UTILIZAÇÃO DE MINERAÇÃO DE TEXTOS NA PREVISÃO DE INTERNAÇÕES HOSPITALARES COM BASE EM REGISTROS MÉDICOS DA EMERGÊNCIA HOSPITALAR

Filipe R. Lucini
Universidade Federal do Rio Grande do Sul
Avenida Oswaldo Aranha, 99, Porto Alegre, RS
filipe.lucini@gmail.com

Flavio S. Fogliatto
Universidade Federal do Rio Grande do Sul
Avenida Oswaldo Aranha, 99, Porto Alegre, RS
ffogliatto@producao.ufrgs.br

Giovani J.C. Da Silveira
Haskayne School of Business, University of Calgary
Scurfield Hall, 2500 University Dr NW, T2N 1N4, Calgary, AB, Canada
giovani.dasilveira@haskayne.ucalgary.ca

Jeruza Neyeloff
Universidade Federal do Rio Grande do Sul
Rua Ramiro Barcelos, 2350, Porto Alegre, RS
jneyeloff@hcpa.edu.br

Michel J. Anzanello
Universidade Federal do Rio Grande do Sul
Avenida Oswaldo Aranha, 99, Porto Alegre, RS
michel.anzanello@gmail.com

Rafael Calegari
Universidade Federal do Rio Grande do Sul
Avenida Oswaldo Aranha, 99, Porto Alegre, RS
rafael.calegari@ufrgs.br

Ricardo de S. Kuchenbecker
Universidade Federal do Rio Grande do Sul
Rua Ramiro Barcelos, 2350, Porto Alegre, RS
rkuchenbecker@hcpa.edu.br

Beatriz D. Schaan
Universidade Federal do Rio Grande do Sul
Rua Ramiro Barcelos, 2350, Porto Alegre, RS
bschaan@hcpa.edu.br

https://proceedings.science/p/85299?lang=pt-br

Resumo

Nesse estudo, são utilizados métodos de mineração de texto para processar dados de pacientes de uma emergência hospitalar (EH) para prever internações e altas hospitalares. Sacos-de-palavras são obtidos via representação binária, frequência de termos e frequência de termos-inverso da frequência de documentos. Unigramas, bigramas e trigramas são testados para formação de variáveis. A seleção de variáveis é feita com base em métricas de χ^2 e F-score. No módulo de previsão, oito métodos são testados: Árvores de Decisão, Random Forest, Extremely Randomized Tree, AdaBoost, Regressão Logística, Multinomial Naïve Bayes, Máquina de Suporte Vetorial (Kernel linear) e Nu-Máquina de Suporte Vetorial (Kernel linear). A performance de previsão é avaliada via F_1-scores. Precisão e sensibilidade também são calculados para todos os métodos avaliados. Nu-Máquina de Suporte Vetorial foi o método com o melhor desempenho médio. Seu F_1-score médio na previsão de hospitalizações foi de 77.70%, com um desvio-padrão de (DP) de 0.66%.

PALAVRAS CHAVE: mineração de textos; emergências hospitalares; suporte à decisão.

Abstract

In this study, we use text mining methods to process data from early emergency department ED patient records using the SOAP framework, and predict future hospitalizations and discharges. Sets-of-words are obtained via binary representation, term frequency, and term frequency-inverse document frequency. Unigrams, bigrams and trigrams are tested for feature formation. Feature selection is based on and F-score metrics. In the prediction module, eight text mining methods are tested: Decision Tree, Random Forest, Extremely Randomized Tree, AdaBoost, Logistic Regression, Multinomial Naïve Bayes, Support Vector Machine (Kernel linear) and Nu-Support Vector Machine (Kernel linear). Prediction performance is evaluated by F1-scores. Precision and Recall values are also informed for all text mining methods tested. Nu-Support Vector Machine was the text mining method with the best overall performance. Its average F1-score in predicting hospitalization was 77.70%, with a standard deviation (SD) of 0.66%.

KEYWORDS: text mining; emergency department; decision support.

https://proceedings.science/p/85299?lang=pt-br

1. Introdução

A lotação de emergências hospitalares (EH) é um problema sério para hospitais, o qual representa riscos para a segurança dos pacientes [1]. Sabe-se que esse fenômeno ocorre principalmente quando a demanda por leitos excede a capacidade da EH [2]. Nos últimos anos, o número de leitos hospitalares vem reduzindo [3–6], apesar do aumento da demanda e da expectativa por melhores níveis de atendimento [7,8]. No Brasil, os hospitais têm um papel importante, visto que promovem um dos principais acessos ao sistema de saúde. Além do atendimento especializado, os hospitais oferecem atendimento emergencial, seja para pessoas que acessaram o atendimento por conta própria, seja para pessoas enviadas por outras unidades de atendimento de saúde [9,10].

