Caracterização do Consumo de Água Residencial utilizando Redes Neurais Artificiais

Lucas Nunes Monteiro* Fernando Pinhabel Marafão*
Alexandre da Silva Simões* Ivando Severino Diniz* Caíque Vendemiatti*

*Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba
BR (Tel: (15)3238-3409; e-mail: lucasnunes_tec@hotmail.com, alexandre.simoes@unesp.br).

Abstract: Water leaks in residential pipes are usually noticed only after 30 days, when the consumer receives the bill regarding the use of water and realizes that there was an atypical consumption on site. In such scenario, this article proposes the use of an artificial neural network to propose a methodology for identifying possible water leaks in residential buildings, thus allowing a better understanding and awareness of water consumption in the property. The proposed water monitoring system is based on four main stages: data acquisition; exploration and manipulation of data; data characterization and classification. Through the proposed system, from the classification / forecast of one week, it was possible to identify dates with abnormal consumption, which could be caused by excess water consumption or failures in water lines.

Resumo: Os vazamentos de água que ocorrem nas tubulações das residências são notados apenas após 30 dias, quando o proprietário recebe a conta referente a utilização de água e percebe que houve um consumo atípico no local. Diante deste cenário, este artigo propõe o uso de uma rede neural artificial para proposição de uma metodologia de identificação de possíveis vazamentos de água em uma residência, permitindo assim um melhor entendimento e conscientização do consumo de água na propriedade. O sistema de monitoramento de água proposto é baseado em quatro estágios principais: aquisição de dados; exploração e manipulação de dados; caracterização dos dados e classificação. Por meio do sistema proposto, a partir da classificação/previsão de uma semana, foi possível identificar datas com consumos anormais, podendo ser causados por excesso de consumo de água ou avarias relacionadas a vazamentos.

Keywords: Water Loss Control; Machine Learning; Water Leaks; ANN.

Palavras-chaves: Controle de Perdas de Água; Aprendizagem de Máquina; Vazamentos de Água; RNA.

1. INTRODUÇÃO

A demanda por água potável vem crescendo no cenário mundial devido ao crescimento populacional que ultrapassa 7 bilhões de habitantes. No entanto, mesmo com a crescente demanda de água, muitos países desperdiçam água, seja por uso indevido, excesso de consumo ou por meio de falhas na rede de distribuição, causando os vazamentos (Marcato et al. 2018). No ano de 2014, o Estado de São Paulo vivenciou a pior crise hídrica de sua história, quando os principais reservatórios da região metropolitana quase foram esgotados. O sistema Cantareira atingiu um nível de 14,6%, o menor desde a sua criação em 1974 (ABES et al. 2015). Atualmente, o estado de São Paulo desperdiça quase 40% de água potável, sendo por fraudes, furtos e atividades clandestinas, falta e/ou erros de medição e vazamentos de água (ABES et al. 2015). Uma pesquisa realizada pela Agência Nacional de Águas (ANA) aponta que o consumo de água no Brasil irá crescer 24% nos próximos dez anos. O Brasil utiliza, em média, 2 milhões e 83 mil litros de água por segundo. Em 2030, a utilização de água atingirá os 2,5 milhões de litros por segundo (Marcato et al. 2018).

No cenário internacional, de acordo com uma pesquisa realizada pelo *World Resources Institute (WRI)*, mais de 160 países experimentarão crise hídrica nos próximos anos (FUNASA et al. 2014).

2. RNAs (Redes Neurais Artificiais)

No aprendizado de máquinas, as abordagens conexionistas – ou RNAs – se caracterizam pela presença de um conjunto de unidades simples de processamento que são abstrações matemáticas inspiradas no sistema nervoso de seres vivos (Zambiasi et al. 2002). Essas unidades de processamento são conhecidas como neurônios artificiais, e são interligadas por um grande número de interconexões, chamada de sinapses artificiais (Zambiasi et al. 2002). As sinapses artificiais são representadas computacionalmente utilizando-se vetores ou matrizes de pesos sinápticos, que são sintonizados para representar um determinado mapeamento entrada-saída por meio de um processo denominado aprendizado. Os pesos sinápticos são utilizados, portanto, para o armazenamento do conhecimento adquirido (Zambiasi et al. 2002).

