PROPOSTA DE USO DE REDES NEURAIS PARA CLUSTERIZAÇÃO DE DADOS NO CONTEXTO DE ANÁLISE QUALITATIVA DE INFORMAÇÕES DE RECLAMAÇÕES EM SERVIÇOS DE TELECOMUNICAÇÕES

Willian Santos Silva[1], Aniel Silva de Morais[1] e Wender de Oliveira Silva[1]

[1] Universidade Federal de Uberlândia, Brasil. williansantosptc@gmail.com; morais.aniel@gmail.com wender4j@gmail.com

Resumo. Este artigo apresenta um trabalho em desenvolvimento no qual se utiliza dos conhecimentos desenvolvidos até o momento nas áreas de redes neurais para uma aplicação prática. A relação entre a exponencial quantidade de dados gerados seja por equipamentos ou pelas interações humanas precisam de ferramentas que possam auxiliar na correta interpretação. A análise qualitativa possui um grande potencial no auxílio e no entendimento de problemas, tendências, conclusões e pode direcionar de forma mais assertiva as ações necessárias para as resoluções de problemas, melhorias e evolução. Dentro deste contexto é abordada a proposta do uso de uma das técnicas de redes neurais como uma ferramenta no auxílio destas análises. Dentro do contexto da teoria de redes neurais é abordado a utilização de clusterização de informações tomando como exemplo e caso prático, informações relacionadas a qualidade e serviços prestados em serviços de telecomunicações. Conclui-se com os estudos preliminares a possibilidade de clusterização e assim a possibilidade da evolução dos estudos no contexto proposto.

Palavras-chave: Tecnologia; Redes Neurais; Clusterização; Análise Qualitativa; Telecomunicações.

PROPOSAL FOR THE USE OF NEURAL NETWORKS FOR DATA CLUSTERING IN THE CONTEXT OF QUALITATIVE ANALYSIS OF COMPLAINTS INFORMATION IN TELECOMMUNICATIONS SERVICES

Abstract. This article presents a work in progress in which it uses the knowledge developed so far in the areas of neural networks for practical application. The relationship between the exponential amount of data generated either by equipment or by human interactions needs tools that can assist in the correct interpretation. Qualitative analysis has great potential in helping and understanding problems, trends, conclusions and can more assertively direct the actions needed to solve problems, improve and evolve. Within this context, the proposal to use one of the neural network techniques as a tool to aid these analyzes is approached. Within the context of neural network theory, the use of information clustering is approached taking as an example and practical case information related to quality and services provided in telecommunications services. It concludes with the preliminary studies the possibility of clustering and thus the possibility of the evolution of the studies in the proposed context.

Keywords: Technology; Neural Networks; Clustering; Qualitative Analysis; Telecommunications.

1. INTRODUÇÃO

A construção de mecanismos autônomos dotados de inteligência vem sendo publicados desde os anos 50. Apresenta grande potencial nas mais diversas áreas, como classificação de padrões, reconhecimento e outros (Silva et al., 2010).

DOI: https://doi.org/10.36367/ntqr.4.2020.499-506

Dentro das possibilidades de utilização em problemas de categorização, que por vezes são confundidas com classificação, engloba a descoberta de categorias ou classes (Braga et al., 2003).

O objetivo científico é descrever, interpretar, explicar e predizer a realidade. Cada abordagem possui uma contribuição e a integração de diferentes abordagens possibilita que uma área inexplorada seja incorporada a análise resultando em uma pesquisa mais robusta (Paranhos et al., 2016).

Os dados quantitativos (números, indicadores) podem ser analisados com o auxílio da estatística, revelando informações rápidas, úteis e confiáveis sobre determinado número de observações (Creswell, 2007). As técnicas qualitativas fornecem informações sobre os entrevistados e oferecendo diferentes perspectivas sobre os fenômenos. As integrações das análises qualitativas e quantitativas possibilitam em tirar o melhor de cada método e assim responder por uma questão específica, trazendo confirmação e complementaridade (Paranhos et al., 2016).

Este artigo traz então a perspectiva de RNAs como uma ferramenta de complementaridade na pesquisa qualitativa. Para buscar inserir seus resultados ou possibilidades que venham a ser possíveis durante o desenvolvimento deste trabalho foi selecionado como área de aplicação serviços de telecomunicações, mais especificamente a análise de qualidade de redes banda larga.

A internet provavelmente é o hoje o maior sistema de engenharia já criado pela humanidade (Kurose & Ross, 2013). Para a prestação de serviços é necessário que haja padrões de qualidade e compromisso por parte dos provedores de serviço de internet ou do inglês *Internet Serviçe Providers* (ISPs). Em muitos casos cabe aos profissionais de atendimento receber através de entrevistas com os clientes e apontar e investigar possíveis problemas baseadas na percepção e de outras informações. Neste contexto, as análises quantitativas e qualitativas se inserem como ferramentas que podem auxiliar a observar diversos problemas e melhorar a interação entre operadora e clientes.

