

COMPORTAMENTO DE REDES NEURAIS ARTIFICIAIS MLP PARA CLASSIFICAÇÃO DE QUESTÕES DO ENEM BASEADO EM TRI

RESUMO

O Exame Nacional do Ensino Médio (ENEM) ganha a cada ano mais importância, substituindo vestibulares tradicionais. Com isso, o desenvolvimento de ferramentas que auxiliem a preparação dos estudantes é de grande importância. Muitos simulados são feitos de forma praticamente aleatória, com questões escolhidas sem critério. Neste contexto, este trabalho apresenta o desenvolvimento de uma rede neural para classificação de questões do ENEM, com base nos parâmetros de Teoria de Resposta ao Item. Para isso, será utilizado o algoritmo *k-means* para gerar um *dataset* contendo a classificação dessas questões em 3 grupos: Fáceis, Médias e Difíceis. Com esse *dataset* uma rede neural será treinada para realizar a classificação das questões e verificar os resultados. Dessa maneira esse estudo contribui permitindo a elaboração de simulados de forma balanceada. Além disso, após o treinamento da rede a mesma será capaz de realizar a classificação de novas questões com base no TRI. Caso o *k-means* continuasse a ser utilizado a classificação teria que ser feita novamente caso novas questões fossem adicionadas a base de questões. A rede desenvolvida teve praticamente o mesmo desempenho que o *k-means* na tarefa de classificação, isso mostra a capacidade que as redes neurais possuem de classificação.

Palavras-chave: Enem. Algoritmos de classificação. Teoria de Resposta ao Item.

1 INTRODUÇÃO

Neste capítulo será apresentada uma introdução a assuntos que são relevantes para o entendimento do trabalho proposto. A ordem de apresentação do capítulo será a seguinte: Contextualização, Justificativa, Objetivos e Contribuições.

1.1 Contextualização

Em 2009 o ENEM adotou o sistema de Teoria de Resposta ao Item (TRI) que substituiu a Teoria Clássica dos Testes (TCT), o qual considerava somente a quantidade de acertos e erros para o cálculo da nota. Com o passar dos anos o ENEM ganhou espaço em meio aos tradicionais vestibulares que eram aplicados por cada instituição. Ainda assim, atualmente existem poucas ferramentas que utilizam TRI para auxiliar na preparação dos estudantes, por exemplo, grande parte dos simulados feitos em escolas e cursinhos preparatórios para o ENEM são feitos por professores que montam as provas escolhendo questões de maneira aleatória. Entretanto, essa não é a maneira ideal de se montar um caderno de provas, já que os parâmetros do TRI de cada questão podem ser calculados com base nos dados de acertos e erros referentes a questão. E com base nesses parâmetros é possível classificar as questões em grupos de dificuldade, possibilitando a geração de simulados de forma inteligente.

35 O presente trabalho é apresentado neste contexto, onde a construção dos cadernos de
36 questões para simulações do ENEM é feita de forma balanceada, baseando-se nos parâmetros
37 de cada questão e não aleatoriamente. Na próxima seção é abordada a justificativa da pesquisa.

38 **1.2 Justificativa e proposta**

39 O ENEM é utilizado por 1434 instituições de ensino superior no Brasil para selecionar
40 os candidatos às vagas, sendo o maior vestibular do país (MATIAS; TOLEDO, 2016). Diante
41 disso, o desenvolvimento de uma ferramenta dotada dos recursos oferecidos por redes neurais
42 para classificar questões em grupos de dificuldade é um grande passo para evolução dos métodos
43 de ensino e de métodos de avaliação que se baseiam no TRI. Dessa maneira, estudantes podem
44 fazer sua preparação de forma orientada, tendo a possibilidade de iniciar seus estudos com
45 questões de dificuldade mais baixa e conforme forem evoluindo com os estudos elevem os
46 níveis de dificuldade das questões.

47 **1.3 Objetivos**

48 Este estudo propõe atender ao objetivo geral e aos objetivos específicos descritos a se-
49 guir.

50 **1.3.1 Objetivo Geral**

51 Propor o desenvolvimento de uma rede neural que seja capaz de classificar questões com
52 base nos parâmetros de Teoria de Resposta ao Item.

53 **1.3.2 Objetivos específicos**

54 São objetivos específicos deste trabalho:

- 55 1. Possibilitar a geração simulados com base na dificuldade das questões e não aleatoria-
56 mente.
- 57 2. Verificar a capacidade das redes neurais em classificação de questões com base nos parâ-
58 metros de Teoria de Resposta ao Item.

