

# Rede Neural Artificial para Atribuição de Notas em Relatórios Técnicos num Contexto de Avaliações por Pares e Rubricas

Tiago de Oliveira - Unifesp - tiago.oliveira@unifesp.br  
Denise Stringhini - Unifesp - dstringhini@unifesp.br

**Resumo:** Num ambiente educacional, o processo de atribuição de notas para uma determinada atividade didática é uma tarefa importante e recorrente realizada por um professor ou instrutor responsável pela condução de uma disciplina ou de um curso. Esse processo de atribuição de notas envolve fatores como o contexto educacional onde a metodologia de ensino-aprendizagem está sendo aplicada e os critérios, estratégias e o estilo do professor ou instrutor no estabelecimento das notas. Neste artigo, busca-se investigar a possibilidade de personalização e ao mesmo tempo de automatização de um processo de atribuição de notas para relatórios técnicos utilizando redes neurais artificiais (RNAs) num contexto educacional composto de avaliações por pares e rubricas. A arquitetura de RNA proposta foi submetida à avaliação em dados educacionais reais provenientes de avaliações por pares realizadas por alunos de um curso de Engenharia de Computação nos anos de 2017, de 2018 e de 2019. Os resultados obtidos com o uso das RNAs levando em consideração os dados educacionais reais coletados são promissores, inclusive em comparação com atribuições automáticas de notas disponibilizadas por Ambientes Virtuais de Aprendizagem (AVA) como o MOODLE.

**Palavras-chave:** Rede neural artificial, relatórios técnicos, avaliações por pares, rubricas.

## Artificial Neural Network for Assigning Scores to Technical Reports in Educational Peer and Rubric Assessments

**Abstract:** In an educational environment, the process of assigning a score to a certain didactic activity accomplish by a student is an important and recurring task performed by a teacher or instructor responsible for conducting a course. This scoring process involves factors such as the educational context in which the teaching-learning methodology is being applied and the criteria, strategies and style of the teacher or instructor in setting grades. In this paper, we seek to investigate the possibility of personalizing and at the same time automating the process of assigning scores to technical reports using artificial neural networks (ANNs) in an educational context composed of peer assessments and rubrics. The proposed ANN architecture was submitted to evaluation in real educational data from peer assessments carried out by students of a Computer Engineering course in 2017, 2018 and 2019. The results obtained with the use of ANNs taking into account the real educational data collected are promising, even compared to automatic score assignments made available by Virtual Learning Environments (VLE) such as MOODLE.

**Keywords:** Artificial neural network, technical reports, peer assessments, scoring rubrics.

## 1 Introdução e Contextualização

Na Universidade Federal de São Paulo, encontra-se um curso de Engenharia de Computação cuja metodologia de ensino-aprendizagem adotada em determinadas disciplinas faz uso das técnicas pedagógicas de avaliação por pares (TENÓRIO et al., 2016) e rubricas (REDDY; ANDRADE, 2010). Nesse ambiente educacional, cada aluno da disciplina deve confeccionar um relatório técnico relacionado à execução de experimentos no desenvolvimento de determinados projetos.

Por sua vez, cada relatório técnico produzido por um aluno deve ser distribuído para outros cinco alunos da turma quando da realização do processo de avaliação por pares. Cada aluno-avaliador deverá preencher uma ficha de avaliação composta por rubricas cujo objetivo é ajudá-lo e direcioná-lo na averiguação da qualidade técnica e de escrita do relatório que ele está analisando.

A ficha de avaliação do relatório estrutura-se em nove dimensões, quais sejam: cabeçalho (CAB.), introdução (INT.), objetivos (OBJ.), fundamentação teórica (FUN.), desenvolvimento do trabalho (DES.), resultados obtidos e discussão (RES.), considerações finais (FIN.), referências bibliográficas (REF.) e aspectos gerais (GER.). Para cada dimensão existe um conjunto de perguntas, sendo que em cada uma delas deve-se assinalar, dependendo da qualidade das respostas encontradas no relatório técnico que está em análise, uma dentre cinco estratificações possíveis (Muito Bom, Bom, Razoável, Ruim ou Muito Ruim).

Para informações mais detalhadas sobre a metodologia de ensino-aprendizagem que vem sendo aplicada e os resultados que estão sendo obtidos, consultar os artigos publicados em (OLIVEIRA et

al., 2018), (OLIVEIRA; STRINGHINI; CORRÊA, 2018), (OLIVEIRA et al., 2017) e (OLIVEIRA et al., 2018). Em (OLIVEIRA et al., 2018), demonstra-se a melhora na habilidade de escrita dos alunos submetidos a essa metodologia; em (OLIVEIRA; STRINGHINI; CORRÊA, 2018), demonstra-se a melhora no desenvolvimento dos projetos realizados pelos alunos e; em (OLIVEIRA et al., 2017) e (OLIVEIRA et al., 2018), encontram-se, respectivamente, os pontos de vista dos alunos e dos instrutores da disciplina em relação aos processos de ensino e de aprendizagem da disciplina.

Tendo, nos parágrafos anteriores, contextualizado o ambiente educacional onde se insere a pesquisa descrita aqui neste artigo, suponha agora o cenário hipotético apresentado na Tabela 1 referente ao resultado de um processo de avaliação por pares para um determinado relatório técnico contendo as rubricas assinaladas por cada aluno-avaliador (Aval.) em sua ficha de avaliação correspondente.