Como objetivo, este artigo apresenta o trabalho em desenvolvimento do uso de Redes Neurais Artificiais (RNAs) para clusterização de dados de qualidade de serviços prestados de telecomunicações para uso na pesquisa qualitativa destes.

O artigo está estruturado nos próximos tópicos apresentando primeiramente os objetivos do trabalho em andamento, em sequência a metodologia e por fim, são apresentados os resultados preliminares e conclusões parciais.

2. OBJETIVOS

O principal objetivo deste artigo é trazer a comunicade acadêmica uma visão do trabalho parcialmente desenvolvido com a utilização de redes neurais. Este artigo apresenta o trabalho em desenvolvimento de clusterização, servindo como uma ferramenta complementar a investigação qualitativa de informações. Foi tomado como estudo de caso dados relacionados a qualidade de serviços prestados por operadoras de telecomunicações, cenário de atuação do investigador.

3. MÉTODOS

Para o estudo foi utilizado uma base de dados de classificação de análise de reclamações e os principais motivadores relacionados de reclamação em serviços de telecomunicações, mais especificamente de banda larga. Estes dados são coletados no período mensal através de chamadas recebidas e pelo fechamento de ordens de serviço de campo. O instrumento de coleta consiste no roteiro semiestruturado.

Nesta base são informadas em cada fechamento de chamada ou de ordem motivadores principais. Para o estudo foi selecionado uma base de 1 mês contendo 30 diagnósticos e 5 informações principais conforme Quadro 1, com redução de dimensionalidade e tratamento dos dados que eventualmente estejam com erros de preenchimento ou com valores nulos. A base total possui 7378 entrevistas referentes a 1 mês utilizado a fim de se testar a proposta metodológica.

Quadro 1. Amostra de exemplo dos dados trabalhados como entradas para SOM com dados de fechamento de chamados de reclamação.

Análise Final	Grupo de atendimento	Diagnóstico	Possui Bloqueio	Problema Modem
Validações iniciais	Empresarial	Finalizado sem abertura de ordem de serviço	Não	Não
Validações iniciais	Empresarial	Finalizado com abertura de ordem de serviço	Sim	Não
Modem Alarmado	Corporativo	Finalizado sem abertura de ordem de serviço	Não	Sim

DOI: https://doi.org/10.36367/ntqr.4.2020.499-506

Uma vez definida a base, as palavras são substituídas por valores numéricos a fim de se formar uma matriz que servirá como entrada no software de processamento da informação que fará a clusterização dos dados a ser fornecido como ferramenta de método misto ao investigador.

Para a programação e desenvolvimento do código foi utilizada a linguagem de programação Phyton e as bibliotecas numpy, matplotlib, o código desenvolvido seguiu os princípios do algoritmo de rede neural Self Organizes Maps (SOM) (Kohonen, 1990). A base de dados de reclamação é inserida pela leitura do próprio código Phyton através de uma janela de seleção de arquivo, sendo este definido em formato de texto simples (.txt).

O processamento destes dados utiliza o princípio do cérebro humano que possui em torno de 10^{11} neurônios que são a célula fundamental que processa e se comunica com milhares de outros de forma paralela e contínua. Essa unidade individual e sua topologia são a base para o estudo das redes neurais artificiais (Braga et al., 2003). A estrutura de um neurônio é apresentada na Figura 1.

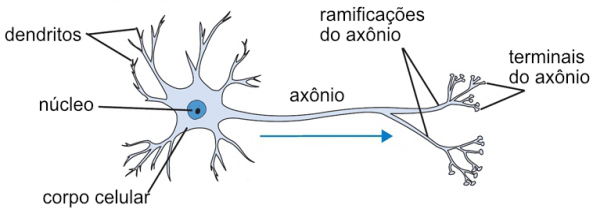

Figura 1. Estrutura de um neurônio biologico

RNAs são processadores simples paralelamente distribuídos e que tem a propensão de armazenar conhecimento experimental. Sua inspiração no cérebro humano se dá no aspecto de o conhecimento ser adquirido pela rede neural através de um processo de aprendizagem; o segundo aspecto trata-se das forças de conexão entre os neurônios, ou pesos sinápticos, que armazenam o conhecimento adquirido (Haykin, 2017).

As RNAs tentam reproduzir funções das redes biológicas através de similaridades entre esses dois tipos de sistemas. Apesar da pouca similaridade do ponto de vista biológico, essas permitem a reprodução de várias funções cognitivas dos seres humanos (Braga et al., 2003). A Figura 2 apresenta a estrutura básica de um neurônio artificial (Silva et al., 2010).