59 **1.4 Contribuições**

60 Espera-se que com o desenvolvimento do trabalho proposto que simulados possam ser
61 gerados com base nos grupos de dificuldade das questões e não de forma aleatória. Com isso,
62 é possível por exemplo avaliar se alunos estão errando questões consideradas fáceis e buscar

a melhor forma de corrigir as deficiências no ensino. Aliado a isso, avaliar a capacidade das RNAs na classificação de questões e contribuir para trabalhos futuros relacionados a área.

Na seção a seguir serão tratados tópicos importantes para o entendimento do trabalho.

2 DESENVOLVIMENTO

Alguns pontos fundamentais para o entendimento da dinâmica da classificação de questões são discutidos na sequência. A ordem de apresentação do capítulo é a seguinte: Teoria de Resposta ao Item, Rede Neural Artificial, Perceptron, MLP e Trabalhos Relacionados.

2.1 Referencial Teórico

2.1.1 Algoritmos de classificação

O reconhecimento de padrões (RP) é uma tarefa da qual nós, humanos, dependemos ao longo dos anos para sobreviver. É algo que nos trouxe vantagem competitiva no meio natural e, por isso, sabemos que seu uso pode resolver diversos tipos de problema e ser útil nas mais variadas situações. O reconhecimento de padrões, por suas vantagens, tem sido utilizado nos últimos anos como estratégia para melhorar a qualidade do serviço produzido por máquinas. É um subtópico da aprendizagem de máquina, que tem como objetivo classificar objetos em categorias. Para isso análise estatística das aglomerações de dados são realizadas. (PASSOS, 2005). RP pode ser aplicado a processamento de sinais e imagens, sistemas adaptativos, mineração de dados e outros contextos onde existam informações disponíveis e passíveis de análise (PASSOS, 2005).

Realizar o reconhecimento de padrões em determinados datasets pode não ser uma tarefa trivial, por isso foram criados os algoritmos de classificação. Eles tem como objetivo dividir os elementos da amostra em grupos (clusters), de forma que os elementos de um mesmo grupo tenham características semelhantes entre si [Mingoti \(2005\)](#).

2.1.2 Teoria de Resposta ao Item

No ano de 2009 o TRI foi implantado no cálculo das notas do ENEM, deixando para trás a Teoria Clássica dos Testes (TCT) que faz o cálculo das notas levando em consideração apenas a quantidade de acertos. Segundo [Andrade, Tavares e Valle \(2000\)](#) a Teoria de Resposta ao Item (TRI) é um conjunto de modelos matemáticos que representam a chance do indivíduo que está respondendo à questão acertá-la com base nos parâmetros da mesma e da sua habilidade. Ou seja, o modelo é baseado em uma função crescente que indica que quanto maior o traço latente (habilidade), maior a chance de acerto no item [Moreira \(2014\)](#).

Entre as principais vantagens da TRI em relação a TCT é que ela permite que duas populações sejam comparadas, desde que as provas tenham itens em comum [Moreira \(2014\)](#). Com TRI isso é possível de ser feito graças ao modelo teórico baseado nos itens e não nos testes, onde cada resposta correta apresenta uma informação sobre o traço latente [Castro \(2017\)](#).

As funções de probabilidade do tipo logístico são as mais utilizadas atualmente e dão origem aos modelos logísticos [Castro \(2017\)](#). No presente trabalho será trabalhado o modelo 3PL, que descreve a relação entre a habilidade do indivíduo examinado e a probabilidade de uma resposta correta considerando três parâmetros: dificuldade, discriminação e casualidade [Moreira \(2014\)](#)

2.1.3 Rede Neural Artificial

A ideia da construção de tecnologias autônomas, que sejam dotadas de inteligência, sempre foi um dos grandes desafios no âmbito de desenvolvimento tecnológico na história da humanidade. Essa ideia surgiu através da necessidade de solução de problemas em diversas áreas do conhecimento, desde a microbiologia até as engenharias de um modo geral e que eram muito difíceis ou exarcebadamente complexos para as capacidades dos seres humanos. A Rede Neural Artificial (RNA) é uma dessas tecnologias, e através de modelos computacionais simula o funcionamento do sistema nervoso dos seres vivos. As primeiras publicações sobre RNA começaram em meados de 1940 com os trabalhos de McCulloch & Pitts (1943), contudo em 1990 o tema começou a ser fortemente pesquisado devido aos grandes avanços da computação e posteriormente devido a revolução da Internet e consequentemente o *Big Data*.

