

CHATTERBOT COMO APOIO À TOMADA DE DECISÃO NO CONTEXTO DO GERENCIAMENTO DE RECURSOS NATURAIS

BRUNO CASCAES ALVES¹; **LARISSA ASTROGILDO DE FREITAS¹**; **MARILTON SANCHOTENE DE AGUIAR¹**

¹*Universidade Federal de Pelotas – {bcalves, larissa, marilton}@inf.ufpel.edu.br*

1. INTRODUÇÃO

O gerenciamento de recursos naturais é uma área que busca por melhores formas de gerenciar terras, águas, plantas e animais, baseado na qualidade de vida da sociedade. Esta área ganhou visibilidade dos governos devido à noção de desenvolvimento sustentável, que é um princípio de como eles veem e compreendem o mundo. A gestão de recursos naturais se dedica no estudo de recursos e em como esses recursos podem dar suporte à vida (HOLZMAN, 2007).

Neste contexto, o Aprendizado de Máquina (AM) aparece como uma forma de utilização de ferramentas de computação como apoio à previsão de riscos e auxílio na tomada de decisão em sistemas reais complexos (ADAMATTI; AGUIAR, 2012). Segundo ALPAYDIN (2010), o AM tem como objetivo designar um computador a aprender de acordo com a experiência passada, sendo definido como não-supervisionado – quando não há informação sobre as respostas corretas, e supervisionado – quando há conhecimento das saídas corretas.

Considerando a crescente presença de *chatterbots* no cotidiano, utiliza-se deste meio de comunicação para apoiar usuários em suas decisões. RAJ (2019) define *chatterbot* como uma máquina inteligente capaz de compreender solicitações e, em seguida, formular uma resposta que esteja de acordo com a pergunta. Neste cenário, aplicam-se técnicas de Processamento de Língua Natural (PLN) a fim de extrair significado das mensagens enviadas por usuários.

Este trabalho foi desenvolvido no contexto de um projeto de pesquisa em andamento com financiamento do Programa de Apoio ao Ensino e à Pesquisa Científica e Tecnológica em Regulação e Gestão de Recursos Hídricos – Pró-Recursos Hídricos Chamada 16/2017 da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Agência Nacional de Águas (ANA). Neste projeto, está sendo feita a modelagem de um jogo computacional baseado em sistemas multiagentes e jogos de papéis (RPG) para o gerenciamento de recursos naturais, mais especificamente para a gestão participativa dos recursos hídricos, considerando a Bacia Hidrográfica do São Gonçalo e da Lagoa Mirim.

Neste jogo sério, os jogadores interpretam personagens, modelada a partir de interações com o comitê de bacia da região, dentro de uma história construída a partir de regras, onde tomam decisões e realizam interações com outros agentes, a fim de buscar seus objetivos individuais e/ou coletivos.

Os agentes do jogo são divididos em três grupos: fiscalizadores, produtores e reguladores. Agentes fiscalizadores têm como tarefa fiscalizar produtores que infrinjam alguma regra imposta por reguladores. Produtores exploram o ambiente com o objetivo de obter recursos financeiros, através da utilização/venda de equipamentos e insumos. Por fim, agentes reguladores são responsáveis por gerenciar os recursos financeiros, obtidos através da aplicação de taxas, a fim de controlar os índices de poluição.

Tendo em vista este contexto, neste trabalho desenvolvemos um *chatterbot* capaz de auxiliar diferentes papéis do RPG na tomada de decisão, considerando características do ambiente e interação entre indivíduos, no cenário do

gerenciamento de recursos naturais. Desta forma, os agentes do jogo podem consultar tendências de mercado, a partir de estatísticas, e realizar previsões sobre os níveis de poluição, com base na aplicação de algoritmos de AM.

2. METODOLOGIA

Para o desenvolvimento do *chatterbot* foram definidas as seguintes etapas: i) modelagem de banco de dados; ii) construção de fluxograma conversacional; iii) elaboração de sistema utilizando técnicas de PLN e similaridade textual; iv) implementação de modelos com uso de AM; e, v) desenvolvimento de interface. A Figura 1 apresenta um diagrama com fluxo de atividades do sistema proposto.

