Projeto PIBIC

Adaptação do algoritmo de aprendizado online K-NN e SAM-KNN para permitir o direito ao esquecimento

Orientador:
Flavia Cristina Bernardini
SIAPE: **1671775**
Lotação: **Departamento de Ciência da Computação – Instituto de Computação**

Aluno:
Ariadne Gonçalves Pinheiro
Matrícula: **119071114**

Local onde será desenvolvido: **Instituto de Computação**

Grande Área: **Ciências Exatas e da Terra**
Área: **Ciência da Computação**
Subárea: **Metodologia e Técnicas de Computação**

Resumo

O aprendizado de máquina online, ou stream learning, tem se mostrado efetivo em diversos problemas nos quais fluxos de dados estão presentes, como detecção de invasão em redes, previsão de demanda de energia, dentre outros. No entanto, com as leis sendo divulgadas internacionalmente que garantem o direito ao esquecimento de dados, não somente os dados precisam ser excluídos, mas os modelos que utilizam tais dados precisam ser adaptados para esquecer os dados. O objetivo deste trabalho é adaptar os algoritmos de aprendizado online K-NN e SAM-KNN para permitir o direito ao esquecimento. Este projeto faz parte de um projeto maior, que tem por objetivo estudar os efeitos do direito ao esquecimento no aprendizado de máquina, tanto em lote como online.

1. Introdução e Justificativas

O Processamento de Dados (PD) é um dos passos importantes em Aprendizado de Máquina (WITTEN et al, 2016). O PD como um fluxo de dados constitui uma maneira alternativa de lidar com grandes quantidades de dados. Uma das abordagens de aprendizado de máquina mais pesquisadas nos últimos anos é a que se baseia em construção de modelos adaptativos, denominada aprendizado de máquina online ou *stream learning*, para dados em fluxo em tempo real (FACELI et al, 2011). Tipicamente, esses modelos são aplicados principalmente em cenários onde a distribuição dos dados acontece de forma não-estacionária e que, consequentemente, na maior parte dos casos ocorrem mudança de conceito (do inglês *concept drift*, oriundo da mudança da distribuição dos dados ao longo do tempo (HOLZINGER et al., 2017). Previsão de demanda de energia (FERNANDES, 2019), detecção de invasão em redes (ALVES et al, 2020), dentre outros, são exemplos de problemas em que aprendizado de máquina online tem mostrado bons resultados. Em aprendizado online, um dos tipos de problemas presentes relacionados a mudança de conceito é a detecção do esquecimento de conceito (*concept forgetting*) (BIFET et al, 2018). O esquecimento de conceito surge a partir da evolução do conceito, onde estes ficaram desatualizados e não são mais relevantes para o domínio de destino. Esses conceitos exigem um mecanismo adaptativo para o modelo esquecer conceitos não utilizados.

No problema de detecção do esquecimento do conceito, um exemplo de cenário é o Direito de Ser Esquecido – DSE ("The Right to Be Forgotten") (VILLARONGA; KIESEBERG; TIFFANY, 2018), que é essencialmente o conceito de que os indivíduos têm o direito de solicitar que seus dados (coletados por outros) sejam excluídos. Inclusive, no Brasil foi sancionada a Lei Geral de Proteção de Dados (PRESIDÊNCIA DA REPÚBLICA DO BRASIL, 2018), que também prevê o DSE.

Embora a exclusão de dados possa parecer um tópico direto do ponto de vista de muitos reguladores, essa questão aparentemente simples apresenta muitos problemas práticos em ambientes reais de aprendizado de máquina. Assim, os modelos de inferência presentes nestes sistemas precisam de estratégias para esquecer esses dados que devem ser excluídos. O problema com o DSE e sua inaplicabilidade à Inteligência Artificial (IA) pode ser devido ao nosso entendimento impreciso da privacidade em relação à IA (VILLARONGA;

KIESEBERG; TIFFANY, 2018). Muitas vezes, as pessoas veem a privacidade como escondendo suas informações de outras pessoas. Isso é especialmente aparente ao examinar o princípio do DSE, sob o qual os indivíduos podem solicitar que as informações tornadas públicas sejam excluídas (e, portanto, tornadas privadas). No caso de informações públicas tornadas privadas, a ideia da mente humana esquecendo uma parte da informação se aplica bem. Quando indivíduos tornam privadas as informações anteriormente públicas, elas solicitam que outras pessoas esqueçam essas informações. No entanto, essa ideia é exclusiva apenas da mente humana e não se traduz necessariamente na era da IA/ aprendizado de máquina. Para entender o direito de ser esquecido no contexto da inteligência artificial, é necessário primeiro mergulhar em uma visão geral dos conceitos de memória e esquecimento humano e da IA. Nossa lei atual parece tratar a memória humana e de máquina da mesma maneira, apoiando um entendimento fictício da memória e esquecendo que não se encaixa na realidade.

