluto (*mean absolute error*) entre os resultados de referência e os resultados fuzzy, segundo a fórmula $\frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$, em que: n = quantidade de entradas do *dataset* (400 no nosso caso); y = valor obtido pelo sistema fuzzy; \hat{y} = valor de referência, oriundo do *dataset* original [3].

E. Variações

Para a análise do presente trabalho, o sistema foi executado em cinco configurações distintas:

- **Configuração 1:** Conjunto completo de regras / funções de pertinência triangulares;
- **Configuração 2:** Conjunto de regras sem **regra 2** / funções de pertinência triangulares;
- **Configuração 3:** Conjunto de regras sem **regra 3** / funções de pertinência triangulares;
- **Configuração 4:** Conjunto de regras sem **regra 4** / funções de pertinência triangulares;
- **Configuração 5:** Conjunto de regras idêntico ao utilizado pela configuração, dentre as anteriores, que apresentou menor erro médio absoluto / funções de pertinência gaussianas.

O objetivo das quatro primeiras configurações é verificar a influência de cada regra no resultado final; para tanto, é necessário manter o restante do sistema idêntico, razão pela qual optou-se por não adotar variações nas funções de pertinência nesses casos.

Em posse da informação sobre qual dessas quatro configurações traz melhores resultados — o que é traduzido, no nosso caso, por um menor erro médio absoluto —, o mesmo conjunto de regras é aplicado à última configuração, para comparar o comportamento do sistema com as funções de pertinência triangulares e gaussianas. Destaca-se que as funções triangulares são substituídas por gaussianas apenas nas variáveis em que há pertinência para tal, conforme

descrito na seção III-B.

Convém mencionar, ainda, que uma configuração em que apenas a primeira regra é eliminada não é aceita pelo sistema, que levanta uma exceção "*Crisp output cannot be calculated, likely because the system is too sparse*" quando isso ocorre.

IV. RESULTADOS

Para cada uma das configurações, são exibidas três tabelas: a primeira contém as cinco primeiras entradas do *dataset* após a execução do sistema; a segunda mostra os melhores resultados (diferenças mínimas), limitados a cinco entradas; a terceira mostra os piores resultados (diferenças máximas), com a mesma limitação. (Os *datasets* resultantes completos encontram-se nos arquivos anexos a este relatório.) O valor do erro médio absoluto também é registrado, arredondado para quatro casas decimais.

Cabe também observar que o índice mostrado nas tabelas de melhores e piores resultados é o *serial number* da entrada em questão no *dataset* original, e não o índice propriamente dito, que é sempre sequencial.

A. Configuração 1

Tabela I
RESULTADOS GERAIS DA CONFIGURAÇÃO 1

| Index | 1 | 2 | 3 | 4 | 5 |
|--------------|------|------|------|------|------|
| GRE | 337 | 324 | 316 | 322 | 314 |
| TOEFL | 118 | 107 | 104 | 110 | 103 |
| U. Rating | 4 | 4 | 3 | 3 | 2 |
| SOP | 4.5 | 4 | 3 | 3.5 | 2 |
| LOR | 4.5 | 4.5 | 3.5 | 2.5 | 3 |
| GPA | 9.65 | 8.87 | 8 | 8.67 | 8.21 |
| Research | 1 | 1 | 1 | 1 | 0 |
| Chance | 0.92 | 0.76 | 0.72 | 0.8 | 0.65 |
| Fuzzy chance | 0.91 | 0.9 | 0.89 | 0.69 | 0.51 |
| Difference | 0.01 | 0.14 | 0.17 | 0.11 | 0.14 |

