

Identificação de Comportamento Anormal em Atividades da Vida Diária do Idoso Através de Algoritmos de Detecção de Novidades

Maurício de Freitas
UNIVALI –SC, Brasil
mauriciopasetto@edu.univali.br

Vinicius Piai
UNIVALI –SC, Brasil
vinicius_piai@edu.univali.br

Anita Fernandes
UNIVALI –SC, Brasil
anita.fernandes@univali.br

Gabriel de Souza
UNIVALI –SC, Brasil
souza.gabriel@edu.univali.br

Adriana de Freitas
UNIVALI –SC, Brasil
adrianawegener@gmail.com

Gabriel Andonini
UNIVALI –SC, Brasil
gabriel.andonini@edu.univali.br

Wemerson Parreira
UNIVALI –SC, Brasil
parreira@univali.br

ABSTRACT

In old age, a series of common health conditions, chronic diseases, and disabilities affect the individual's physical and mental health and prevent the performing of Activities of Daily Living. This paper presents a solution to identify abnormalities in the behavior of the elderly based on ADL (Activities of Daily Living), using Machine Learning, through the Novelty Detection technique. The ADL data were used to create a model that defines the baseline behavior of the elderly, and new observations, to verify significant changes in behavior, are classified as discrepant or abnormal. The Local Outlier Factor, One-class Vector Machine, Robust Covariance, and Isolation Forest novelty detection algorithms were used and evaluated. The model presented reached an accuracy and F1-Score of 96%.

KEYWORDS

Novelty Detection, Activities of Daily Living, Machine Learning.

1 INTRODUÇÃO

De acordo com estimativas da OMS, o número de pessoas com mais de 60 anos quase dobrará, representando um desafio para que países de todo o mundo, propiciem ambientes físicos e sociais, que permitam um envelhecimento saudável e inclusivo [1]. Para o idoso, o envelhecimento traz uma série de desafios, sejam de natureza física, social, econômica, ou de saúde. Algumas condições de saúde são comuns a esta fase da vida, dentre elas as doenças crônicas, as incapacidades e deficiências [2]. Estas condições, podem impedir que o idoso realize as atividades da vida diária (ADL, Activities of Daily Living) de forma independente [3]. No contexto doméstico, a dificuldade de realizar tais tarefas, podem ser minimizadas por meio de tecnologias de Ambiente de Vida Assistida (Ambient Assisted Living, AAL). Estas tecnologias têm por objetivo, o desenvolvimento de sistemas de saúde pessoal e de monitoramento remoto [4]. É essencial o monitoramento das ADL para identificar anomalias na realização destas tarefas, pois mudanças significativas na rotina normal do idoso, podem

representar declínios na saúde física ou mental. Esta detecção permite que a rede de apoio do idoso possa estar ciente destas mudanças, para que seja possível identificar a necessidade de intervenção [5].

No contexto de Aprendizado de Máquina (ML, Machine Learning), uma das formas de identificar anomalias, é por meio da Detecção de Novidades (ND, Novelty Detection). O uso desta técnica está relacionado à detecção de eventos raros e anômalos. Esta técnica é utilizada quando existem dados de treinamento para a classe alvo, entretanto os dados da outra classe são pouco amostrados ou inexistentes. No contexto deste trabalho, utiliza-se dados de ADL disponíveis para criar uma linha base de comportamento do idoso, e as novas observações para verificar variações significativas que são classificadas como discrepantes ou anômalas [6], a fim de apresentar uma solução que identifique anomalias no comportamento diário do idoso, por meio de Detecção de Novidades, através do menor número possível de semanas de dados de treinamento a fim de alcançar uma acurácia maior que 90%.

Para compor a solução, foram avaliados os algoritmos LOF (Local Outlier Factor), OC-SVM (One-class Support Vector Machine) e RCE (Robuste Covariance Estimator). Assim como, avaliados os atributos temporais para uma maior generalização e o número mínimo de semanas de dados de treinamento.