Tabela 1: Resultado hipotético para um determinado relatório técnico quando da realização de um processo de avaliação por pares

Aluno	CAB.	INT.	OBJ.	FUND.	DES.	RES.	FIN.	REF.	GER.
Aval. 1	Muito	Muito	Muito		Muito		Muito	Muito	Muito
	Bom	Bom	Bom	Bom	Bom	Ruim	Bom	Bom	Bom
Aval. 2	Muito	Muito	Muito		Muito		Muito	Muito	Muito
	Bom	Bom	Bom	Bom	Bom	Ruim	Bom	Bom	Bom
Aval. 3	Muito	Muito	Muito	Muito	Muito		Muito	Muito	
	Bom	Bom	Bom	Bom	Bom	Razoável	Bom	Bom	Bom
Aval. 4	Muito	Muito	Muito		Muito		Muito		Muito
	Bom	Bom	Bom	Bom	Bom	Razoável	Bom	Bom	Bom
Aval. 5	Muito	Muito		Muito	Muito	Muito			Muito
	Ruim	Ruim	Ruim	Ruim	Ruim	Ruim	Ruim	Ruim	Ruim

Utilizando os dados provenientes dessa tabela para aferir uma nota (de 0 até 100 pontos) ao relatório técnico correspondente, como poderia ser realizada essa aferição/atribuição de nota?

Uma possível resposta seria mapear cada uma das cinco rubricas (Muito Bom, Bom, Razoável, Ruim ou Muito Ruim) em um valor numérico específico para representar uma determinada pontuação. Com isso, pode-se percorrer, para um determinado avaliador, todas as dimensões da ficha de avaliação, acumulando-se as pontuações obtidas. A soma das pontuações obtidas em todas as nove dimensões comporia a nota atribuída por esse avaliador para o relatório técnico. Desta forma, levando em consideração todos os cinco avaliadores, para obter a nota geral do relatório técnico poderia-se calcular a média aritmética das notas atribuídas por cada um dos cinco avaliadores.

Embora essa seja uma resposta possível, com certeza ela não será a única. Se observamos em mais detalhes a Tabela 1, pode-se perceber que as estratificações escolhidas pelo Avaliador 5 (Aval. 5) diferem bastante dos demais avaliadores. Sendo assim, uma outra resposta para a pergunta seria excluir o resultado apresentado pelo Avaliador 5 e utilizar apenas os resultados dos demais avaliadores para a composição da nota geral do relatório técnico.

Outra resposta poderia ser, ao invés de observarmos as linhas da Tabela 1, pode-se computar a nota por meio das dimensões das fichas de avaliação (observando, neste caso, as colunas da Tabela 1). Com essa abordagem, como derivar a nota, por exemplo, para a dimensão RES.? Uma possibilidade seria levar em consideração apenas os resultados dos Avaliadores 1, 2 e 5, excluindo-se os resultados mais discrepantes produzidos pelos Avaliadores 3 e 4. E para a dimensão FIN.? Uma possibilidade seria realizar a média de todos os cinco avaliadores e outra possibilidade seria realizar a média excluindo o Avaliador 5.

Enfim, os exemplos retratados nos parágrafos anteriores têm a intenção de demonstrar as diversas possibilidades ou estratégias diferentes que podem ser utilizadas pelo professor ou instrutor na atribuição de notas para os relatórios técnicos. Utilizando as informações das rubricas atribuídas nas fichas de avaliação pelos alunos quando de um processo de avaliação por pares, quais estratégias ou critérios você, leitor, estabeleceria para atribuir as notas dos relatórios técnicos?

Obviamente, essa resposta não seria a mesma para todos os leitores, sendo que cada um levaria em consideração às especificidades educacionais onde a metodologia foi aplicada e o seu próprio processo mental de construção da nota, baseando-se, para isso, em experiências anteriores armazenadas em sua memória e em inferências cognitivas particulares relacionadas ao seu raciocínio lógico, a sua interpretação e categorização das informações em seu cérebro (esse processo mental é muitas vezes conhecido pelos alunos como sendo o "estilo de correção ou atribuição de nota do professor").

Na sequência, na seção 2, são introduzidos os objetivos da pesquisa realizada dentro do contexto

apresentado nesta primeira seção. Na seção 3, encontram-se os trabalhos relacionados ao tema da pesquisa. Por sua vez, na seção 4, descreve-se a metodologia de condução da pesquisa definida com o intuito de responder às questões formuladas na próxima seção. Na seção 5, apresentam-se os resultados obtidos e, por fim, na seção 6, encontram-se as considerações finais sobre as análises realizadas no decorrer do artigo.

## 2 Objetivos da Pesquisa Realizada

Devido às diversas especificidades e particularidades apresentadas na seção anterior, o objetivo dessa pesquisa é investigar se a Inteligência Artificial, mais especificamente as redes neurais artificiais (RNAs), podem ajudar no processo de automatização na atribuição de notas para os relatórios técnicos por um determinado instrutor ou professor.

A pergunta principal que se busca responder é: Utilizando as notas atribuídas por um determinado instrutor ou professor, a RNA pode ser treinada para que, em disciplinas ou turmas futuras, ela seja capaz de identificar as notas dos relatórios técnicos que seriam atribuídas por esse mesmo instrutor ou professor quando da avaliação por pares realizada pelos alunos? Em outras palavras, a RNA é capaz de inferir e automatizar o processo mental do instrutor ou professor na atribuição das notas ou de seu "estilo de correção/atribuição"?