Os principais elementos de uma RNA são:

- Neurônios: unidade elementar onde ocorre o processamento das informações;
- Conexão: elemento de transmissão dos sinais de um neurônio para outro;
- Peso sináptico: força do sinal em uma conexão, matematicamente representado por um valor (peso) que multiplica o sinal transmitido;
- Função de ativação (ou função de transferência): função que computa o estado de cada neurônio com base nas entradas recebidas.

Sejam os pesos sinápticos de um neurônio representados por $W = [w_1, w_2, \dots, w_n]$ (*weight*). Esses pesos determinam o grau de importância de cada sinal de entrada em relação àquele neurônio. O sinal de ativação de um neurônio tipicamente é resultante do somatório do produto dos sinais de entrada, representado por $X = [x_1, x_2, \dots, x_n]$, pelos pesos sinápticos (Falkenberg et al. 2003). As funções de transferência típicas nesses neurônios são do tipo semilinear e diferenciáveis. As funções mais utilizadas são a sigmoide e a tangente hiperbólica. A Fig. 1 ilustra esse modelo de neurônio artificial, denominado “perceptron”.

Fig. 1 Neurônio Artificial: Sinais de entrada e pesos sinápticos.

2.1 Arquitetura

A arquitetura de uma RNA – usualmente também denominada topologia – remete à forma como os neurônios encontram-se dispostos, um em relação ao outro em um agrupamento. Quanto à relação entrada-saída, existem dois tipos básicos de organização: *feedforward* e *feedback* (Castro et al. 2005) (Zambiasi et al. 2002). No modelo *feedback*, há conexões entre a saída e os neurônios de entrada, que constitui uma memória dos estados anteriores. Assim, o próximo estado depende não só dos dados de entrada, mas também das saídas prévias da rede. Ou seja, a rede realimenta suas entradas com os sinais de saída (Castro et al. 2005).

Já no modelo *feedforward* não há realimentação das saídas para os neurônios de entrada. Dessa forma a rede não mantém uma memória dos valores de saída passados e dos estados de ativação dos neurônios, logo não há correlação temporal entre as diferentes entradas. As redes neurais *feedforward* podem apresentar uma ou mais camadas. As camadas ocultas – todas as camadas entre a de entrada e a de saída – também são constituídas de neurônios artificiais e podem ser vistas como um extrator de características (Zambiasi et al. 2002). A Fig.2

ilustra uma RNA *single-layer* e a Fig. 3 ilustra uma RNA *multi-layer*.

Fig. 2 RNA Single-layer.

Fig 3 RNA Multi-layer.

2.2 Treinamento

O procedimento para sintonização dos parâmetros livres W em uma RNA é denominado algoritmo de treinamento. Este algoritmo modifica os pesos sinápticos de forma adequada para atingir um objetivo determinado (Silva et al. 2010) (Hussin et al. 2016). Um dos algoritmos mais amplamente utilizados para o treinamento de RNAs do tipo multilayer perceptron é a retropropagação do erro (Rumelhart et. al., 1986). Trata-se de um algoritmo de treinamento que segue o paradigma supervisionado, ou seja, onde é fornecido à RNA um conjunto de exemplos de treinamento que deverá ser aprendido e generalizado pela rede. Quando as RNAs têm o objetivo de atuar como classificadores, as informações de entrada são usualmente chamadas de *features* (características) e as informações de saída são chamadas de *labels* (rótulos) (Ren et al. 2010). Tipicamente, o processamento de um *multilayer perceptron* (MLP) é realizado em duas fases. Na fase *feedforward* ocorre o processamento das entradas para geração de saídas. Na segunda fase, com base no erro da rede – diferença entre a saída obtida e a desejada –, o valor dos pesos sinápticos da rede é atualizado através de uma propagação do erro de trás para a frente (*backward*). O equacionamento para o backpropagation é apresentado em Haykin (2003).

3. METODOLOGIA DE MONITORAMENTO DE ÁGUA

O objetivo deste trabalho é a produção de um sistema capaz de aprender o padrão de consumo de água em uma residência, detectando de forma automática possíveis problemas relacionados a vazamentos de água e/ou excesso de consumo. Esta seção encontra-se dividida da seguinte forma: i)

aquisição de dados; ii) exploração e manipulação de dados; iii) caracterização dos dados e iv) classificação.