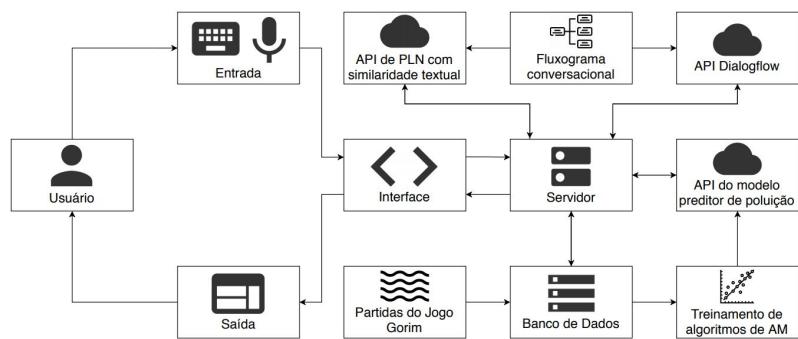


Figura 1: Arquitetura proposta para o *chatterbot*.

A partir da análise das especificações do projeto do jogo, foi modelado um banco de dados SQL para armazenamento de registros gerados durante simulações de partidas do jogo ocorridas em seis sessões, entre 2019 e 2020. Estes registros foram coletados no formato de *log*, diretamente do motor do jogo e, portanto, necessitando uma reestruturação dos dados, previamente à inserção no banco de dados.

Considerando a modelagem do jogo proposta, elaborou-se um fluxograma conversacional com base em questões relevantes a cada tipo de personagem (agricultor, empresário, fiscal, prefeito e vereador). Os diálogos construídos foram implementados no *Dialogflow* (RAJ, 2019), plataforma de criação de *chatterbots*.

Buscando expandir o entendimento do robô, implementou-se um sistema baseado em PLN e similaridade entre frases, o qual se procura a intenção associada à mensagem enviada pelo usuário. Para isto, foi necessário um conjunto de frases de comparação, sendo este retirado dos diálogos construídos.

As sentenças são processadas com apoio de técnicas de PLN, como: normalização, tokenização, remoção de *stopwords* e lematização. Assim, após as aplicações, o documento se encontra escrito com letras minúsculas, onde cada frase é organizada em um conjunto de palavras relevantes em seu formato básico. Para a frase enviada pelo usuário, verifica-se a existência de seus *tokens* em um vocabulário obtido com base nas frases de comparação. Desta forma, caso determinada palavra não conste no vocabulário, uma pesquisa é realizada em um dicionário de sinônimos a fim de buscar por uma equivalência.

Por fim, as palavras das sentenças são mapeadas em vetores de números reais de baixa dimensão através de representação com *Word Embedding*. Para este trabalho foi utilizado o algoritmo não-supervisionado *GloVe* de PENNINGTON (2014), o qual tem como princípio extrair significado baseado nas proporções das probabilidades de co-ocorrências das palavras e de características globais do corpus. Assim, após a conversão para estes vetores,

utilizou-se a similaridade de cosseno para determinar a frase mais similar à frase de entrada.

Considerando o sistema preditor de poluição, foram utilizados registros armazenados no banco de dados durante o treinamento. Estes registros se referem a cerca de 2800 ações ocorridas no jogo, sendo estas rearranjadas em 1050 entradas. Tendo em vista que existem diferentes papéis no RPG, houve a necessidade de tratar valores faltantes, corrigido-os com zeros. Ainda, foi executada conversão de dados categóricos em numéricos e normalização.

A separação dos dados foi realizada através de validação cruzada *StratifiedKFold*, a fim de avaliar a capacidade de generalização. A partir deste método, são realizadas dez iterações com os dados divididos em dez grupos, onde a cada iteração uma destas partes se refere ao conjunto de teste e o restante ao de treinamento. Ainda, foi considerada a execução de trinta ensaios.

Levando em conta que a classe poluição é um valor real, faz-se necessário a utilização de regressão. Para determinar o melhor modelo, foram comparados quatro algoritmos regressores: regressão linear – equação que representa a relação entre uma variável dependente e um conjunto de atributos; regressão de vetores de suporte (SVR) – margem máxima que separa o hiperplano; árvore de regressão – regras com base em atributos previsores; e, por fim, *random forest regression* – média da combinação de árvores de regressão.