Por conta das especificidades levantadas anteriormente na seção 1, atualmente realiza-se a atribuição de notas dos relatórios técnicos de forma manual (*ad-hoc*), embora ambientes virtuais de aprendizagem (AVAs), como o MOODLE<sup>1</sup>, por exemplo, tenham a disposição a possibilidade de se utilizar uma atribuição automática de notas, a qual é realizada computando as médias aritméticas de cada avaliação preenchida pela turma de alunos (mesmo procedimento descrito no oitavo parágrafo da seção 1 como sendo uma das possibilidades de atribuição de notas).

Dentro deste contexto, uma questão interessante que se pode articular é: Qual sistema automático de atribuição de notas produzirá o melhor resultado para os dados a serem testados: a atribuição automática realizada, por exemplo, pelo MOODLE com a média aritmética, ou a atribuição automática realizada com o uso de uma RNA?

Sendo assim, como objetivo secundário, na seção de resultados, buscar-se-á comparar o desempenho obtido com o uso da RNA em relação ao processo automático de atribuição de notas realizado pelo MOODLE.

## 3 Trabalhos Relacionados

Focando especificamente o tema de predição de desempenho na área educacional, no trabalho apresentado em (MASON et al., 2018) os autores implementam RNAs probabilísticas para predição da retenção de alunos de engenharia de uma universidade utilizando atributos relacionados à suas demografias e históricos acadêmicos. Por sua vez, os autores em (COSTA et al., 2017) relatam a efetividade do uso de classificadores, incluindo RNAs, na identificação prévia de alunos com possibilidades elevadas de reprovarem em disciplinas introdutórias de programação utilizando as atividades didáticas que foram completadas e os exames intermediários realizados. Em (ZHANG; JIANG, 2018), os autores implementaram RNAs para predição de desempenho de alunos (*score*) em um ambiente virtual de aprendizagem utilizando como variáveis de entrada a taxa de curso completada, o número de postagens no fórum utilizado, a taxa de atividades completadas, o tempo de interação com o curso e a quantidade de dias de acesso. Os autores em (LUFT et al., 2013) reportam a implementação de uma bateria de testes online em formato de jogo para identificar as habilidades cognitivas do aluno e utilizar os resultados provenientes dessa bateria de teste para o treinamento de uma RNA capaz de inferir o rendimento do aluno em matemática (rendimento típico ou rendimento baixo). Na dissertação de mestrado em (BEZERRA, 2018) propõe-se a implementação de RNAs para determinação do desempenho de um aluno (estratificado em três níveis: desempenho superior, intermediário e inferior) utilizando como pré-processamento técnicas de normalização e redução de dimensionalidade sobre vetores de características, os quais contém, dentre outras informações, o número de questões respondidas, número de postagens no fórum e o número de acesso ao curso. Por fim, os autores em (FREITAS et al., 2018) apresentam uma estrutura de RNA para avaliar a presencialidade dos alunos num fórum de discussão

<sup>1</sup>O MOODLE (*Modular Object-Oriented Dynamic Learning Environment*) é uma plataforma computacional de código aberto utilizado como apoio à aprendizagem do aluno. Mais detalhes sobre essa plataforma podem ser encontrados no endereço eletrônico <https://moodle.org>, acessado em 31/08/2019.

do MOODLE, a qual recebe como entrada os atributos de quantidade de postagens enviadas, recebidas, relacionadas e referenciadas e produz como saída a classificação do aluno quanto à sua presencialidade (não satisfatório, fraco, regular, bom ou muito bom).

Como primeira contribuição deste trabalho, pode-se indicar a utilização e análise de RNAs como regressores/estimadores na predição de desempenho ao tentar inferir a nota dos relatórios técnicos no valor de 0 até 100 pontos, diferenciando-o de trabalhos que buscam utilizar as RNAs para categorizar o desempenho em duas ou mais classes (sucesso/insucesso, rendimento típico/rendimento baixo, não satisfatório/fraco/regular/bom/muito bom). Como relatado nos artigos científicos publicados em (CAZAREZ; MARTIN, 2018), (XING et al., 2015) e (CHEN; HSIEH; DO, 2014), há escassez de estudos nessa linha de atuação na área educacional.

Uma segunda contribuição deste trabalho refere-se a investigação da possibilidade de automatizar e ao mesmo tempo personalizar o processo de atribuição de notas, buscando levar em consideração as especificidades e particularidades do instrutor ou professor responsável pela disciplina, como comentado na seção 1.

Por fim, uma terceira contribuição é a aplicação de RNAs em um novo contexto educacional. Os artigos científicos têm retratado contextos educacionais extraindo atributos dos alunos relacionados às suas demografias, históricos acadêmicos, atividades didáticas que foram completadas, exames realizados, o número de postagens no fórum, quantidade de dias de acesso, entre outros, para classificar o desempenho dos alunos. Neste trabalho, busca-se utilizar RNAs em um contexto educacional referente a aplicação de metodologias de ensino-aprendizagem diferenciadas, sendo compostas de avaliações por pares e rubricas, onde os atributos extraídos referem-se ao preenchimento das fichas de avaliação e das rubricas selecionadas pelos alunos.

## 4 Metodologia Adotada na Condução da Pesquisa

Embora neste artigo está se aplicando a técnica de RNAs para a construção de um regressor/estimador, deve-se comentar que outros algoritmos de regressão existem, tais como: regressão linear (*Linear Regression*), *k* Vizinhos mais Próximos (*k-Nearest Neighbors*), Árvore de Decisão (*Decision Tree*) e Máquina de Vetores de Suporte (*Support Vector Machine*).