3.1 Aquisição de dados

Para aquisição de dados, um medidor de vazão (sensor de fluxo tipo *hall*) foi instalado em uma residência, no município de São Roque, no estado de São Paulo. A residência possui uma variação populacional, onde nos dias de semana o número de residentes é de 3 pessoas e aos finais de semana de 5 pessoas. O local apresenta consumo médio mensal de 42 m³ (42.000 litros) de água. Os dados de consumo de água do sensor de vazão foram coletados por um microcontrolador (Node MCU) e enviados a um banco de dados (MySQL), por meio da internet do local (WiFi) (Kulkarni at el. 2016) (Gosavi at el. 2017). O sistema de monitoramento de consumo de água permaneceu instalado na residência durante aproximadamente 9 meses, período de março até novembro do ano de 2018. As coletas dos dados foram realizadas de hora em hora. A base de dados totalizou 6.528 leituras de consumo de água da residência.

3.2 Exploração e Manipulação dos Dados

Em primeira instância, os dados coletados são explorados e manipulados. Os dados coletados de hora em hora então, são agrupados em períodos de 24 horas (Ren at el. 2010). Para um melhor desempenho no aprendizado do padrão de consumo da residência, os dados de consumo de água foram agrupados da seguinte forma (Teng at el. 2016) (Walker at el. 2015):

- Dias da semana: o agrupamento em períodos de 24 horas possibilita a identificação de cada dia de consumo de água do local. Dessa forma, é possível identificar as segundas-feiras de consumo de água, por exemplo;
- Semanas do mês: o agrupamento dos dados em dias de semana possibilita a identificação de cada semana do mês que consumiu água. As semanas de cada mês são utilizadas como atributos para o treinamento da rede;
- Meses do ano: consequentemente, os dados agrupados em semanas são agrupados em anos, possibilitando uma visão geral anual do consumo de água no local.

3.3 Caracterização dos Dados

Por meio dos agrupamentos dos dados, foi possível a caracterização do consumo de água, possibilitando a identificação dos padrões de consumo da residência. Dessa forma, foram realizadas identificações detalhadas sobre as características do consumo de água da residência, tais como: dia da semana de maior consumo de água; horário de maior consumo de água e maior valor consumido (em metros cúbicos) no dia (Abadi at el. 2017).

O agrupamento anual mostra um comportamento de pouca dinamicidade, apresentando um consumo pouco variado ao longo do ano, onde o valor de consumo de água de cada mês encontra-se próximo ao consumo médio (Abadi at el. 2017).

3.4 Classificação

O classificador adotado foi uma RNA do tipo Perceptron Multicamada (MLP) treinada com o algoritmo da retropropagação do erro. Foi utilizada uma RNA com 3 camadas, com 100 neurônios em cada camada oculta, além da camada de entrada (4 neurônios) e camada de saída (2 neurônios). A função de ativação adotada nos neurônios foi a sigmoide logística.

O conjunto de treinamento da RNA foi elaborado a partir da caracterização das informações de consumo de água e identificação do padrão de consumo (anual, mensal e diário). Os rótulos foram manualmente inseridos por um humano. A preparação dos dados adotou a seguinte metodologia:

- Anual (período de coleta dos dados): as informações de entradas (*features*) são os consumos de cada mês. As informações de saída (*labels*) são rotuladas como **anomalia** e **normal**, onde **normal** são os meses quais apresentaram consumo menor ou igual que a média de consumo mensal e **anomalia** os meses quais apresentaram consumo maior que a média mensal;
- Mensal: as informações de entradas (*features*) são os consumos de cada dia. As informações de saída (*labels*) são rotuladas como **anomalia** e **normal**, onde **normal** são os dias quais apresentaram consumo menor ou igual que a média de consumo diário e **anomalia** os dias quais apresentaram consumo maior que a média diária;
- Semanal: as informações de entradas (*features*) são os consumos de cada dia da semana (segundas, terças, quartas, quintas e sextas-feiras, sábados e domingos), caracterizando os maiores consumidores (como identificado as quintas-feiras como as maiores consumidoras). As informações de saída (*labels*) são rotuladas como **anomalia** e **normal**, onde **normal** são os dias da semana quais apresentaram consumo menor ou igual que a média de consumo diário e **anomalia** os dias da semana quais apresentaram consumo maior que a média diária;
- Diária: as informações de entradas (*features*) são os consumos de cada dia, de hora em hora. As informações de saída (*labels*) são rotuladas como **anomalia** e **normal**, onde **normal** são os dias quais apresentaram consumo de água menor ou igual que a média de consumo diário e **anomalia** os dias quais apresentaram consumo maior que a média diária.