2.1.4 *Multilayer Perceptron*

Segundo [Silva, Spatti e Flauzino \(2016\)](#) o *Perceptron*, idealizado por Rosenblatt(1958), é a forma mais simples de configuração de uma rede neural artificial, cujo propósito focava em implementar um modelo computacional inspirado na retina, objetivando-se então um elemento de percepção eletrônica de sinais. Uma de suas aplicações consistia em identificar padrões geométricos. A concepção inicial de uma rede *perceptron* consistia em um conjunto de entradas com seus respectivos pesos sinápticos, um único neurônio e uma única saída.

As redes *perceptron* de múltiplas camadas (PMC) também conhecidas como *Multilayer Perceptron* (MLP), são caracterizadas pela presença de pelo menos uma sub camada intermediária (escondida) de neurônios, situada entre a camada de entrada e a respectiva camada neural de saída [Silva, Spatti e Flauzino \(2016\)](#). O algoritmo das PMC é dividido em dois: algoritmo de treinamento e a algoritmo de classificação. No algoritmo de treinamento utiliza-se a ideia de *backpropagation*, dividindo a execução na parte *feedforward* e *backward*. A fase *feedforward* é o próprio algoritmo de classificação, onde a saída de cada neurônio servirá como uma das entradas da próxima camada. A parte *backward* caracteriza-se por repropagar as classificações

129 da rede através de um ajuste de pesos, baseado em equações matemáticas e na classificação
130 advinda da fase *feedforward*.

131 Na seção a seguir serão apresentados trabalhos relacionados com o trabalho e contribui-
132 ram para o seu desenvolvimento.

133 3 TRABALHOS RELACIONADOS

134 Para o desenvolvimento do trabalho foram realizadas buscas por trabalhos relaciona-
135 dos, para desenvolvimento do embasamento teórico e metodológico. A seguir seguem alguns
136 trabalhos relacionados que contribuem para a pesquisa.

137 No trabalho de [Nunes e Primi \(2005\)](#) foi realizado um estudo sobre a quantidade mínima
138 de amostras necessária para utilização da Teoria de Resposta ao Item. Para isso, foram criadas
139 9 amostras de tamanhos diferentes, a partir de uma base de 44 mil respostas de estudantes.
140 Segundo os autores é possível concluir que, amostras com 100 pessoas são instáveis, se tornando
141 estáveis com amostras a partir de 200 pessoas. Com um dataset de 500 sujeitos é possível obter
142 um resultado bem mais próximo em comparação com a amostra completa, sendo que com 20
143 mil estudantes é praticamente igual aos resultados obtidos com a amostra total.

144 Existem alguns programas que podem ser utilizados para do cálculo de TRI, como o
145 BILOG-MG que é um sistema licenciado e o R ([R Core Team, 2017](#)) que é livre. [Barroso e](#)
146 [Rodrigues \(2012\)](#) desenvolveram uma comparação entre os resultados obtidos com a utilização
147 dos sistemas. Para isso, utilizaram 25 questões de matemática e português, que já tinham seus
148 parâmetros calibrados. Para realização dos cálculos foram escolhidos 364 alunos da 5º e 9º
149 série de forma aleatória, que tinham suas habilidades conhecidas. Com essa habilidade foi
150 calculada a probabilidade de cada aluno acertar as questões e após esse processo foi utilizada
151 a distribuição de Bernoulli para gerar a provável resposta de cada aluno a cada questão. As
152 habilidades calculadas pelos sistemas foram parecidas. Ocorreram algumas divergências na
153 estimação de alguns parâmetros, mas, na grande maioria as diferenças foram pequenas. Diante
154 disso, o R é uma ótima ferramenta para fazer análises de TRI.

155 [Castro \(2017\)](#) realizou o cálculo dos parâmetros de TRI com a utilização do pacote ltm
156 do R, com a máxima verossimilhança e com o método “dos intervalos”, utilizando os micro-
157 dados disponibilizados pelo INEP. Segundo a conclusão dos autores, o método dos intervalos
158 é o mais simples para ser implementado, entretanto é menos preciso. O método da máxima
159 verossimilhança foi o mais eficiente para reproduzir as notas do ENEM. Já pacote ltm foi o que
160 conseguiu o melhor resultado na estimação dos parâmetros do TRI.