A escolha de se utilizar as RNAs se deve, basicamente, a dois motivos principais, quais sejam: (1) como o objetivo é verificar a possibilidade de inferir e automatizar o processo mental do instrutor ou professor na atribuição das notas ou de seu "estilo de correção/atribuição", parece-nos coerente utilizar-se de uma técnica de inteligência artificial (neste caso, as RNAs) que, em sua essência, busca emular justamente o sistema nervoso central, modelando computacionalmente os neurônios e suas ligações sinápticas; (2) a técnica de RNA tem sido aplicada com sucesso na área educacional, como pode ser verificado nos artigos científicos publicados em (COSTA et al., 2017), (NADEEM; STANSBURY; MOONEY, 2018) e (BAJAJ; SHARMA, 2018), além disso, comparativamente com outros algoritmos computacionais utilizados para regressão ou para classificação, a RNA também tem se mostrado promissora na área educacional, como pode ser observado nos artigos encontrados em (SHAHIRI; HUSAIN et al., 2015), (BAHADIR, 2016) e (CAZAREZ; MARTIN, 2018).

Os dados para a fase de treinamento e de teste das RNAs foram extraídos de um ambiente educacional real quando da realização de processos de avaliação por pares e do preenchimento das fichas de avaliação pelos alunos na disciplina de Laboratório de Sistemas Computacionais: Arquitetura e Organização de Computadores ministrada na Universidade Federal de São Paulo nos anos de 2017, de 2018 e de 2019.

Ao todo foram ofertadas duas turmas (A e B) em cada ano da disciplina, gerando dados correspondentes a 6 processos de avaliação por pares, totalizando 128 relatórios técnicos e 640 fichas de avaliação, distribuídos de acordo com o apresentado na Tabela 2.

A linguagem de programação Python e a interface de desenvolvimento IDLE foram utilizadas durante a realização dessa pesquisa, permitindo-nos reunir em um único ambiente de programação a execução de diversas tarefas, sendo elas:

- Para o tratamento e adequação dos dados comentados nos parágrafos anteriores, foram utilizadas as bibliotecas Numpy e Openpyxl, as quais possuem classes e funções que permitiram a importação e manipulação dos dados que foram previamente tabulados em planilhas específicas;
- Para as implementações das RNAs e suas fases de treinamento e de teste, foram utilizadas a API Keras juntamente com a biblioteca TensorFlow para processadores de propósito geral,

Tabela 2: Composição dos Dados Utilizados para o Treinamento e Teste da RNA

Ano	Turma	Relatórios Técnicos	Fichas de Avaliação
2017	A	21	105
	B	19	95
2018	A	19	95
	B	20	100
2019	A	25	125
	B	24	120
<b>TOTAL</b>	–	<b>128</b>	<b>640</b>

sendo, inclusive, muito utilizadas atualmente pela comunidade científico-acadêmica (MIGDAL; JAKUBANIS, 2018).

Para a condução da pesquisa, a metodologia adotada se divide em duas etapas. Na primeira etapa selecionam-se dois processos de avaliação por pares (um para a fase de treinamento e outro para a fase de teste) contendo as atribuições de notas realizadas pelo professor da disciplina para realizar uma análise exploratória nesse subconjunto de dados variando-se hiperparâmetros das RNAs como o número de camadas, a quantidade de neurônios em cada camada, tamanho do lote (*batch*) e o número de épocas.

Na segunda etapa, realiza-se a expansão do conjunto de dados disponível para todos os seis processos de avaliação por pares da Tabela 2, avaliando o desempenho das RNAs em diferentes situações com o intuito de verificar a capacidade de generalização da arquitetura de RNA definida. Para cada situação, reserva-se um dos processos de avaliação por pares realizados por uma determinada turma/ano para a fase de teste das RNAs enquanto os demais processos de avaliação por pares são utilizados para a fase de treinamento. Nesta etapa, procura-se verificar os resultados obtidos com o uso das RNAs em relação às atribuições de notas realizadas manualmente pelo professor, comparando-os também com as atribuições de notas realizadas automaticamente pelo MOODLE.

## 5 Resultados Obtidos

### 5.1 Definição da Arquitetura e de seus Hiperparâmetros

A arquitetura implementada classifica-se como sendo um Perceptron Multicamadas (*Multi Layer Perceptron — MLP*) (DU; SWAMY, 2014), tendo camadas totalmente conectadas cujos parâmetros (pesos e *bias*) são calculados por meio do algoritmo de otimização Adam.

A camada de entrada recebe informações das fichas de avaliação referentes aos cinco alunos avaliadores, onde para cada aluno avaliador é necessário informar 9 entradas (representadas por CAB., INT., OBJ., FUN., DES., RES., FIN., REF. e GER.). Essas entradas referem-se às dimensões das fichas de avaliação comentadas anteriormente, sendo assim relacionadas: CAB. — Cabeçalho, INT. — Introdução, OBJ. — Objetivos, FUN. — Fundamentação Teórica, DES. — Desenvolvimento do Trabalho, RES. — Resultados Obtidos e Discussão, FIN. — Considerações Finais, REF. — Referências Bibliográficas e GER. — Aspectos Gerais.

Para cada uma das entradas mencionadas no parágrafo anterior, as rubricas de cada questão das fichas de avaliação são convertidas em números reais e a média aritmética obtida na dimensão correspondente à entrada é utilizada na RNA. Para essa conversão, utiliza-se o seguinte mapeamento: Muito Bom = 1,00, Bom = 0,75, Razoável = 0,50, Ruim = 0,25 e Muito Ruim = 0,00.

Por exemplo, suponha que um determinado aluno avaliador tenha preenchido as duas questões relacionadas à dimensão de Desenvolvimento do Trabalho com as seguintes rubricas: Razoável e Muito Bom. Primeiramente, essas rubricas são convertidas para seus respectivos valores reais. Após o mapeamento, calcula-se a média aritmética obtida (média = 0,75). Essa média é então apresentada como sendo a entrada DES. desse aluno avaliador na RNA.