A seguir serão apresentados os Algoritmos de Detecção de Novidades, a metodologia utilizada, os resultados, e as conclusões.

2 ALGORITMOS DE DETECÇÃO DE NOVIDADES

Para um maior entendimento do trabalho, esta seção aborda os conceitos relacionados aos algoritmos de ND utilizado neste trabalho. LOF [7] é uma abordagem não supervisionada baseada no conceito de densidade local, que é estimada observando os k vizinhos mais próximos. O OC-SVM trabalha com o conceito de limite de decisão entre classes, por meio da criação de hiperplanos e margens associadas. No entanto, no problema de detecção de

anomalias, não existem dados rotulados, o OC-SVM considera que todos os dados pertencem a classe normal [7].

RCOV é um algoritmo baseado em distâncias, que utiliza medidas geométricas para definir a proximidade entre as amostras de dados no espaço multidimensional. Os pontos mais distantes que a maioria, são considerados anômalos [7].

O iForest se baseia na ideia de que, ao definir um conjunto de dados, consegue-se, a partir dos dados mais centrais do *cluster*, organizar uma árvore binária baseada na proximidade entre os *clusters*. Cada nodo recebe um *score*, quantificando sua proximidade em relação ao centro do *cluster*. *Scores* menores representam comportamentos mais comuns ao conjunto de dados, e a posição do *cluster* na árvore será em direção aos nodos folha. Já as anomalias terão *scores* maiores e a tendência será de ficarem mais próximas ao nó raiz da árvore [7].

3 METODOLOGIA

A Figura 1 apresenta o fluxo da solução para identificação de anomalias diárias em ADL, usando N_i semanas (de 2 a 20), para o treinamento de M_i modelos (LOF, iForest, OC-SVM e RCE), que são avaliados para obter uma acurácia maior que 90%. A base de dados utilizada foi a eHealth Monitoring Open Data Project [8], utilizando o *dataset* monitoringPIHRversion. Este *dataset*, apresenta ADL de um idoso no contexto doméstico no período de 365 dias, sobre o perfil de dependência 1 (um) de Functional Autonomy Measurement System. Este perfil, corresponde a um indivíduo autônomo com algum nível de supervisão e ajuda necessários [9]. Dentre as 21 atividades observadas neste *dataset*, selecionou-se 9 atividades de forma arbitrária: Change clothes, Eating, Make coffee, Make hot food, Make sandwich, Sleep, Toileting, Washing - Take shower, e Watching TV.

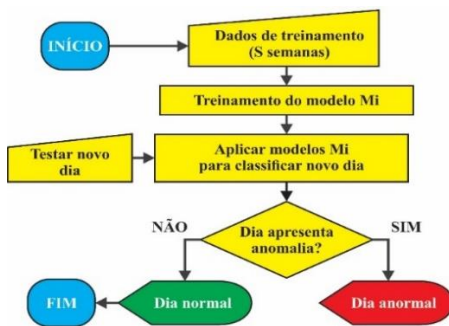


Figura 1: Fluxo de Detecção de Anomalias.

3.1 Pré-processamento

O pré-processamento do *dataset* consistiu em duas etapas: a criação de um novo *dataset*, agrupando e resumizando as atividades por dia. Foram criados os seguintes atributos: dia, atividade, o atributo duração (um somatório dos tempos de duração da de cada atividade, realizada no dia), o atributo frequência (com base no número de vezes que cada atividade foi realizada durante o dia), o atributo duração/frequência (sendo esta, a razão entre a duração e frequência). A segunda etapa consistiu em: criado o *dataset* de treinamento, validar e testar, a partir do pivoteamento do *dataset*

resultante da primeira etapa, utilizando o atributo atividade como coluna e duração/frequência como conteúdo, mantendo a coluna dia. O resultado pode ser visto na Tabela 1.