Tabela II
CONFIGURAÇÃO 1: MELHORES RESULTADOS

| Index | 6 | 33 | 45 | 99 | 108 |
|--------------|------|------|------|------|------|
| GRE | 330 | 338 | 326 | 332 | 338 |
| TOEFL | 115 | 118 | 113 | 119 | 117 |
| U. Rating | 5 | 4 | 5 | 4 | 4 |
| SOP | 4.5 | 3 | 4.5 | 5 | 3.5 |
| LOR | 3 | 4.5 | 4 | 4.5 | 4.5 |
| GPA | 9.34 | 9.4 | 9.4 | 9.24 | 9.46 |
| Research | 1 | 1 | 1 | 1 | 1 |
| Chance | 0.9 | 0.91 | 0.91 | 0.9 | 0.91 |
| Fuzzy chance | 0.9 | 0.91 | 0.91 | 0.9 | 0.91 |
| Difference | 0 | 0 | 0 | 0 | 0 |

Tabela III
CONFIGURAÇÃO 1: PIORES RESULTADOS

| Index | 94 |
|--------------|------|
| GRE | 301 |
| TOEFL | 97 |
| U. Rating | 2 |
| SOP | 3 |
| LOR | 3 |
| GPA | 7.88 |
| Research | 1 |
| Chance | 0.44 |
| Fuzzy chance | 0.88 |
| Difference | 0.44 |

Diferença mínima: 0

Quantidade de melhores resultados: 12

Diferença máxima: 0.44

Quantidade de piores resultados: 1

Erro médio absoluto: 0.1161

B. Configuração 2

Tabela IV
RESULTADOS GERAIS DA CONFIGURAÇÃO 2

| Index | 1 | 2 | 3 | 4 | 5 |
|--------------|------|------|------|------|------|
| GRE | 337 | 324 | 316 | 322 | 314 |
| TOEFL | 118 | 107 | 104 | 110 | 103 |
| U. Rating | 4 | 4 | 3 | 3 | 2 |
| SOP | 4.5 | 4 | 3 | 3.5 | 2 |
| LOR | 4.5 | 4.5 | 3.5 | 2.5 | 3 |
| GPA | 9.65 | 8.87 | 8 | 8.67 | 8.21 |
| Research | 1 | 1 | 1 | 1 | 0 |
| Chance | 0.92 | 0.76 | 0.72 | 0.8 | 0.65 |
| Fuzzy chance | 0.91 | 0.9 | 0.89 | 0.69 | 0.57 |
| Difference | 0.01 | 0.14 | 0.17 | 0.11 | 0.08 |

Tabela V
CONFIGURAÇÃO 2: MELHORES RESULTADOS

| Index | 6 | 33 | 45 | 99 | 108 |
|--------------|------|------|------|------|------|
| GRE | 330 | 338 | 326 | 332 | 338 |
| TOEFL | 115 | 118 | 113 | 119 | 117 |
| U. Rating | 5 | 4 | 5 | 4 | 4 |
| SOP | 4.5 | 3 | 4.5 | 5 | 3.5 |
| LOR | 3 | 4.5 | 4 | 4.5 | 4.5 |
| GPA | 9.34 | 9.4 | 9.4 | 9.24 | 9.46 |
| Research | 1 | 1 | 1 | 1 | 1 |
| Chance | 0.9 | 0.91 | 0.91 | 0.9 | 0.91 |
| Fuzzy chance | 0.9 | 0.91 | 0.91 | 0.9 | 0.91 |
| Difference | 0 | 0 | 0 | 0 | 0 |

Tabela VI
CONFIGURAÇÃO 2: PIORES RESULTADOS

| Index | 93 |
|--------------|------|
| GRE | 298 |
| TOEFL | 98 |
| U. Rating | 2 |
| SOP | 4 |
| LOR | 3 |
| GPA | 8.03 |
| Research | 0 |
| Chance | 0.34 |
| Fuzzy chance | 0.89 |
| Difference | 0.44 |

Diferença mínima: 0
Quantidade de melhores resultados: 12
Diferença máxima: 0.55
Quantidade de piores resultados: 1
Erro médio absoluto: 0.1144