161 [Anjos e Andrade \(2012\)](#) utilizaram o programa R, que é um software livre, para realizar
162 o cálculo da TRI. Para isso eles selecionaram a base de dados da SARESP para realização do
163 procedimento. Ela possui uma amostra de 3 testes (manhã, tarde e noite) de Língua Portuguesa,
164 que foram aplicados para alunos do ensino médio. Existem no R alguns pacotes para cálculo

165 de TRI, como o ltm (RIZOPOULOS, 2006). No estudo foi feita a análise com o pacote ltm,
166 com o modelo de 2 e 3 parâmetros. Os autores concluíram que até o término do trabalho ainda
167 existiam algumas funções que poderiam ser desenvolvidas para facilitar os trabalhos com o R.
168 Não existe, por exemplo, funções para estimação de parâmetros em modelos normais. Além dos
169 pacotes utilizados no estudo Andrade e Anjos ainda citam outro importante pacote que pode ser
170 utilizado para análise de modelos multidimensionais, o mirt (CHALMERS, 2012).

171 Chagas et al. (2009) utiliza de dois algoritmos de classificação, redes neurais artificiais
172 (RNA) e máxima verossimilhança (Maxver) respectivamente, para classificação do uso da terra
173 no município de Viçosa, MG. O estudo identificou 3 níveis diferentes de degradação das pas-
174 tagens (moderado, forte e muito forte). O simulador de redes neurais empregado foi o “Java
175 Neural Network Simulator” e o algoritmo de aprendizado, o *backpropagation*. Os resultados do
176 trabalho apresentam que o algoritmo de redes neurais conseguiu atingir um índice *Kappa* de 0,8.

177 4 Metodologia

178 Nesta seção serão explicados os passos seguidos para o desenvolvimento desse trabalho.
179 O tópico a seguir mostra a metodologia usada para a seleção da amostra de dados.

180 4.1 Seleção da amostra de dados

181 No trabalho desenvolvido por (NUNES; PRIMI, 2005) foram realizados testes para ave-
182 riguar o tamanho mínimo da amostra, para convergência no sistema de TRI. Isso se torna impor-
183 tante pelo fato do custo computacional para o cálculo do TRI aumentar de acordo com tamanho
184 da base. No ENEM de 2012 cerca de 5791065 participantes realizaram o exame, e caso fosse
185 feito o cálculo levando em consideração toda a amostra o tempo gasto seria grande. No trabalho
186 de (CASTRO, 2017), foram gastas 45 horas para calcular o TRI referente ao ano de 2012, com
187 a utilização da função tpm.

188 Com isso o trabalho desenvolvido por (NUNES; PRIMI, 2005) ganha importância para
189 o trabalho em questão, sustentando que não é necessário a utilização de toda a amostra de
190 dados para o cálculo do TRI, o que gera economias em recursos computacionais e tempo gasto
191 na operação. Eles concluíram que amostras a partir de 200 indivíduos começam a se tornar
192 estáveis, para o cálculo do TRI. Então para o presente trabalho foram geradas 3 amostras de
193 tamanhos diferentes de dados: 30, 40 e 50 mil indivíduos, para escolher a amostra que atende
194 aos critérios. O critério de exigência para escolha da amostra seria quando a oscilação entre o
195 desvio padrão e a média dos parâmetros a , b e c do TRI entre uma amostra e outra se tornasse
196 menor que 0.01. Após os testes verificou-se que a amostra de 30 mil já atendia aos requisitos.

197 Para evitar algum tipo de tendenciamento na escolha dos dados foi desenvolvido um
198 código para fazer o sorteio de 30 mil candidatos. Após o sorteio é feita uma verificação de cada

199 índice sorteado, para verificar se não existem repetidos e se o indivíduo compareceu em todos
200 os dias da prova e não foi desclassificado do exame. Caso alguma das premissas anteriores
201 sejam quebradas um novo índice é sorteado.

202 Com o tamanho da amostra de dados definido será apresentado na próxima seção a
203 metodologia utilizada para o cálculo do TRI.

204 **4.2 Cálculo do TRI**

205 Para realizar o cálculo do TRI foi utilizado os microdados disponibilizados pelo INEP.
206 Eles contém dados sobre a aplicação do ENEM, como dados sociais e econômicos dos partici-
207 pantes e respostas dadas em cada questão. Para o presente trabalho foram selecionados os anos
208 de 2013 à 2015 de aplicação do exame e feito o cálculo do TRI das áreas de Ciências Humanas,
209 Ciências da Natureza e Matemática e suas Tecnologias, totalizando 405 questões.

210 Para isso foi utilizado a biblioteca Mirt do R. Ela espera como parâmetro para realizar o
211 cálculo uma matriz de respostas 0s e 1s, que significa respectivamente uma resposta incorreta e
212 uma correta. Foi desenvolvido então um script para converter os gabaritos da forma alfabética
213 que estão presentes nos microdados para a forma binária. Com isso foi realizado o cálculo do
214 TRI e os resultados foram exportados para um arquivo CSV.