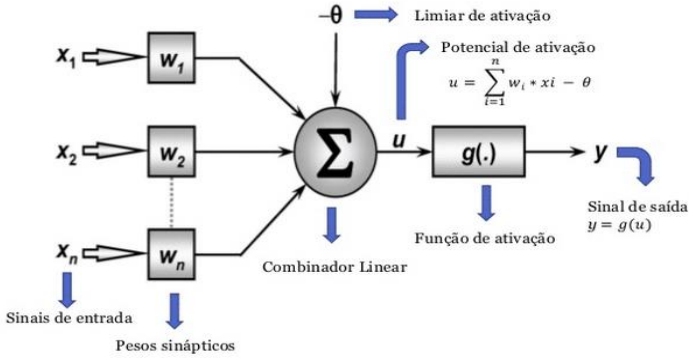

Figura 2. Estrutura de um neurônio artificial.

O neurônio artificial é um modelo simplificado e simulado do neurônio real. Os sinais de entrada representados por Xn são as variáveis ou informações a serem consideradas e processadas pelo neurônio. Similarmente essas entradas representam o que biologicamente seriam os dentritos em um neurônio biológico (Fig.1). As sinapses, onde ocorre a transmissão de impulsos nervosos de uma célula para outra é emulada por wn ou pesos sinápticos acoplados às entradas dos neurônios. O combinador linear realiza o somatório de todas estas entradas, multiplicadas por suas respectivas forças de conexão sináptica (os pesos), assim se este valor atingir mais ou menos que seu potencial de ativação gera uma saída yn.

O SOM está sendo utilizado no trabalho em desenvolvimento como ferramenta para clusterização dos dados semelhantes inspirado em trabalhos como (Affonso, 2011; Gonçalves et al., 2008; McCaffrey, n.d.; Pinho, 2008). O resultado serão duas imagens em que são agrupados os dados conforme a proximidade dos atributos referidos na Quadro 1. Estes atributos formam um vetor (x1, x2, x3, xn). O SOM os agrupa em clusters cada um com seu centróide (Gonçalves et al., 2008; Kohonen, 1990; Pinho, 2008; Rodrigues, n.d.).

O primeiro mapa formado servirá então como referência aos grupos formados indicando os volumes de informações mais próximas, ou seja, a formação de clusters de padrões encontrados nos dados analisados. O segundo mapa trará uma redução de dimensionalidade para facilitação da visualização dos clusters formados de acordo com a escala dos valores dos atributos. Uma vez com os padrões processados e apresentados espera-se que a análise qualitativa seja facilitada assim como com outros softwares específicos, porém utilizando uma abordagem adicional baseada em aprendizado de máquina.

4. RESULTADOS PRELIMINARES

Para o teste realizado, considerando cinco principais informações relevantes das bases de dados de reclamação, foi obtido um mapa de 30x30 (aqui representado com eixo x e y bidimensional), com resultados que são as saídas processadas pela rede neural.

Pelos dados de saída para classificação obteve-se quatro principais clusters, ou seja, dos dados da base poderia se considerar como quatro os grandes motivos de reclamações dos clientes de serviços de banda larga processados nesse teste. Para o código desenvolvido e considerando na metodologia apresentada, quanto a distribuição e visualização é possível obter uma amostra visual apresentada na Figura 3. Está é uma amostra em relação aos dados processados.

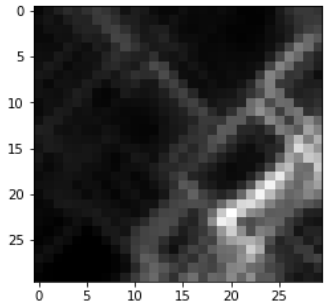

Figura 3. Mapa de Kohonem 30x30 pontos.

As células pretas indicam itens de dados que são semelhantes entre si e as células brancas indicam bordas entre grupos de itens semelhantes. Como os dados de entrada possuem mais de 2 dimensões não é possível representar graficamente toda a distribuição dos dados processados, sendo a visualização em 2 dimensões uma forma de redução de dimensionalidade.

Para uma visualização melhor e representação em duas dimensões foi aplicado à redução de dimensionalidade e apresentado na Figura 4 com correções e utilizando técnicas gráficas próprias da linguagem de programação Phyton e da biblioteca matplotlib.

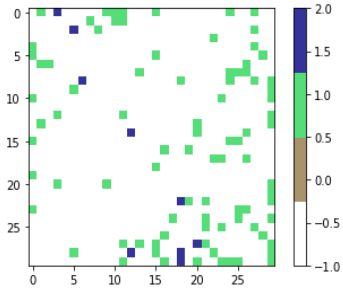

Figura 4. Redução de dimensionalidade para os dados considerados.