A Fig. 4 ilustra a estrutura da RNA utilizada. O conjunto de dados foi particionado em três conjuntos: i) treinamento, ii) validação e iii) teste, na proporção de 70:20:10, e o desempenho da rede para classificação dos dados de teste foi avaliado. O algoritmo foi implementado em ambiente Python.

Fig. 4 Estrutura da RNA.

4. RESULTADOS

4.1. Caracterização dos Dados

O consumo médio de água no local foi de 43,89 m³ (43890 litros). A caracterização do consumo anual possibilita também identificação do mês de maior consumo de água no ano. A Fig. 5 mostra o consumo de água do local no período de março a novembro, onde é possível caracterizar o mês de janeiro como o maior consumidor de água.

Fig. 5 Consumo de Água Anual: Consumo em *Metros Cúbicos (m³)* ao longo dos meses *Mês*.

Posteriormente a análise anual, o agrupamento semanal é analisado. Por meio da análise das semanas de cada mês, foi possível a identificação do dia da semana de maior consumo de água. A Fig. 6 mostra a distribuição semanal do consumo de água, apresentando o consumo de cada dia da semana ao longo dos 9 meses de coleta de dados, sendo possível a identificação de quinta-feira como maior consumidor.

Fig. 6 Consumo de Água dos Dias da Semana: Consumo em *Metros Cúbicos (m³)* e dias da semana aos longos dos meses *Dias das Semanas*.

Diante da contextualização dos dados, uma análise junto ao proprietário foi possível caracterizar as quintas-feiras como dias de limpeza na residência. Apresentar os dados ao proprietário aprimorou o treinamento da rede neural, com finalidade de melhorar o detalhamento dos dados para a realização do processo de aprendizagem. O detalhamento dos dados atribuídos ao treinamento é de suma importância para a análise do padrão de consumo da residência, diante de que se fossem utilizados apenas os dados do agrupamento anual, analisando consumo total de cada mês, o treinamento seria pobre (Abadi et al. 2017).

A visão mensal de cada mês mostra os dias de consumo de água, possibilitando a identificação dos dias de maior consumo da residência em cada mês. Cada mês apresenta um dia diferente de maior consumo no mês. Por exemplo, em março o dia de maior consumo na residência foi dia 10, um sábado, com consumo de 7,56 m³ (7560 litros). A Fig. 7 mostra o consumo de água do mês de março.

Fig. 7 Consumo de Água mês de Março: Consumo em *Metros Cúbicos (m³)* e dias ao longo do mês *Dia*.

Por outro lado, o mês de novembro apresentou dia 15 como dia de maior consumo na residência, uma quinta-feira, com o consumo de 4,2 m³ (4200 litros) de água. A Fig. 8 mostra o consumo de água no mês de novembro.

Fig. 8 Consumo de Água mês de Novembro: Consumo em Metros Cúbicos (m^3) e dias ao longo do mês Dia.

Por fim, uma análise das horas de cada mês foi realizada. A Fig. 9 mostra as horas de consumo no mês de março. Por meio da análise das horas de consumo do mês, foi possível a identificação dos horários de maior consumo de água na residência. Por exemplo, o mês de março apresentou um grande consumo de água no período vespertino, porém com pico de consumo às 22 horas.

Fig. 9 Horário de Consumo de Água em Março: Consumo em Metros Cúbicos (m^3) e horas de consumo Horas do Mês.

4.2. Classificação da RNA

A rede neural treinada apresentou acurácia de 96,77%. A Fig. 10 exemplifica o processo de classificação da rede após o treinamento com dados de uma semana. Com base nos dados de entrada, a rede classificou o desempenho da quinta-feira como “com avarias” (anormal) enquanto os demais foram classificados como “sem avarias” (normal).

Fig. 10 Caracterização do Consumo de Água: Consumo em Metros Cúbicos (m^3) e dias da semana Dia.

5. CONCLUSÕES

Por meio da rede neural, a partir da classificação/previsão de uma semana, foi possível comprovar a quinta-feira como dia anormal de consumo de água na residência. Os experimentos comprovaram que a rede neural treinada é capaz de classificar os dias que apresentam anomalias de consumo de água, podendo ser causadas por excesso de consumo de água ou avarias perante presença de um vazamento de água.

Foi possível a caracterização do consumo de água na residência no período dos 9 meses (anual), ao longo das semanas, dias e horas, identificando períodos com maiores níveis de consumo e períodos sazonais de consumo de água. Consequentemente, foi possível a contextualização dos dias que ocorriam limpezas na residência, caracterizando as quintas-feiras como maiores consumidores de água no local.