215 **4.3 Tratamento dos dados**

216 Como os parâmetros do TRI possuem dimensões diferentes de tamanho foi feita a nor-
217 malização dos dados. Essa etapa é importante para a classificação dos dados que será feita com
218 o k-means seja feita da forma correta. Optou-se por não fazer a remoção de outliers já que
219 eles podem representar uma questão fácil ou difícil por exemplo. Foi realizada apenas o em-
220 baralhamento dos dados, para remover algum tipo de determinismo. O algoritmo K-means foi
221 utilizado para fazer a classificação das questões com base nos parâmetros a, b e c do TRI. Foi
222 desenvolvido um script em R e usada a função k-means para realizar essa etapa. Para o presente
223 trabalho foram criados 3 grupos de questões, Fáceis, Médias e Difíceis. Essas classificações
224 serão utilizadas para fazer o treinamento da Rede Neural e para comparar os resultados obtidos
225 pelo K-means e pela RNA.

226 **4.4 Classificação em grupos de dificuldade**

227 As provas do ENEM possuem questões de nível fácil, intermediário e difícil, contudo
228 estas questões não estão classificadas em grupos distintos. Esta classificação foi feita através
229 de uma rede neural artificial do tipo *Multilayer Perceptron*, cujo treinamento se deu com 60 %
230 dos dados, 20% para ajustes dos parâmetros e o restante para validação. Para isso, utilizou-se

231 da biblioteca do R "RSNNS" que contém as mais diversas implementações padrões sobre redes
232 neurais. Destas, foi escolhida a função "mlp" que gera uma multilayer perceptron e a treina,
233 através das fases feedforward e backpropagation. A seguir são descritos os passos realizados
234 para classificação das questões do ENEM em 3 grupos de dificuldade, utilizando a rede neural
235 MLP e um *dataset* contendo questões de diferentes áreas (Ciências Humanas, Ciências Naturais
236 e Matemática).

237 O primeiro passo foi ler o dataset das questões, que contém os cálculos do TRI já clas-
238 sificadas pelo k-means. Após a leitura os dados foram embaralhados para evitar que a rede
239 ficasse viciada durante os testes. Em seguida, os dados lidos foram divididos em três amostras,
240 uma com 60% para o treinamento e as outras duas com 20% cada, como foi explicado anterior-
241 mente. Em seguida, foi realizada a chamada da função "mlp" que espera como parâmetro uma
242 matriz com as entradas de treinamento, os valores esperados de classificação para a matriz de
243 entrada, o número de neurônios da rede, o número máximo de iterações da rede, a função de
244 inicialização e seus parâmetros, a função de aprendizado e seus parâmetros, a função de atuali-
245 zação e seus parâmetros, a função de ativação e outros dois parâmetros que definem se a saída
246 da rede será feita linearmente e se ela será embaralhada.

247 Após o treinamento foi construído um gráfico para visualização da taxa de acertos e
248 erros da rede treinada. Após essa fase foi realizada a classificação da rede com 20% da amostra
249 e um gráfico foi gerado para análise dos resultados obtidos. Em seguida, alguns parâmetros
250 da rede foram modificados no intuito de aumentar o desempenho da mesma. Esta modificação
251 nos parâmetros se deu através de estratégias empíricas, resultando nas seguintes configurações
252 finais para rede:

- 253 • Número de neurônios: 6
- 254 • Número de épocas: 20000
- 255 • Função de inicialização: (-0.25, 0.25)
- 256 • Função de aprendizado: *Std_Backpropagation*
- 257 • Parâmetros da função de aprendizado: (0.1)
- 258 • Função de atualização: *Topological_Order*

259 Em seguida foi feito a classificação final com os últimos 20% do *dataset*, a fim de
260 validar corretamente os resultados gerados pela rede com o auxílio de gráficos e uma matriz de
261 confusão.

262 Com base nos passos descritos acima, também foram realizados outros experimentos
263 mudando apenas o *dataset*. Estes experimentos consistiram em treinar uma rede e fazer a clas-
264 sificação com a mesma metodologia utilizada para todas as questões, só que agora o *dataset* foi

reduzido em apenas uma área de conhecimento para cada experimento, cada um possuindo 135 questões. Os experimentos foram divididos em classificações de questões do ENEM em 3 grupos de dificuldades para os *datasets* contendo questões das áreas de Ciências Humanas, Ciências Naturais e Matemática.