DOI: https://doi.org/10.36367/ntqr.4.2020.499-506

Pela Figura 4 nota-se que predominantemente temos maior presença do grupo 3 (verde) porém de forma mais espalhada na representação e baixa presença do grupo 4 (azul) e nenhuma representatividade dos demais grupos. Considerando a redução de dimensionalidade, podendo representar um único e forte motivo da causa das reclamações dos usuários de serviços de banda larga. Este espalhamento é provocado devido a quantidade de entradas.

Com essas informações então seria possível entender e a melhor opção do teste de hipótese definida e explorar o comportamento do consumidor quanto a qualidade do serviço. Sendo utilizadas juntamente como observações, entrevistas em profundidade, estudos de caso (Godoy, 1995).

5. CONCLUSÕES

Este ainda é um trabalho preliminar com o intuito de apresentar uma visão e a abordagem de novas perspectivas do uso de RNAs (Silva et al., 2010) como uma ferramenta de auxílio a investigação qualitativa. Ainda se faz necessário prosseguir com desenvolvimento de melhorias para ampliação da matriz de dados a ser processada, melhoria das apresentações gráficas dos mapas SOM abordando novas arquiteturas e novos atributos dos dados já coletados. Melhorar a visão dos grupos de respostas da entrevista para que sejam apresentados de forma clara os clusters mostrando agrupamentos semelhantes entre as informações coletadas e assim apresentar uma forma de entender a relação indissociável entre o mundo objetivo e a subjetividade.

Este seria um método adicional ao pesquisador uma vez que o método qualitativo não emprega método estatístico (Godoy, 1995). Estes resultados servirão como ponto de atenção para a melhoria do estudo e como possibilidade a trabalhos futuros que venham a utilizar softwares ou metodologias no apoio à análise qualitativa.

Não é possível descartar a sua utilização em um contexto diferente do proposto neste trabalho, tendo em vista que as possibilidades de utilização de RNAs são as mais diversas e nos mais abrangentes campos da ciência (Haykin, 2017). Outro fato, como apontado na Fig. 4, que poderia ser já utilizado como uma ferramenta adicional na análise do investigador.

Assim, como exposto em visões da bibliografia em relação à análise qualitativa e quantitativa é importante a integração das análises para que se possa tirar o melhor de cada método e assim trazer confirmação e complementaridade (Paranhos et al., 2016)

O principal objetivo deste artigo é trazer à comunidade acadêmica uma visão do trabalho parcialmente desenvolvido com a utilização de RNAs para clusterização, servindo como uma ferramenta complementar a investigação qualitativa de informações e tomado como estudo de caso dados relacionados a qualidade de serviços prestados por operadoras de telecomunicações.

6. REFERÊNCIAS

Affonso, G. S. (2011). Mapas Auto-organizáveis de Kohonen (SOM) aplicados na avaliação dos parâmetros da qualidade da água. In *Autarquia Associada à universidade de São Paulo*. Universidade de São Paulo.

Braga, A., Carvalho, A., & Ludermir, T. (2003). Redes neurais artificiais. *Sistemas Inteligentes: Fundamentos e Aplicações*.

Creswell, J. W. (2007). Qualitative enquiry & research design, choosing among five approaches. In *Book: Vol. 2nd ed*. https://doi.org/10.1016/j.aenj.2008.02.005

Godoy, A. S. (1995). Introdução à pesquisa qualitativa e suas possibilidades. *Revista de Administração de Empresas*, *35*(2), 57–63. https://doi.org/10.1590/s0034-75901995000200008

Gonçalves, M. L., Netto, M. L. de A., Zullo Jr., J., & Costa, J. A. F. (2008). Classificação Não-Supervisionada de Imagens de Sensores Remotos Utilizando Redes Neurais Auto-Organizáveis e Métodos de Agrupamentos Hierárquicos. *Revista Brasileira de Cartografia*.

Haykin, S. (2017). Redes neurais: princípios e prática. *Bookman*. https://doi.org/8573077182

Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE*. https://doi.org/10.1109/5.58325

Kurose, J. F., & Ross, K. W. (2013). Redes de Computadores e a Internet. In *Editora PEARSON*.

McCaffrey, J. (n.d.). *Visual Studio Magazine Self-Organizing Maps Using Python*.

Paranhos, R., Filho, D. B. F., Da Rocha, E. C., Júnior, J. A. da S., & Freitas, D. (2016). Uma introdução aos métodos mistos. *Sociologias*. https://doi.org/10.1590/15174522-018004221

Pinho, A. G. de. (2008). Mineração de Dados com Mapas de Kohonen: uma Abordagem no Setor Financeiro. *Revista Pensamento Contemporâneo Em Administração*. https://doi.org/10.12712/rpca.v2i1.140

Rodrigues, N. M. (n.d.). *Fatores de Hipertensão usando Redes Neurais Kohonen – SOM*.

Silva, I. N., Spatti, D. H., & Flauzino, R. A. (2010). Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas. In *São Paulo: Artliber*.