De acordo com a metodologia de trabalho relatada na seção 4, para a etapa 1, selecionaram-se os dados provenientes do processo de avaliação por pares da turma A de 2019 para a fase de treinamento e o processo de avaliação por pares da turma A de 2018 para a fase de teste na implementação de RNAs, variando-se as configurações de hiperparâmetros com o intuito de se analisar o desempenho das RNAs geradas nesse subconjunto de dados.

Sendo assim, uma pesquisa exploratória inicial foi realizada tomando-se como padrão as recomendações da API Keras quanto ao algoritmo de otimização a ser utilizado (Adam) e seus parâmetros associados, bem como as funções de ativação (Relu), com exceção da última camada, onde foi adotada como padrão a função linear por se tratar de um estimador/regressor. Para a pesquisa exploratória, as seguintes variações de configuração ocorreram: número de neurônios na primeira camada oculta = [10, 20, 30, 40, 45, 50, 60]; números de neurônios na segunda camada oculta = [0, 6, 12, 24, 36, 48, 60], número de épocas = [50, 80, 150, 300, 500] e tamanho do lote = [1, 5, 10]. Todas as combinações possíveis geraram 735 RNAs que foram analisadas quanto aos seus desempenhos obtidos.

Como resultado dessa pesquisa exploratória, tem-se na Tabela 3 uma sumarização dos desempenhos obtidos pelas RNAs implementadas com a variação no número de neurônios da primeira e segunda camadas. A título de exemplo, a primeira entrada da tabela representa a média do desempenho obtido (primeiro número) e seu respectivo desvio padrão (segundo número na primeira entrada da tabela) de todas as RNAs implementadas onde o número de neurônios na primeira camada oculta é 10 e o número de neurônios na segunda camada oculta é 0. Cada entrada da tabela representa a média de desempenho de 15 RNAs (total de épocas x total de lotes).

Tabela 3: Sumarização dos Desempenhos Obtidos com as Variações dos Hiperparâmetros em relação à conjugação dos números de neurônios da Primeira e da Segunda Camadas Ocultas

Cam. 1/2	0	6	12	24	36	48	60	
10	409,01	284,73	228,84	152,75	122,65	104,77	66,78	(média)
	534,47	464,84	382,37	303,09	234,62	166,95	20,96	(desvio)
20	295,94	269,31	143,88	126,35	71,42	62,96	62,59	(média)
	432,25	440,20	287,50	241,15	35,64	6,33	6,06	(desvio)
30	242,30	158,77	128,55	90,83	62,89	<b>61,25</b>	61,34	(média)
	371,86	301,23	253,59	114,10	9,86	9,33	7,91	(desvio)
40	217,01	125,80	111,03	62,67	62,04	62,16	63,27	(média)
	336,29	242,53	188,32	8,31	6,76	8,84	9,53	(desvio)
45	181,51	124,80	95,19	82,42	62,96	64,10	62,86	(média)
	280,64	236,81	125,13	78,01	8,54	8,78	9,60	(desvio)
50	153,05	114,82	101,72	62,47	62,24	61,97	62,41	(média)
	239,43	200,34	151,20	8,22	8,33	7,88	9,39	(desvio)
60	154,06	115,56	93,45	61,69	62,31	61,69	61,67	(média)
	242,19	205,42	120,18	6,66	7,57	8,51	9,31	(desvio)

O desempenho obtido na fase de teste de cada RNA implementada está expresso em erro absoluto (soma dos erros obtidos na saída da RNA em relação à nota atribuída pelo professor de todos os vetores de entrada da fase de teste), sendo que os valores compreendidos para a nota de um relatório técnico variam de 0 até 100 pontos.

De acordo com a tabela apresentada, a melhor média de desempenho obtida foi de 61,25 pontos (destacada em negrito na tabela) e refere-se a 30 neurônios na primeira camada oculta e 48 neurônios na segunda camada oculta, com um desvio padrão de 9,33 pontos.

Nos resultados apresentados na Tabela 3, as RNAs foram implementadas para retornar as notas dos relatórios técnicos na mesma ordem de magnitude do MOODLE, ou seja, as redes foram treinadas para produzirem saídas na faixa de valores de 0 até 100 pontos para os relatórios técnicos (denominadas aqui de saída não-normalizada).

Visando verificar o desempenho das RNAs com saídas normalizadas de maneira semelhante ao realizado com as entradas, a pesquisa exploratória comentada nos parágrafos anteriores foi repetida treinando-se as redes para produzirem saídas na faixa de valores de 0 até 1. Sendo assim, as RNAs produzem indiretamente as notas dos relatórios técnicos, devendo-se multiplicar a saída produzida por 100 para se obter a nota do relatório técnico.

Na Tabela 4, encontra-se o resultado obtido com a realização dessa nova pesquisa exploratória, sendo que, para efeitos de comparação, todas as entradas dessa tabela estão na mesma ordem de magnitude da Tabela 3 (as saídas das RNAs foram multiplicadas por 100).