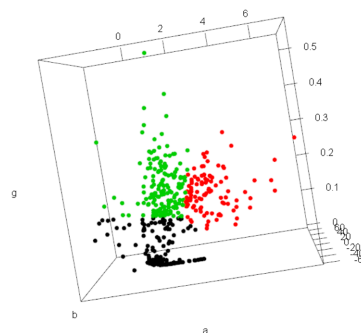
5 Resultados

Os resultados apresentados nesta seção estão divididos em diferentes experimentos. O primeiro é referente a classificação de questões de diferentes áreas de conhecimento em 3 grupos de dificuldade. Os outros também denotam o comportamento da rede na classificação em diferentes grupos de dificuldade, contudo com questões de uma única área do conhecimento.

5.1 Classificação das questões de todas as áreas feitas pelo k-means

A Figura 1 mostra a classificação gerada com a utilização do algoritmo k-means. Para essa classificação foram utilizadas questões das áreas de Ciências da Natureza, Matemática e Ciências Humanas.

Figura 1 – Questões classificadas com o k-means



Fonte: (Os autores 2018)

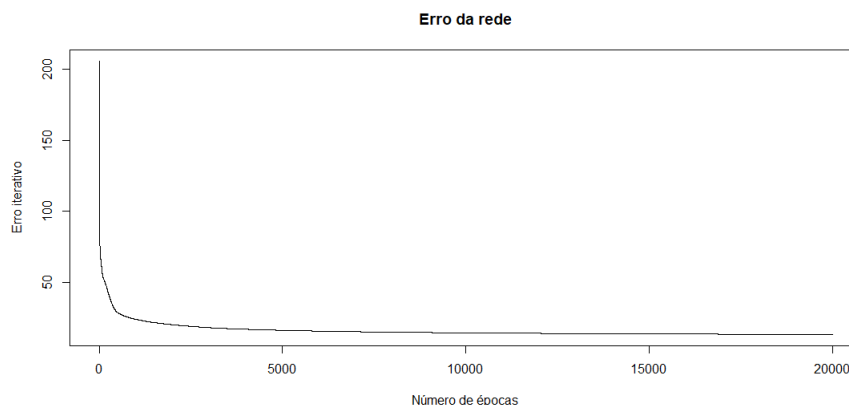
Com as classificações realizadas pelo *k-means* foi gerado o dataset para treinamento da rede neural. Na próxima seção serão apresentados os resultados obtidos.

5.2 Classificação para diferentes Áreas de conhecimento realizadas com RNA

Durante a fase de treinamento, a rede obteve um comportamento de erro aceitável, sendo necessário apenas alguns ajustes como descrito na seção 4. Após os ajustes necessários, foi obtido o gráfico de erro da rede ao longo das épocas como mostra a Figura 2.

Após treinada e ajustada, a rede classificou as questões do ENEM com base no *dataset* das questões de diferentes áreas de conhecimento, como mencionado anteriormente. A figura 3

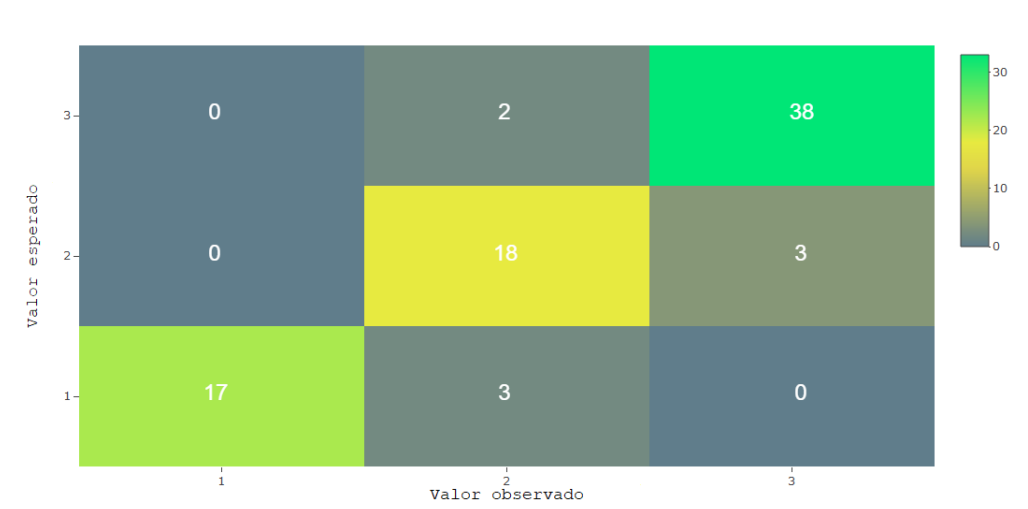
Figura 2 – Erro da rede ao longo das épocas para rede treinada em diferentes áreas de conhecimento.