No entanto, a variedade de ferramentas ou técnicas disponíveis de DSE neste contexto ainda é escassa atualmente. Em cenários de DSE, a abordagem dominante tipicamente é refazer o treinamento de modelos em lote. Neste cenário, havia a necessidade de armazenamento de todos os registros para a geração de novos modelos. Malle et al. (2016) usa técnicas de clusterização para a manutenção dos sistemas. Villaronga, Kieseberg e Tiffany (2018) utiliza técnicas de k-anonimato em banco de dados, para a proteção e exclusão de informações. Para o domínio de aprendizado online, Mirzasoleiman, Karbasi e Krause (2017) propôs um algoritmo para a exclusão de informações em modelos online para a tarefa de recomendação de notícias. Essa abordagem apresenta resultados robustos, mas não foi aplicado em diferentes cenários para validação deste técnica. No entanto, em nenhum dos trabalhos encontrados, há uma disponibilidade de algoritmos de aprendizado de máquina que incorporem tais técnicas de esquecimento, facilitando o uso dos algoritmos de aprendizado de máquina neste contexto.

Em particular, no contexto de *stream learning*, o algoritmo de aprendizado SAM K-NN, proposto por Losing, Hammer e Wersing (2016), apresenta um modelo de memória de autoajuste (*Self Adjusting Memory* – SAM) para o algoritmo k-NN (*k-Nearest Neighbor*, ou k-vizinhos mais próximos). Isso é devido ao k-NN constituir um algoritmo de classificação com bons resultados em *stream learning*. O SAM-kNN pode lidar com mudanças heterogêneas de conceito, ou seja, diferentes tipos de mudanças e com diferentes frequências. Além disso, o SAM K-NN utiliza modelos de memória de curto e longo prazo, que tem inspiração biológica inclusive nas questões de coordenação dessas memórias. O SAM K-NN pode ser facilmente aplicado na prática, não é necessária a otimização dos seus meta-parâmetros não é necessária. A ideia por trás do algoritmo é construir modelos dedicados para cobrir conceitos atuais e antigos, e aplicá-los de acordo com as demandas dos fluxos de dados. Essas estratégias de manipulação das memórias podem ser adaptadas não somente para atualizar os conceitos com os novos dados apresentados no fluxo, mas também para remover dados da memória.

2. Objetivos

O objetivo geral deste trabalho é adaptar o algoritmo K-NN e SAM K-NN para stream learning para permitir o esquecimento de dados, conforme vem sendo exigido por leis de privacidade de dados.

Os objetivos específicos para o bolsista são:

- Aprender conceitos de aprendizado em lote, ou seja, quando todos os dados estão disponíveis para aprendizado de um conceito, e online, quando os dados devem ser utilizados conforme são apresentados, e o conceito é adaptado com o passar do tempo;
- Modificar o algoritmo K-NN e SAM K-NN para esquecimento dos dados;
- Implementar a modificação do algoritmo no Multiflow, e disponibilizá-lo na ferramenta;
- Avaliar o algoritmo com conjuntos de dados existentes para aprendizado online.

3. Metodologia e forma de análise dos resultados

Para a execução deste projeto, serão utilizadas as ferramentas Scikit Multiflow[1] e Scikit Learn[2], implementadas em Python. O Scikit learn tem sido amplamente utilizado para atividades que envolvam o uso de aprendizado de máquina pois seu código fonte é aberto, permitindo tanto o seu reuso quanto o seu uso como biblioteca em sistemas que requerem o uso de aprendizado de máquina. Além disso, a prototipação em Python é uma característica bastante interessante para validação de conceitos. O Scikit Multiflow foi inspirado no MOA (BIFET et al, 2018), e utiliza o Scikit Learn como base. Assim como o Scikit Learn, o Scikit Multiflow também foi implementado em Java e seu código fonte também é aberto.