Interessante observar que, independentemente da saída estar no formato não-normalizado ou normalizado, a melhor configuração conjugada de números de neurônios para as camadas ocultas foi de 30 neurônios na primeira camada oculta e 48 neurônios na segunda camada oculta (30 x 48 neurônios). No entanto, embora a melhor configuração tenha sido a mesma, o desempenho obtido foi diferente, onde claramente percebe-se uma melhora na média do desempenho quando se utiliza a saída normalizada

Tabela 4: Repetição da Pesquisa Exploratória com a normalização da saída das RNAs para valores entre 0 e 1

Cam. 1/2	0	6	12	24	36	48	60	
10	69,29	49,80	49,96	55,25	50,74	50,72	59,93	(média)
	8,94	10,67	5,78	8,43	10,13	9,19	9,99	(desvio)
20	64,79	50,78	54,76	58,66	57,30	57,94	59,05	(média)
	8,92	7,55	7,80	10,48	5,70	7,19	8,79	(desvio)
30	60,47	53,59	52,63	54,55	53,98	<b>46,51</b>	53,38	(média)
	12,31	10,56	7,62	12,93	9,39	12,23	10,19	(desvio)
40	64,40	56,64	53,99	56,93	60,91	61,19	59,94	(média)
	12,79	12,87	10,81	9,81	11,64	12,02	13,00	(desvio)
45	78,89	58,02	61,59	59,49	61,24	64,84	61,99	(média)
	8,10	6,63	9,69	11,23	12,24	7,81	13,77	(desvio)
50	77,00	56,93	55,00	51,93	58,20	59,72	56,52	(média)
	8,95	12,35	9,63	8,81	10,81	9,79	9,35	(desvio)
60	65,11	55,17	58,85	58,00	61,39	58,24	58,09	(média)
	8,46	12,03	10,80	11,47	9,41	11,20	12,68	(desvio)

para a fase de treinamento das RNAs. Para a melhor configuração mencionada de 30 x 48 neurônios, a saída normalizada produziu uma média de desempenho de 46,51 pontos contra 61,25 pontos da saída não-normalizada, apresentando-se um desvio padrão de 12,23 pontos contra 9,33 pontos da saída não-normalizada.

Observe agora os gráficos apresentados na Figura 1, onde se pode comparar a média de desempenho obtido (a) e seu respectivo desvio padrão (b) para as RNAs treinadas com saída não-normalizada versus saída normalizada em relação às épocas variadas. De acordo com esses gráficos, pode-se perceber que com a saída normalizada as RNAs treinadas com uma quantidade menor de épocas conseguem atingir resultados melhores do que com a saída não-normalizada. Também percebe-se que, em média, a saída não-normalizada acaba sofrendo por um tempo maior (em mais épocas) de *underfitting* do que a saída normalizada, enquanto a saída normalizada acaba sofrendo por um tempo maior de *overfitting*, produzindo-se resultados de desempenho piores em *underfitting* do que em *overfitting*. Além disso, em média, também ocorreu uma maior variação desses desempenhos (maior desvio-padrão) em épocas que sofreram mais com o *underfitting* do que em épocas que sofreram mais com o *overfitting*, como pode ser observado ao se confrontar os dois gráficos (a) e (b) da Figura 1.

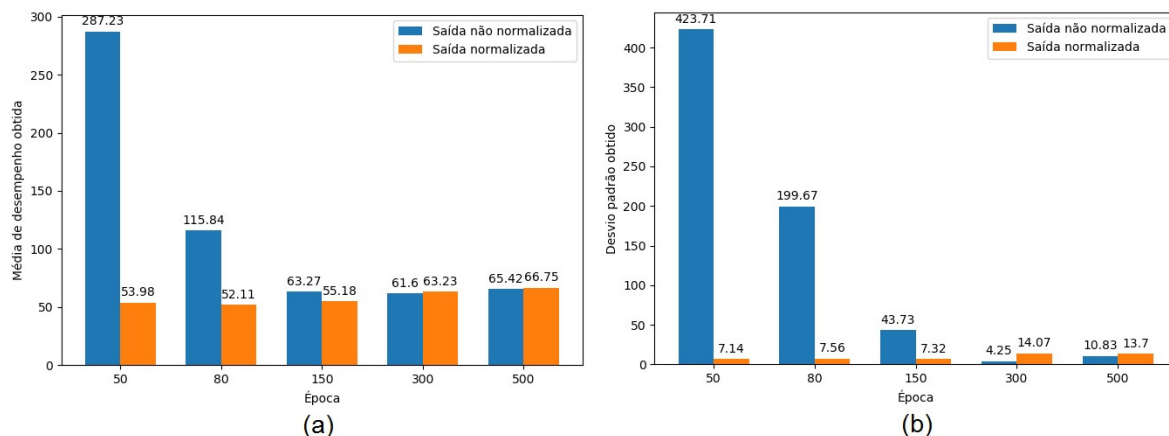


Figura 1: Média de desempenho (a) e seu respectivo desvio padrão (b) para as RNAs treinadas com saída não-normalizada versus saída normalizada.

Tendo como panorama a análise de dados realizada nos parágrafos anteriores, definiu-se a arquitetura da RNA com uma camada de entrada contendo 45 neurônios, uma primeira camada oculta contendo 30 neurônios, uma segunda camada oculta contendo 48 neurônios e uma última camada contendo um único neurônio com saída normalizada, buscando-se produzir números reais entre 0 e 1.

## 5.2 Desempenho na Etapa 2 da Metodologia

Na etapa 2 da metodologia de condução da pesquisa relatada na seção 4, utilizaram-se todos os seis processos de avaliação por pares do conjunto de dados apresentado na Tabela 2 para a análise de desempenho, buscando-se variar os dados de teste das RNAs onde cada teste realizado continha dados provenientes de um processo de avaliação por pares diferente (variando-se ano e turma de alunos). Após uma pesquisa exploratória inicial e buscando-se evitar os problemas relatados de *underfitting* e de *overfitting* utilizaram-se 100 épocas e lotes de tamanho igual a 1 para o treinamento das RNAs, indo em direção ao padrão identificado na análise realizada na subseção 5.1 como melhores parâmetros.