Fonte: (Os autores 2018)

286 ilustra a matriz de confusão gerada a partir da classificação da rede, onde a diagonal secundária
 287 representa os valores esperados e as demais, os erros da rede. Esses resultados mostram que as
 288 redes neurais são ferramentas poderosas para classificação de objetos em categorias distintas,
 289 Conforme é possível observar, a rede teve 86% de acerto nas classificações realizadas mostrando
 290 ótimos resultados para classificação esperada das questões.

Figura 3 – Matriz de confusão para questões de diferentes áreas de conhecimento.



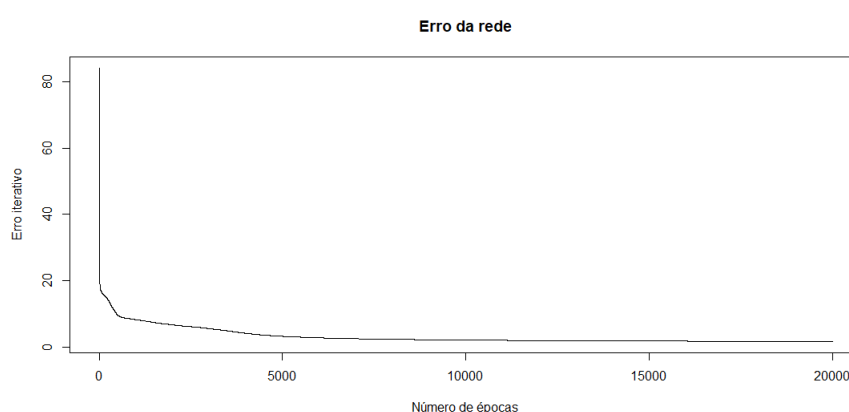
Fonte: (Os autores 2018)

291 5.3 Classificação para áreas únicas de conhecimento

292 Foram criados novos experimentos de classificação com uma única rede para cada área
 293 de conhecimento. A taxa de erro em Ciências Humanas 100%, em Ciências Naturais 96,29%

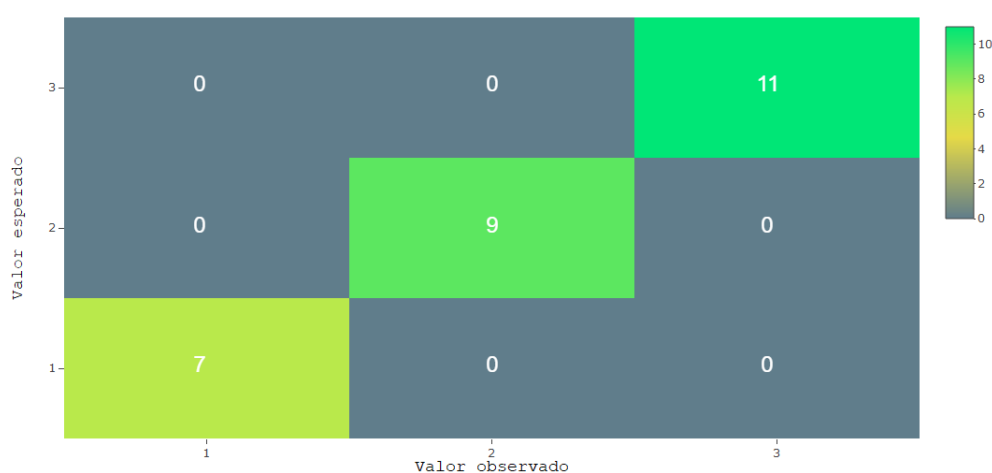
294 e em Matemática foi de 100%. Estes resultados demonstram que redes neurais artificiais trei-
 295 nadas em um ambiente homogêneo possuem maior precisão de acerto quando utilizadas para
 296 classificação. Os desempenhos das redes em cada um destes experimentos foram bastante simi-
 297 lares, logo apenas o resultado de uma área será mostrado, sendo esta Ciências Humanas. Os
 298 resultados das demais áreas estão presentes no Apêndice. Sendo assim, as figuras 4 e 5 mostram
 299 respectivamente, o erro ao longo das épocas e a matriz de confusão gerada a partir da classifi-
 300 cação realizada pela rede para área de Ciências Humanas. Conforme é possível observar a rede
 301 não teve nenhum erro durante a classificação.