C. Configuração 3

Tabela VII
RESULTADOS GERAIS DA CONFIGURAÇÃO 3

| Index | 1 | 2 | 3 | 4 | 5 |
|--------------|------|------|------|------|------|
| GRE | 337 | 324 | 316 | 322 | 314 |
| TOEFL | 118 | 107 | 104 | 110 | 103 |
| U. Rating | 4 | 4 | 3 | 3 | 2 |
| SOP | 4.5 | 4 | 3 | 3.5 | 2 |
| LOR | 4.5 | 4.5 | 3.5 | 2.5 | 3 |
| GPA | 9.65 | 8.87 | 8 | 8.67 | 8.21 |
| Research | 1 | 1 | 1 | 1 | 0 |
| Chance | 0.92 | 0.76 | 0.72 | 0.8 | 0.65 |
| Fuzzy chance | 0.91 | 0.9 | 0.89 | 0.69 | 0.51 |
| Difference | 0.01 | 0.14 | 0.17 | 0.11 | 0.14 |

Tabela VIII
CONFIGURAÇÃO 3: MELHORES RESULTADOS

| Index | 6 | 9 | 29 | 33 | 45 |
|--------------|------|-----|------|------|------|
| GRE | 330 | 302 | 295 | 338 | 326 |
| TOEFL | 115 | 102 | 93 | 118 | 113 |
| U. Rating | 5 | 1 | 1 | 4 | 5 |
| SOP | 4.5 | 2 | 2 | 3 | 4.5 |
| LOR | 3 | 1.5 | 2 | 4.5 | 4 |
| GPA | 9.34 | 8 | 7.2 | 9.4 | 9.4 |
| Research | 1 | 0 | 0 | 1 | 1 |
| Chance | 0.9 | 0.5 | 0.46 | 0.91 | 0.91 |
| Fuzzy chance | 0.9 | 0.5 | 0.46 | 0.91 | 0.91 |
| Difference | 0 | 0 | 0 | 0 | 0 |

Tabela IX
CONFIGURAÇÃO 3: PIORES RESULTADOS

| Index | 94 |
|--------------|------|
| GRE | 301 |
| TOEFL | 97 |
| U. Rating | 2 |
| SOP | 3 |
| LOR | 3 |
| GPA | 7.88 |
| Research | 1 |
| Chance | 0.44 |
| Fuzzy chance | 0.88 |
| Difference | 0.44 |

Diferença mínima: 0
Quantidade de melhores resultados: 15
Diferença máxima: 0.44
Quantidade de piores resultados: 1
Erro médio absoluto: 0.1162

D. Configuração 4

Tabela X
RESULTADOS GERAIS DA CONFIGURAÇÃO 4

| Index | 1 | 2 | 3 | 4 | 5 |
|--------------|------|------|------|------|------|
| GRE | 337 | 324 | 316 | 322 | 314 |
| TOEFL | 118 | 107 | 104 | 110 | 103 |
| U. Rating | 4 | 4 | 3 | 3 | 2 |
| SOP | 4.5 | 4 | 3 | 3.5 | 2 |
| LOR | 4.5 | 4.5 | 3.5 | 2.5 | 3 |
| GPA | 9.65 | 8.87 | 8 | 8.67 | 8.21 |
| Research | 1 | 1 | 1 | 1 | 0 |
| Chance | 0.92 | 0.76 | 0.72 | 0.8 | 0.65 |
| Fuzzy chance | 0.91 | 0.9 | 0.89 | 0.9 | 0.51 |
| Difference | 0.01 | 0.14 | 0.17 | 0.09 | 0.14 |

Tabela XI
CONFIGURAÇÃO 4: MELHORES RESULTADOS

| Index | 6 | 33 | 45 | 99 | 108 |
|--------------|------|------|------|------|------|
| GRE | 330 | 338 | 326 | 332 | 338 |
| TOEFL | 115 | 118 | 113 | 119 | 117 |
| U. Rating | 5 | 4 | 5 | 4 | 4 |
| SOP | 4.5 | 3 | 4.5 | 5 | 3.5 |
| LOR | 3 | 4.5 | 4 | 4.5 | 4.5 |
| GPA | 9.34 | 9.4 | 9.4 | 9.24 | 9.46 |
| Research | 1 | 1 | 1 | 1 | 1 |
| Chance | 0.9 | 0.91 | 0.91 | 0.9 | 0.91 |
| Fuzzy chance | 0.9 | 0.91 | 0.91 | 0.9 | 0.91 |
| Difference | 0 | 0 | 0 | 0 | 0 |