Os algoritmos K-NN e SAM K-NN estão implementados no Scikit Multiflow. Essa versão será utilizada como base para adaptação dos algoritmos, de forma a permitir que o esquecimento dos dados possa ser realizado. Deve ser observado que ambas as ferramentas são baseadas em Frameworks de aprendizado de máquina, que envolvem os passos de pré-processamento dos dados, aprendizado e avaliação dos modelos (conceitos) construídos, o que requer uma curva de aprendizado por parte do estudante de graduação doss framework. Assim sendo, é importante observar que a adaptação do algoritmo não é uma tarefa trivial, pois envolve o estudo do impacto da remoção de dados do conceito do ponto de vista estatístico.

Os resultados serão analisados utilizando técnicas estatísticas de amostragem de dados para avaliação de modelos de aprendizado de máquina online, os quais são *holdout* e *prequential*, conforme apresentado por (BIFET, 2018). Essas técnicas levam em consideração o fato que os dados chegam em fluxo de dados, apresentando aspecto temporal, o que deve ser respeitado no processo de avaliação. Além disso, serão utilizados datasets comumente utilizados para

[1] Disponível em https://scikit-multiflow.github.io/
[2] Disponível em https://scikit-learn.org/stable/

avaliação de modelos construídos com algoritmos de aprendizado online. Em https://github.com/vlosing/driftDatasets estão disponíveis os conjuntos de dados utilizados por Losing, Hammer e Wersing (2016), utilizados para avaliar o SAM K-NN, que também serão utilizados neste trabalho.

Este projeto faz parte de um projeto de maior abrangência, que tem por objetivo estudar os efeitos do direito ao esquecimento no aprendizado de máquina. O aluno de doutorado por mim coorientado, Leandro Botelho, também atua nas discussões deste trabalho, pois auxilia no seu projeto de pesquisa na questão de evolução de modelos preditivos. Além disso, este projeto de maior abrangência está sendo realizado em parceria com o Prof. Albert Bifet, da Universidade Paris-Tech, autor do livro Machine Learning for Data Streams with Practical Examples in MOA (BIFET et al, 2018).

Potenciais veículos de apresentação de resultados:

Conferências internacionais de aprendizado de máquina e/ou inteligência artificial;
Periódicos internacionais de aprendizado de máquina e/ou inteligência artificial.

Deve ser observado que a orientadora tem publicado internacionalmente trabalhos oriundos de graduação. Os mais recentes, relacionados ao tema deste trabalho, foram (FAIAL et al 2019) e (ALVES et al 2020).

Local de Execução da Pesquisa:

O projeto será executado no Instituto de Computação, que apresenta uma boa infra-estrutura de laboratórios, incluindo o LabESI (Laboratório de Engenharia de Sistemas de Informação). O LabESI conta com diversos equipamentos adquiridos com diversos projetos aprovados por órgãos de fomento, como a FAPERJ e CNPq, e parcerias com instituições privadas, como a Petrobras.

4. Plano de trabalho do bolsista e cronograma de atividades

As seguintes atividades serão realizadas pelo bolsista:

- Aprender conceitos de aprendizado em lote, ou seja, quando todos os dados estão disponíveis para aprendizado de um conceito, e online, quando os dados devem ser utilizados conforme são apresentados, e o conceito é adaptado com o passar do tempo;
- Familiarizar com o framework que possui a implementação dos algoritmos K-NN e SAM K-NN - o Scikit Multiflow, inspirado no MOA (BIFET et al, 2018);
- Adaptar os algoritmos K-NN e SAM K-NN para esquecimento dos dados, conforme as teorias de esquecimento apresentadas na literatura;
- Implementar a modificação do algoritmo no Scikit Multiflow, e disponibilizá-lo na ferramenta;
- Avaliar o algoritmo com conjuntos de dados existentes para *stream learning*.

Na Tabela 1, é mostrado o cronograma das atividades a serem realizadas pelo

bolsista.