Figura 4 – Erro da rede ao longo das épocas



Fonte: (Os autores 2018)

Figura 5 – Matriz de confusão para as questões de Ciências Humanas



Fonte: (Os autores 2018)

302 Comparando as matrizes de confusão geradas a partir da classificação de questões de

diferentes áreas e de áreas únicas de conhecimento, é possível perceber que as redes treinadas para classificação de questões em uma única área obtiveram um melhor desempenho. O gráfico de erro ao longo das épocas não teve tanta diferença, contudo ainda é possível observar que a rede treinada em diferentes áreas do conhecimento converge um pouco mais devagar para zero se comparada com as outras redes.

Diante dos resultados obtidos com o presente trabalho, conclui-se que a utilização do algoritmo *k-means* para classificações de questões do ENEM em grupos de dificuldade é bem interessante em relação a seus resultados, todavia, o custo para realizar esta classificação pode tornar o seu uso não viável dependendo da aplicação. As redes neurais mostraram ser uma boa alternativa ao *k-means* para esse tipo de classificação, uma vez que seu custo de classificação é linear, enquanto o custo do *k-means* para espaço e sementes fixas (d e k respectivamente) é de $O(n^{dk+1})$.

REFERÊNCIAS

- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria de resposta ao item: Conceitos e aplicações. Caxambu, p. 1–164, 2000.
- ANJOS, A. dos; ANDRADE, D. F. de. Teoria da resposta ao item com uso do r. 2012.
- BARROSO, N. C.; RODRIGUES, M. A. Comparações dos softwares bilog-mg e r no uso da teoria de resposta ao item (pp. 152). **Revista da Estatística da Universidade Federal de Ouro Preto**, v. 2, 2012.
- CASTRO, L. G. M. de. Análise dos microdados do enem a partir da teoria da resposta ao item. UFMG, Belo Horizonte, 2017.
- CHAGAS, C. S. et al. Utilização de redes neurais artificiais na classificação de níveis de degradação em pastagens. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 3, p. 319–327, 2009.
- CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1–29, 2012.
- MATIAS, L.; TOLEDO, S. **Enem: veja como 1.434 instituições usam a nota do exame**. 2016. Disponível em: <https://guiadoestudante.abril.com.br/enem/enem-como-1-434-instituicoes-usam-a-nota-do-exame-2/>. Acesso em: 20 jul. 2018.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**. Belo Horizonte: Editora UFMG, 2005.
- MOREIRA, F. G. Cálculo das proficiências dos alunos de mg na prova de matemática do enem de 2011 através de um modelo da teoria de resposta ao item. UFMG, Belo Horizonte, 2014.
- NUNES, C. H. Sancineto da S.; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. **Avaliação Psicológica**, Instituto Brasileiro de Avaliação Psicológica, v. 4, n. 2, 2005.

339 R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria,
340 2017. Disponível em: <<https://www.R-project.org/>>.

341 RIZOPOULOS, D. ltm: An r package for latent variable modelling and item response
342 theory analyses. **Journal of Statistical Software**, v. 17, n. 5, p. 1–25, 2006. Disponível em:
343 <<http://www.jstatsoft.org/v17/i05/>>.

344 SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para**
345 **engenharia e ciências aplicadas**. second. [S.l.]: Artliber Editora Ltda, 2016.

346 **BEHAVIOR OF MLP ARTIFICIAL NEURAL NETWORKS FOR CLASSIFICATION**
347 **OF TRI-BASED ENEM ISSUES**

348 **ABSTRACT**

349 The National High School Examination (ENEM) in Brazil gains each year more importance,
350 replacing traditional vestibular. With this, the development of tools that help prepare students
351 is of great importance. Many simulations are done in a practically random way, with questions
352 chosen without criterion. In this context, this work presents the development of a neural network
353 to classify ENEM issues, based on the parameters of Item Response Theory. For this, the
354 k-means algorithm will be used to generate a dataset containing the classification of these
355 questions into 3 groups: Easy, Medium and Difficult. With this dataset a neural network
356 will be trained to sort the questions and verify the results. In this way this study contributes
357 allowing the elaboration of simulated in a balanced way. In addition, after the network training
358 the same will be able to perform the classification of new issues based on TRI. If k-means
359 continued to be used the classification would have to be made again if new questions were
360 added on the basis of questions. The developed network had practically the same performance
361 as the k-means in the task of classification, this shows the capacity that the neural networks
362 possess of classification.

363 **Keywords:** National School Examination in Brazil. Classification algorithms. Item Response
364 Theory.