Tabela XII
CONFIGURAÇÃO 4: PIORES RESULTADOS

| Index | 59 |
|--------------|------|
| GRE | 300 |
| TOEFL | 99 |
| U. Rating | 1 |
| SOP | 3 |
| LOR | 2 |
| GPA | 6.8 |
| Research | 1 |
| Chance | 0.36 |
| Fuzzy chance | 0.91 |
| Difference | 0.55 |

Diferença mínima: 0
Quantidade de melhores resultados: 12
Diferença máxima: 0.55
Quantidade de piores resultados: 1
Erro médio absoluto: 0.1283

E. Configuração 5

Tabela XIII
RESULTADOS GERAIS DA CONFIGURAÇÃO 5

| Index | 1 | 2 | 3 | 4 | 5 |
|--------------|------|------|------|------|------|
| GRE | 337 | 324 | 316 | 322 | 314 |
| TOEFL | 118 | 107 | 104 | 110 | 103 |
| U. Rating | 4 | 4 | 3 | 3 | 2 |
| SOP | 4.5 | 4 | 3 | 3.5 | 2 |
| LOR | 4.5 | 4.5 | 3.5 | 2.5 | 3 |
| GPA | 9.65 | 8.87 | 8 | 8.67 | 8.21 |
| Research | 1 | 1 | 1 | 1 | 0 |
| Chance | 0.92 | 0.76 | 0.72 | 0.8 | 0.65 |
| Fuzzy chance | 0.89 | 0.89 | 0.87 | 0.71 | 0.61 |
| Difference | 0.03 | 0.13 | 0.15 | 0.09 | 0.04 |

Tabela XIV
CONFIGURAÇÃO 5: MELHORES RESULTADOS

| Index | 38 | 48 | 135 | 171 | 172 |
|--------------|------|------|------|------|------|
| GRE | 300 | 339 | 333 | 312 | 334 |
| TOEFL | 105 | 119 | 113 | 101 | 117 |
| U. Rating | 1 | 5 | 5 | 2 | 5 |
| SOP | 1 | 4.5 | 4 | 2.5 | 4 |
| LOR | 2 | 4 | 4 | 3.5 | 4.5 |
| GPA | 7.8 | 9.7 | 9.28 | 8.04 | 9.07 |
| Research | 0 | 0 | 1 | 1 | 1 |
| Chance | 0.58 | 0.89 | 0.89 | 0.68 | 0.89 |
| Fuzzy chance | 0.58 | 0.89 | 0.89 | 0.68 | 0.89 |
| Difference | 0 | 0 | 0 | 0 | 0 |

Tabela XV
CONFIGURAÇÃO 5: PIORES RESULTADOS

| Index | 93 |
|--------------|------|
| GRE | 298 |
| TOEFL | 98 |
| U. Rating | 2 |
| SOP | 4 |
| LOR | 3 |
| GPA | 8.03 |
| Research | 0 |
| Chance | 0.34 |
| Fuzzy chance | 0.87 |
| Difference | 0.53 |

Diferença mínima: 0

Quantidade de melhores resultados: 19

Diferença máxima: 0.53

Quantidade de piores resultados: 1

Erro médio absoluto: 0.1071

V. DISCUSSÃO E CONCLUSÕES

De modo geral, o trabalho realizado mostrou que a abordagem *fuzzy* provê soluções razoáveis para o problema descrito, com ressalvas.

Parece ser acertada a intuição segundo a qual as notas do GRE e de GPA são os parâmetros mais relevantes para admissão. Essa tendência está presente no *dataset* original e se mantém na solução *fuzzy*, que sequer admite a execução

do sistema sem esses parâmetros, alegando esparsidade.