Tabela 1: Dataset resultante do pré-processamento. Parcial.

dia	Change Clothor	Eating	...	Watching TV
1	7.327778	36.725000	...	86.383333
2	6.900000	41.161111	...	47.325000
...
363	7.827778	40.405556	...	77.675000
364	7.516667	39.394444	...	53.845833

3.2 Dados de treinamento, validação e teste

Para o treinamento dos modelos foi utilizado o protocolo *holdout*, separando os dados em treinamento, validação e teste. Utilizou-se as 20 primeiras semanas para os dados de treinamento; as 8 seguintes semanas para dados de validação; e para o teste foram utilizados os mesmos dados de validação adicionando anormalidades artificiais. Estas anormalidades foram adicionadas aos dados de teste obedecem às seguintes regras: para as primeiras 2 semanas, foi selecionado de 1 a 9 atividades, aplicando um acréscimo de 20 a 200% sobre os valores; da mesma forma para as próximas 2 semanas, adicionado um decréscimo de 20 a 200%; e para as 4 semanas restantes os dados não foram modificados. Mantendo assim a proporção entre dias com e sem anormalidade. Neste processo os valores negativos foram substituídos por zero.

4 RESULTADOS OBTIDOS

O desempenho dos modelos é apresentado na Figura 3. Pode-se observar que os modelos LOF e RCOV obtiveram um desempenho superior dentre os quatro modelos. Os modelos OC-SVM e iForest apresentam desempenhos similares, porém inferiores aos demais. Entretanto, o OC-SVM apresentou um menor desempenho dentre os modelos observados. Importante observar que, os modelos LOF e RCOV apresentaram acurácias e F1-Score superiores a 90% a partir da primeira e segunda semana respectivamente. Já os modelos iForest e OC-SVM não conseguiram obter acurácia e F1-Score maiores que 90%, limitados a 87% e 83% de acurácia.

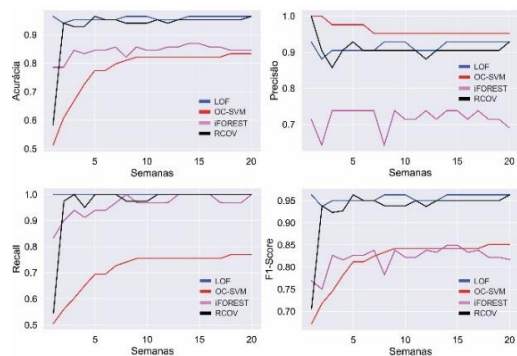


Figura 3: Desempenho dos modelos LOF, OC-SVM, iForest e RCOV.

A Tabela 2 apresenta o desempenho dos modelos, na semana em que a acurácia foi maior que 90%, para os modelos que não ultrapassaram este valor, é apresentado a semana em com melhor desempenho. Observando estes valores, pode-se afirmar que o modelo LOF obteve o melhor desempenho dentre os modelo, com apenas uma semana de dados de treinamento.

Tabela 2: Desempenho dos modelos.

modelo	Semanas trein.	acurácia	precisão...	Recall	F1-Score
LOF	1	0,96	0,93	1	0,96
RCOV	2	0,94	0,90	0,97	0,94
iForest	14	0,87	0,74	1	0,85
OC-SVM	18	0,83	0,95	0,77	0,85

4.1 Modelo Proposto

Levando em consideração os resultados apresentados na seção anterior, para o modelo proposto foram selecionados: os algoritmo LOF, que apresentou melhor desempenho dentre os modelos; uma semana para o número mínimo de dados de treinamento alcançando uma acurácia de 96% e F1-Score de 93%. Também se considera o algoritmo RCOV, como algoritmo viável para a detecção de anomalias com um baixo número de semanas de dados de treinamento. Os algoritmos iForest e OC-SVM foram descartados por não terem alcançado 90% de acurácia e F1-Score, mesmo com uma quantidade maior de semanas de treinamento.

5 CONCLUSÕES

Este trabalho teve por objetivo criar uma solução para identificar anormalidade em ADL do idoso no contexto doméstico com o uso de Detecção de Novidades. Buscando uma acurácia e F1-Score maior que 90% com o menor número possível de dados de treinamento.