Na Tabela 5, encontram-se os desempenhos obtidos pelas RNAs geradas, comparando-os aos resultados obtidos ao se utilizar a atribuição automática disponível no MOODLE. Para cada entrada da tabela, o desempenho está expresso em erro absoluto total (soma dos erros obtidos na saída da RNA em relação à nota atribuída pelo professor de todos os vetores de entrada da fase de teste correspondente). Além disso, como a saída de cada RNA gerada está normalizada, os resultados apresentados na tabela foram multiplicados por 100 para que estejam na mesma ordem de magnitude dos resultados apresentados pelo MOODLE.

Tabela 5: Desempenho Alcançado pelas RNAs em comparação ao Desempenho Obtido pelo MOODLE

Teste	RNA - Erro Total	MOODLE - Erro Total
2019 A	63,83	217,33
2019 B	60,94	199,11
2018 A	43,53	175,26
2018 B	42,77	173,48
2017 A	60,54	186,96
2017 B	41,10	160,00
Média Total	52,11	185,35
Desvio Total	10,66	20,49

De acordo com a tabela apresentada, em todas as seis situações de teste executadas, as RNAs treinadas obtiveram desempenho superior às atribuições de notas realizadas pelo MOODLE. Na primeira linha da tabela, mostra-se o resultado obtido utilizando como conjunto de dados para a fase de teste da RNA gerada as fichas de avaliação produzidas no processo de avaliação por pares da turma de alunos A do ano de 2019 (2019 A). Neste caso, o erro absoluto total da RNA gerada foi de 63,83 pontos em relação à atribuição de notas realizada pelo professor, enquanto o erro absoluto total do MOODLE foi de 217,33 pontos. Esse resultado indica uma melhora de 70,62% ao se utilizar a RNA ao invés do MOODLE quando da previsão ou inferência de notas para a turma A de 2019.

Realizando deduções semelhantes para as demais linhas da tabela, tem-se: para a previsão de notas da turma B de 2019, a utilização da RNA trouxe uma melhora de desempenho de 69,39% em comparação com o uso do MOODLE; para a previsão de notas da turma A de 2018, a melhora ao usar a RNA foi de 75,16%; para a turma B de 2018, o desempenho da previsão foi de 75,34% em relação ao MOODLE; para a turma A de 2017, o desempenho melhorou 67,61% com a RNA gerada e; por fim, para a turma B de 2017, a RNA trouxe uma melhora de 74,31%.

Em geral, de acordo com os resultados apresentados nos parágrafos anteriores, a utilização de RNAs para inferência das notas dos relatórios técnicos produziu resultados superiores com uma média total de erro de 52,11 pontos levando em consideração as seis situações de teste realizadas, enquanto o MOODLE produziu uma média total de erro de 185,35 pontos, o que é equivalente a uma melhora de desempenho médio de 71,88% da RNA em relação ao MOODLE.

## 6 Considerações Finais

O processo de atribuição de notas num ambiente de ensino-aprendizagem leva em consideração diversos fatores que estão relacionados desde ao contexto educacional onde a metodologia é aplicada até aos critérios, estratégias e ao estilo do professor ou instrutor responsável pela aplicação da metodologia no estabelecimento das notas.

Neste artigo, tendo como referencial um contexto educacional composto de avaliações por pares e rubricas, propõe-se o uso de RNAs para buscar automatizar e também personalizar a atribuição



de notas de relatórios técnicos levando em considerações as especificidades comentadas no parágrafo anterior.

A arquitetura de RNA proposta refere-se a um perceptron multicamadas totalmente conectadas cuja primeira camada recebe os dados provenientes de um processo de avaliação por pares de um determinado relatório técnico (para cada aluno-avaliador são informados 9 números reais entre 0 e 1, um para cada dimensão da ficha de avaliação indicando a rubrica selecionada pelo aluno-avaliador). Como camada de saída, tem-se um único neurônio que produz resultados normalizados entre 0 e 1 representando a nota que deve ser atribuída ao relatório técnico (saída da RNA x 100 para notas entre 0 e 100 pontos).

Utilizando dados reais provenientes de processos de avaliação por pares ocorridos com turmas de alunos entre os anos de 2017 até 2019 de um curso de Engenharia de Computação da Universidade Federal de São Paulo, as RNAs geradas mostraram ser promissoras na tentativa de automatizar e de personalizar a atribuição de notas de relatórios técnicos no contexto educacional mencionado anteriormente. As RNAs geradas de acordo com a arquitetura proposta obtiveram desempenhos superiores em relação às atribuições automáticas de nota realizadas pelo MOODLE.

Como trabalho futuro, pretende-se implementar um sistema/software inteligente baseado em RNAs para automatizar a atribuição de notas dos relatórios técnicos de um determinado instrutor ou professor no contexto educacional descrito na seção 1. Contudo, antes disso, planeja-se realizar uma análise com mais processos de avaliação por pares provenientes de novas turmas de alunos, bem como buscar utilizar a arquitetura definida na subseção 5.1 para a construção de RNAs capazes de inferir atribuições de notas de mais professores ou instrutores visando estudar em mais detalhes a sua capacidade em generalizar as mais diversas estratégias e critérios de atribuição de notas.