Quanto aos demais parâmetros, na modelagem adotada, a experiência de pesquisa (*rExp*) e a classificação da universidade (*uRating*) não mostraram grande impacto nos resultados. No caso da primeira, a supressão da penalização por sua ausência não apenas não refinou as previsões como pareceu piorá-las; nas entradas em que *rExp* = 1, naturalmente, não houve alterações. A retirada do fator-bônus associado a *uRating*, por sua vez, se traduziu em um mínimo aumento do erro absoluto em relação à referência (configuração 1) e na obtenção de alguns poucos resultados ótimos a mais, porém, globalmente, os resultados não foram afetados.

A penalização associada a valores baixos de *toefl*, *sop* e *lor*, representada pela regra 4, suprimida na quarta configuração, parece fazer sentido. Globalmente, essa configuração obteve erro médio absoluto ligeiramente maior que a configuração de referência; além disso, os resultados ótimos concentram entradas com altos valores associados a essas variáveis, conforme esperado. O pior resultado obtido também ilustra essa conclusão, dado que, conforme exibido na Tabela XII, trata-se de entrada com tais valores baixos. Assim, ao não aplicar a penalização, a previsão ficou prejudicada especialmente nesses casos.

Dentre as quatro primeiras configurações, a que obteve melhor resultado global foi aquela em que apenas a regra 2 foi suprimida, o que reforça as conclusões mencionadas acima. Ao executar a quinta configuração, com as funções de pertinência gaussianas quando aplicável, o sistema apresentou resultados ligeiramente mais refinados, sugerindo melhor adequação da gaussiana para modelar tais funções. A alteração nas funções de pertinência foi a variação que mais impactou o valor do erro médio absoluto.

Destaca-se, ainda, uma incompatibilidade entre os valores do domínio — usados para alimentar o sistema *fuzzy* — e os valores efetivamente encontrados no *dataset* original para as variáveis *toefl*, *gre* e *gpa*. Embora a todas elas possam ser atribuídos valores a partir de zero, o conjunto de entradas não tem valores abaixo de 92 para o TOEFL, 290 para o GRE e 6.8 para o GPA. Trata-se de uma limitação da solução, o que pode comprometer a análise.

De modo geral, os resultados apresentados pelo sistema são compatíveis com o que se esperava, mantendo-se dentro de margens de erro de até 15% — considerando-se que, nas configurações testadas, o maior erro médio absoluto observado foi de 0.1283. Em todos os casos, houve erros altos, em torno de 50%; contudo, essa não parece ser uma tendência, visto que os erros médios absolutos não chegam perto desse valor. Assim, caso a aplicação a utilizar tal solução exija precisão e seja sensível a erros, o sistema *fuzzy* modelado conforme descrito não seria a opção ideal;

no entanto, para aplicações que tolerem os graus de erro mencionados, a solução seria aceitável — por exemplo, para que se pudesse construir uma interface de banco de dados em que as buscas considerassem os resultados mostrados.

REFERÊNCIAS

- [1] H. Okahana and E. Zhou, *Graduate Enrollment and Degrees: 2006 to 2016*. Council of Graduate Schools, 2017. [Online]. Available: <https://bit.ly/2HjxFDT> 1
- [2] National Science Board, *Graduate Education, Enrollment, and Degrees in the United States*. Science and Engineering Indicators 2018, 2018. [Online]. Available: <https://bit.ly/2GswG1U> 1
- [3] M. S. Acharya, A. Armaan, and A. S. Antony, “Graduate Admissions - Predicting admission from important parameters,” 2018. [Online]. Available: <https://bit.ly/2KDNa9O> 1, 3
- [4] Scikit-Fuzzy development team, “Fuzzy Control Systems: The Tipping Problem.” [Online]. Available: <https://bit.ly/2MMtsLC> 1
- [5] —, “Fuzzy Control Primer: Overview and Terminology.” [Online]. Available: <https://bit.ly/2Kpmr1t> 1
- [6] —, “Scikit-fuzzy API reference.” [Online]. Available: <https://bit.ly/2WFmQhF> 1, 2
- [7] —, “skfuzzy - Fuzzy Logic Toolbox for Python.” [Online]. Available: <https://bit.ly/2ILdNXP> 1, 2