As implementações foram realizadas em *Python*, utilizando a biblioteca *Scikit-Learn*. Os parâmetros dos algoritmos se referem ao padrão proposto na biblioteca, exceto no que diz respeito ao número de árvores do *random forest* e à regularização do SVR linear, sendo definidos como 100 e 10, respectivamente.

Na última etapa, foi desenvolvida uma interface web capaz de viabilizar a troca de mensagens. Considerando a crescente utilização de áudio como forma de comunicação, a interface disponibiliza aos usuários este meio, além de texto. Para as mensagens enviadas pelo chatterbot, são utilizados textos e gráficos.

3. RESULTADOS E DISCUSSÃO

Foram gerados quatro modelos de regressão para poluição. Para definir a melhor performance entre os algoritmos, foram aplicadas as seguintes medidas: Erro Absoluto Médio (MAE – *Mean Absolute Error*) e Erro Quadrático Médio (MSE – *Mean Squared Error*). O MAE se refere ao cálculo da média do erro absoluto ocorrido entre os valores previstos e medidos. De mesma forma, o MSE calcula a média dos erros, porém os eleva ao quadrado. Para estas medidas, quanto menor o valor obtido, mais próximo a regressão representa o modelo com precisão.

A Tabela 1 apresenta os resultados para os algoritmos utilizados. Ao refletir sobre estas medidas, destaca-se os melhores resultados de forma geral na seguinte ordem: regressão linear, regressão de vetores de suporte, *random forest regression* e árvore de regressão. Desta maneira, pode-se constatar que o modelo de regressão linear foi o que melhor se adequou aos dados.

Tabela 1 – Resultado das métricas aplicadas nos modelos de poluição.

Algoritmo	MAE	MSE
Regressão Linear	8,286	909,28
Regressão de Vetores de Suporte	7,503	975,672
Árvore de Regressão	11,594	2358,83
<i>Random Forest Regression</i>	9,051	1381,447

Em relação ao *chatterbot*, também foram realizados testes conversacionais, conforme exemplos apresentados na Figura 2. Tendo em vista que a implementação do jogo está em fase de finalização, ainda não foi realizada integração como o modelo preditor de poluição e, também, pesquisa de opinião.



Figura 2: Exemplos de diálogos produzidos pelo *chatterbot*.

4. CONCLUSÕES

Para este trabalho, foram realizadas diversas aplicações para a construção do *chatterbot*. Deste modo, quatro algoritmos de AM foram treinados com diferentes bases matemáticas, sendo utilizados registros armazenados no banco de dados construído. Considerando os resultados, considera-se satisfatório o modelo de regressão linear como preditor de poluição.

Os diálogos desenvolvidos serviram para o sistema de compreensão da língua natural. Nesta aplicação foram utilizadas técnicas de PLN e a representação *GloVe* associada com a similaridade de cosseno. Como trabalhos futuros, pretendemos comparar outros métodos de representação textual.

No contexto do gerenciamento de recursos hídricos, esta pesquisa continua sendo um importante tópico de estudo, pois a partir da interação entre jogadores em jogos sérios é possível analisar o comportamento humano e apoiar suas decisões, através da simulação do gerenciamento de recursos naturais.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMATTI, D. F; AGUIAR, M. S. de. Computational Techniques to Manage Natural Resources. *Int. J. Organ. Collect Intell.* USA, v.3, n.2, p.117-137, 2012.
- ALPAYDIN, E. **Introduction to Machine Learning**. Cambridge: The MIT Press, 2010. 2v.
- HOLZMAN, B. A. **Natural Resource Management**. 2007. Acessado em 14 set. 2020. Online. Disponível em: <https://bit.ly/3c0dsia>
- PENNINGTON, J.; SOCHER, R.; MANNING, C. *GloVe: Global Vectors for Word Representation*. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2014., Doha, 2014. **Proceedings...** Association for Computational Linguistics, 2014. p.1532–1543.
- RAJ, S. **Building Chatbots with Python**. Berkeley: Apress, 2019. 1v.