Além disso, para evitar ou reduzir os problemas decorrentes de *underfitting* e de *overfitting* relatados na subseção 5.1, também pretende-se analisar o comportamento das RNAs geradas com a implementação de abordagens como a parada antecipada (*early stopping*) (PRECHELT, 2012) que utiliza o conceito de dados de validação além dos dados de treinamento e de teste, a regularização L1/L2 (GOODFELLOW; BENGIO; COURVILLE, 2016) e o Dropout (SRIVASTAVA et al., 2014), bem como verificar a possibilidade de se aumentar artificialmente os dados da fase de treinamento das RNAs (*data augmentation*) (GOODFELLOW; BENGIO; COURVILLE, 2016) (DAO et al., 2018), visando, inclusive, comparar essas abordagens em relação aos seus desempenhos alcançados.

## Referências

- BAHADIR, E. Using neural network and logistic regression analysis to predict prospective mathematics teachers' academic success upon entering graduate education. *Educational Sciences: Theory and Practice*, ERIC, v. 16, n. 3, p. 943–964, 2016.
- BAJAJ, R.; SHARMA, V. Smart education with artificial intelligence based determination of learning styles. *Procedia computer science*, Elsevier, v. 132, p. 834–842, 2018.
- BEZERRA, A. A. *Predição de desempenho de estudantes em ambientes virtuais de aprendizagem*. Dissertação (Mestrado) — Dissertação de mestrado do Programa de Pós-graduação em Engenharia Elétrica e de Computação da Universidade Federal do Ceará, 2018.
- CAZAREZ, R. L. U.; MARTIN, C. L. Neural networks for predicting student performance in online education. *IEEE Latin America Transactions*, IEEE, v. 16, n. 7, p. 2053–2060, 2018.
- CHEN, J.-F.; HSIEH, H.-N.; DO, Q. Predicting student academic performance: A comparison of two meta-heuristic algorithms inspired by cuckoo birds for training neural networks. *Algorithms*, Multidisciplinary Digital Publishing Institute, v. 7, n. 4, p. 538–553, 2014.
- COSTA, E. B. et al. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, Elsevier, v. 73, p. 247–256, 2017.
- DAO, T. et al. A kernel theory of modern data augmentation. *arXiv preprint arXiv:1803.06084*, 2018.
- DU, K.-L.; SWAMY, M. Multilayer perceptrons: Architecture and error backpropagation. In: *Neural Networks and Statistical Learning*. [S.l.]: Springer, 2014. p. 83–126.



- FREITAS, A. et al. O uso de redes neurais artificiais para análise da presencialidade em um fórum lv. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2018. v. 29, n. 1, p. 1903.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016.
- LUFT, C. D. B. et al. Using online cognitive tasks to predict mathematics low school achievement. *Computers & Education*, Elsevier, v. 67, p. 219–228, 2013.
- MASON, C. et al. Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. *Research in Higher Education*, Springer, v. 59, n. 3, p. 382–400, 2018.
- MIGDAL, P.; JAKUBANIS, R. *Keras or PyTorch as your first deep learning framework*. 2018. Disponível em: <<https://deepsense.ai/keras-or-pytorch>>. Acesso em: 2019-09-05.
- NADEEM, M.; STANSBURY, D.; MOONEY, S. Neural educational recommendation engine (nere). In: *IEEE 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. [S.l.], 2018. p. 343–348.
- OLIVEIRA, T. et al. Metodologia de Ensino baseada em Avaliações Colaborativas e Rubricas para o Aprimoramento da Habilidade de Escrita de Relatórios Técnicos em Cursos de Graduação. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*. [s.n.], 2018. p.1603. Disponível em: <<http://br-ie.org/pub/index.php/sbie/article/view/8121>>.
- OLIVEIRA, T. de et al. O ensino e aprendizado de Arquitetura e Organização de Computadores num currículo de Engenharia de Computação estruturado em uma abordagem prática e sistêmica. *International Journal of Computer Architecture Education (IJCAE)*, v. 6, n. 1, p. 27–36, 2017.
- OLIVEIRA, T. de; STRINGHINI, D.; CORRÊA, D. G. M. Online Peer Assessment and Scoring Rubric to Produce Better Digital Systems Designs in an Undergraduate Computer Engineering Curriculum. In: *Latin American Conference on Learning Technologies (LACLO)*. [S.l.: s.n.], 2018.
- OLIVEIRA, T. de et al. Projetando arquiteturas de computadores numa abordagem de ensino prática e sistêmica. *International Journal of Computer Architecture Education (IJCAE)*, v. 7, n. 1, p. 01–10, 2018.
- PRECHELT, L. Early stopping — but when? In: \_\_\_\_\_. *Neural Networks: Tricks of the Trade: Second Edition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 53–67. ISBN 978-3-642-35289-8. Disponível em: <[https://doi.org/10.1007/978-3-642-35289-8\\\\_5](https://doi.org/10.1007/978-3-642-35289-8\\_5)>.
- REDDY, Y. M.; ANDRADE, H. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, v. 35, n. 4, p. 435–448, jul 2010. ISSN 0260-2938. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/02602930902862859>>.
- SHAHIRI, A. M.; HUSAIN, W. et al. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, Elsevier, v. 72, p. 414–422, 2015.
- SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.
- TENÓRIO, T. et al. *Does peer assessment in on-line learning environments work? A systematic review of the literature*. [S.l.]: Pergamon, 2016. 94–107 p.
- XING, W. et al. Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, Elsevier, v. 47, p. 168–181, 2015.
- ZHANG, Y.; JIANG, W. Score prediction model of moocs learners based on neural network. *International Journal of Emerging Technologies in Learning (iJET)*, v. 13, n. 10, p. 171–182, 2018.