O conjunto de dados utilizado para o treinamento da rede neural foi pequeno. Ainda assim, os resultados apresentaram êxito na identificação dos dias com anomalias no consumo de água na residência. Para que o treinamento da rede neural seja feito com melhor desempenho, a quantidade de dados presentes no processo de aprendizagem deve ser maior. Portanto, o ideal seria um conjunto de dados de 24 meses, onde seria possível a análise e comparação do perfil de consumo do usuário ao longo dos anos. Ainda, a aquisição de dados de um número maior de residências permitirá a elaboração de RNAs com melhor capacidade de generalização, ou seja, capazes de permitir a identificação de padrões mais complexos de anomalia (viagens, padrões sazonais ou alternados de utilização, diferentes perfis de usuários, etc.).

Para trabalhos futuros, há a proposta de uma análise de um bairro, a fim de caracterização a necessidade de distribuição de água na região, possibilitando também a identificação de possíveis avarias da rede de distribuição, seja por excesso de consumo dos moradores, ou por vazamentos de água presentes na rede.

AGRADECIMENTOS

A empresa Yamatec, pela disponibilidade e amparo na implementação do estudo.

REFERÊNCIAS

- Abadi L. M., Samé A., Oukhellou L., Cheifetz N., Mandel P., Féliers P. e Chesneau O. (2017). Predictive Classification of Water Consumption Time Series Using Non-homogeneous Markov Models. IEEE International Conference on Data Science and Advanced Analytics (DSAA). Tóquio, Japão.
- ABES. (2015). Controle e Redução de Perdas nos Sistemas Públicos de Abastecimento de Água. São Paulo.
- Castro L.N. e Zuben, J.V. (2005). Redes Neurais Artificiais. Campinas.
- Falkenberg, A.V., Dyminski A.S. e Ribeiro, E.P. (2003). Redes Neurais Artificiais Aplicadas à Previsão de Consumo de Água. VI Brazilian Conference on Neural Networks - VI Congresso Brasileiro de Redes Neurais pp. 319–324. Centro Universitário da FEI, São Paulo.
- FUNASA. (2014). *Redução de Perdas em Sistemas de Abastecimento de Água*, 172 p.. Ministério da Saúde. Brasília.
- Gosavi G., Gawde G. e Gosavi G. (2017). Smart water flow monitoring and forecasting system. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). Bangalore, India.
- Haykin, S. (2003). Redes neurais: princípios e prática. Bookman, 2a edição.
- Hussin S.N.H.S., Malek M.A., Jaddi N.S. and Hamid Z.A. (2016). Hybrid Metaheuristic of Artificial Neural Network - Bat Algorithm in Forecasting Electricity Production and Water Consumption at Sultan Azlan Shah Hydropower Plant. IEEE International Conference on Power and Energy (PECon). Melaka, Malaysia.
- Kulkarni P. e Farnham T. (2016). Smart City Wireless Connectivity Considerations and Cost Analysis: Lessons Learnt From Smart Water Case Studies. *IEEE Access*, volume 4, pp. 660 – 672.
- Marcato, F.S., Oliveira G., Scazufca P. e Pires, R.C. (2018). Perdas de Água 2018 (SNIS 2016): Desafios para Disponibilidade Hídrica e avanço da eficiência do Saneamento Básico. São Paulo.
- Ren Y., Xing T., Chen X., E X. e Zhao T. (2010). Based on Process Neural Network Learning Algorithm for Prediction of Urban Water Consumption. International Conference on Computing, Control and Industrial Engineering. Wuhan, China.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986) Learning representations by back-propagating errors. *Nature*, 323, 533-536.
- Silva I.N., Spatti D.H. e Flauzino (2010). Redes Neurais Artificiais para engenharia e ciências aplicadas. DCA-FEEC-Unicamp. Campinas.
- Teng W. L., Wang L., Wang Y. e Chang Q. (2016). Housing price prediction using neural networks. 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Changsha, China.
- Walker D., Creaco E., Vamvakeridou-Lyroudia L., Farmani R., Kapelan Z. e Savić D. (2015). Forecasting Domestic Water Consumption from Smart Meter Readings Using Statistical Methods and Artificial Neural Networks. *Procedia Engineering*, volume 119, 2015, p. 1419-1428. Leicester, UK.
- Zambiasi, S.P. (2002). Introdução as Redes Neurais Artificiais. Florianópolis.