Tabela 1 - Cronograma de Atividades

No.	ATIVIDADES	1	2	3	4	5	6	7	8	9	10	11	12
1	Estudo de referencial teórico de aprendizado de máquina em lote e online	X	X										
2	Estudo de trabalhos da literatura que discutem o esquecimento de dados no contexto de aprendizado de máquina online (*stream learning*)		X	X	X	X	X	X					
3	Estudo das ferramentas Scikit Learn e Scikit Multiflow				X	X	X	X					
4	Adaptação dos algoritmos K-NN e SAM K-NN					X	X	X	X				
5	Avaliação dos algoritmos K-NN e SAM K-NN considerando dados mais recentes e mais antigos e os parâmetros dos algoritmos							X	X	X	X	X	
6	Escrita de artigo científico					X	X	X	X	X	X	X	X

5. Referências bibliográficas

ALVES, A.; BERNARDINI, F.; SOUSA, L.; MITAAC, E. (2020) Evaluating the behavior of stream learning algorithms for detecting invasion on wireless networks. International Journal of Security and Networks. In Press.

BIFET, A.; GAVALDA, R.; HOLMES, G.; PFAHRINGER, B. (2018) Machine Learning for Data Streams with Practical Examples in MOA. MIT Press.

FACELI, K.; LORENA, A.C.; GAMA, J.; CARVALHO, A. (2011) Inteligência Artificial: Uma abordagem de aprendizado de máquina. LTC.

FAIAL, D.; BERNARDINI, F.; MIRANDA, L.; VITERBO, J. (2019) Anomaly Detection in Vehicle Traffic Data Using Batch and Stream Supervised Learning. In: EPIA Conference on Artificial Intelligence. Lecture Notes in Computer Science, vol 11804. Springer, Cham.

FERNANDES, L. (2019) Predição de Consumo de Energia Utilizando Aprendizado Online. Monografia de Conclusão de Curso de Graduação – Bacharelado em Ciência da Computação, Universidade Federal Fluminense.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, Vol. 11, N. 1, pgs. 10-18.

HOLZINGER, A. et al. (2017) Machine learning and knowledge extraction in digital pathology needs an integrative approach. In: Holzinger A., Goebel R., Ferri M., Palade V. (eds) Towards Integrative Machine Learning and Knowledge Extraction. Lecture Notes in Computer Science, vol 10344. Springer, Cham.

FRANK, E.; HALL, M.A.; WITTEN, I.H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

LEMAIRE V.; SALPERWYCK C.; BONDU A. (2015) A Survey on Supervised Classification on Data Streams. In: Zimányi E., Kutsche RD. (eds) Business Intelligence. eBISS 2014. Lecture Notes in Business Information Processing, vol 205. Springer, Cham.

LOSING, V.; HAMMER, B.; WERSING, H. (2016) KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift. In: Proc. 2016 IEEE Int. Conf. on Data Mining.

MALLE, B.; KIESEBERG, P.; WEIPPL, E.; HOLZINGER, A. (2016). The right to be forgotten: towards machine learning on perturbed knowledge bases. In International Conference on Availability, Reliability, and Security (pp. 251-266). Springer, Cham.

MATUSZYK, P.; VINAGRE, J.; SPILIOPOULOU, M.; JORGE, A. M.; GAMA, J. (2018). Forgetting techniques for stream-based matrix factorization in recommender systems. Knowledge and Information Systems, 55(2), 275-304.

MIRZASOLEIMAN, B.; KARBASI, A.; KRAUSE, A. (2017). Deletion-robust submodular maximization: Data summarization with the right to be forgotten. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 2449-2458). JMLR.org.

POLITOU, E.; ALEPIS, E.; PATSAKIS, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. Journal of Cybersecurity, Vol. 4, N. 1.

PRESIDÊNCIA DA REPÚBLICA DO BRASIL. (2018) Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acessado em 05/01/2020. /

ULLAH, M. M.; ORABONA, F.; CAPUTO, B. (2009). You live, you learn, you forget: Continuous learning of visual places with a forgetting mechanism. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 3154-3161).

VILLARONGA, E. F.; KIESEBERG, P.; LI, T. (2018). Humans forget, machines remember: Artificial intelligence and the right to be forgotten. Computer Law & Security Review, 34(2), 304-313.

VILLARONGA, E. F.; KIESEBERG, P.; TIFFANY L. (2018) "Humans forget, machines remember: Artificial intelligence and the right to be forgotten." Computer Law & Security Review, Vol 34, N. 2, pgs 304-313.

WITTEN, I.; FRANK, E.; HALL, M.A.; PAL, C.J. (2016) Data Mining: Practical Machine Learning Tools and Techniques, 4th edition. Morgan Kaufmann.