Os resultados mostram que os algoritmos LOF e RCOV alcançaram valores de acurácia e F1-Score acima de 90% a partir da primeira semana de dados de treinamento respectivamente. Os algoritmos iForest e OC-SVM foram descartados pelo baixo desempenho.

Para o modelo de detecção de anomalias, foram selecionados: o algoritmo LOF; uma semana para dados de treinamento. Este modelo alcançou 96% de acurácia e 93% de F1-Score. Desta forma, considera-se a Detecção de Novidade um método eficaz para detecção de anormalidades em ADL do idoso no contexto doméstico.

Como trabalhos futuros, considera-se aplicar os experimentos deste estudo em outras bases de dados de ADL, mesmo que em diferentes contextos. Assim como, abordar outras técnicas para a detecção de anomalias.

AGRADECIMENTOS

Agradecemos a FAPESC pelo fomento a este projeto através das chamadas: Fapesc nº 15/2021 - 2021TR001236 - Programa de Ciência, Tecnologia e Inovação de Apoio aos Grupos de Pesquisa da Associação Catarinense das Fundações Educacionais - ACAFE; Fapesc nº 29/2021 - 2021TR001758 - Programa Estruturante Acadêmico de Apoio à Infraestrutura de Laboratórios do Estado de Santa Catarina; e Fapesc nº 21/2021 - Programa de Apoio aos Programas de Pós-graduação Emergentes e em Consolidação em Áreas Prioritárias nos Estados. Também agradecemos a CAPES pelo fomento através do Edital Nº 58/2022/PPGIELA - Concessão De Bolsas - Programa De Desenvolvimento Da Pós-graduação Estratégico De Consolidação - Capes - Programa De Pós-graduação Interdisciplinar Em Estudos Latino-americanos.

REFERÊNCIAS

- [1] OMS – Organização Mundial da Saúde, 2021. Ageing and Health. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>>. Acessado em: 25 de março de 2022.
- [2] NCOA - National Council on Aging, 2021. Chronic Conditions for Older Adults. Disponível em: <<https://www.ncoa.org/article/the-top-10-most-common-chronic-conditions-in-older-adults>>. Acessado em: 25 de março de 2022.
- [3] JCHS – Join Center for Housing Studies of Harvard University, 2016. Disponível em: <<https://www.jchs.harvard.edu/research-areas/reports/projections-and-implications-housing-growing-population-older-households>>. Acessado em: 25 de março de 2022.
- [4] Gomathi, R. M., Nithya, M. Human activity recognition with smartphones using machine learning process. International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-10, doi: 10.1109/ICCCI54379.2022.9740773.
- [5] Arifoglu, D., Bouchachia, A. *Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. Artificial intelligence in medicine*, v. 94, p. 88-95, 2019, doi: 10.1016/j.artmed.2019.01.005.
- [6] Yahaya, S. W., Lotfi, A., Mahmud, M. *Detecting anomaly and its sources in activities of daily living. SN Computer Science*, v. 2, n. 1, p. 1-18, 2021, doi: 10.1007/s42979-020-00418-2.
- [7] Yang, C. , Wang, Y. , Zhou, Y. , Ruan, J. and Liu, W. (2018) False Data Injection Attacks Detection in Power System Using Machine Learning Method. Journal of Computer and Communications, 6, 276-286. Ddoi: 10.4236/jcc.2018.611025.
- [8] eHealth Monitoring Open Data Project, 2018. Disponível em: <<https://sourceforge.net/projects/ehealthmonitoringproject/>>. Acessado em: 22 de maio de 2022.
- [9] Mshali, H., Lemlouma, T., Magoni, D. *Analysis of dependency evaluation models for ehealth services. In: 2014 IEEE Global Communications Conference. IEEE, 2014. p. 2429-2435, doi: 10.1109/GLOCOM.2014.7037172.*