O uso de modelos de previsão de demanda é a abordagem clássica para realizar a gestão de capacidade e alocação de recursos em uma EH [4,11,12]. A demanda futura é estimada através da análise de preditores relevantes obtidos através dos registros dos pacientes na EH [13,14]. Entretanto, os modelos tradicionais de previsão de demanda dependem de dados quantitativos, normalmente obtidos por resultados de laboratório e exames ambulatoriais, os quais raramente estão disponíveis no início do atendimento. Por outro lado, predições com o uso de mineração de textos (MT) dependem somente de dados textuais. Nesse sentido, podem gerar informações úteis a partir das anotações e registros realizados nas primeiras etapas do atendimento. Esse estudo investiga a hipótese de que técnicas de MT podem ser utilizadas para analisar registros médicos em formato texto-livre de pacientes de uma EH, de modo a prever a necessidade de internação dos pacientes, mesmo antes dos resultados de exames estejam disponíveis [15].

2. Métodos

Mineração de textos (MT) é um método que estrutura textos originalmente em formato livre, de modo que as informações possam ser analisadas quantitativamente [16]. Ele utiliza técnicas de processamento de linguagem natural, gestão do conhecimento, mineração de dados e aprendizado de máquina para processar textos. O método pode ser aplicado para classificação de documentos, extração de informações, extração de terminologias e reconhecimento de entidades (ex.: identificação de entidades de categorias pré-definidas), além de outras aplicações [17,18]. MT tem sido usado em estudos médicos com distintos objetivos, incluindo classificação automática de doenças [19], identificação de níveis de obesidade [20,21], análise de documentos biomédicos e clínicos para identificação de associações entre drogas e doenças [22], além de outros [23–31].

Considerando o contexto de uma EH, MT tem sido utilizado para estruturar informações sobre sinais vitais a partir de texto-livre, de modo a possibilitar análises com uso de métodos convencionais [32]. Por exemplo, St-Maurice *et al.* [33] converteu dados textuais para o formato numérico, de modo a prever o uso inapropriado da EH através de regressão logística. Mowery *et al*. [15] avaliou registros médicos através da estrutura SOIC (Subjetivo, Objetivo, Impressão e Conduta). O estudo mostrou que o uso de classificadores baseados em máquinas de suporte vetorial para mapear sentenças na estrutura SOIC resultou em alta concordância entre observadores e excelente cobertura de sentenças nos relatórios.

Nesse estudo, são aplicados métodos de MT em registros de consultas médicas de uma EH, de modo a desenvolver um modelo de previsão de demanda de leitos hospitalares. Nesse sentido, é utilizado um grande banco de dados real, o que possibilita a análise e comparação de diferentes configurações de estratégias de pré-processamento e métodos de classificação.

Esse é um estudo retrospectivo que utiliza métodos de MT para prever demanda de leitos hospitalares baseado somente em registros textuais gerados no primeiro encontro médico-paciente em uma EH, antes de exames médicos estarem disponíveis. O estudo foi aplicado na EH do Hospital de Clínicas de Porto Alegre (HCPA). O HCPA é um hospital universitário de atenção terciária em Porto Alegre, Brasil e possui 842 leitos. Os registros eletrônicos disponibilizados contêm informações relevantes sobre o atendimento clínico que foi oferecido aos pacientes, incluindo registros médicos, prescrições e exames clínicos e laboratoriais. O Comitê Ético do

https://proceedings.science/p/85299?lang=pt-br

HCPA aprovou esse estudo e os autores cumpriram com as recomendações da Declaração de Helsinki.

Para realizar a previsão de demanda de leitos, foi utilizado o registro textual escrito durante a primeira avaliação médica na EH. A Figura 1 apresenta uma visão geral do método, o qual está dividido em três módulos: (i) pré-processamento, (ii) predição e (iii) avaliação de performance.

As analises foram realizadas em Python 3. 4 [34]. Essa linguagem de programação apresenta bibliotecas que suportam o processamento dos dados, além da aplicação e validação do método de previsão. As bibliotecas utilizadas foram NLTK [35] e scikit-learn [36].

2.1 Módulo de Pré-Processamento

Nesse módulo, o banco de dados é preparado de modo a permitir o uso dos métodos de classificação do módulo de predição. O módulo inclui seis etapas, incluindo recuperação de texto informativo, normalização, tokenização, seleção de variáveis, conversão para saco-de-palavras e divisão do banco de dados em porções de treino e teste.

Os dados dos registros dos pacientes foram disponibilizados pelo HCPA em formato *comma-separated values* (.csv), estando escritos em português do Brasil. O banco de dados foi estruturado com dois campos de informação. O primeiro campo é composto por uma variável binária que informa o status final do paciente (internou ou recebeu alta); essa informação foi utilizada para classificar os registros do estudo. Já o segundo campo apresenta textos em formato livre escritos pelos médicos durante o atendimento aos pacientes da EH, sem nenhum intermediário (ex. escreventes).

Somente o primeiro registro médico de cada atendimento foi incluído na análise, sendo desconsiderado o tempo entre a chegada do paciente e o registro médico. Somente os dados do segundo campo foram tratados no módulo de pré-processamento. Dados numéricos (ex.: resultados de exames) não estavam presentes no banco de dados, estando a classificação baseada somente nos registros textuais médicos.

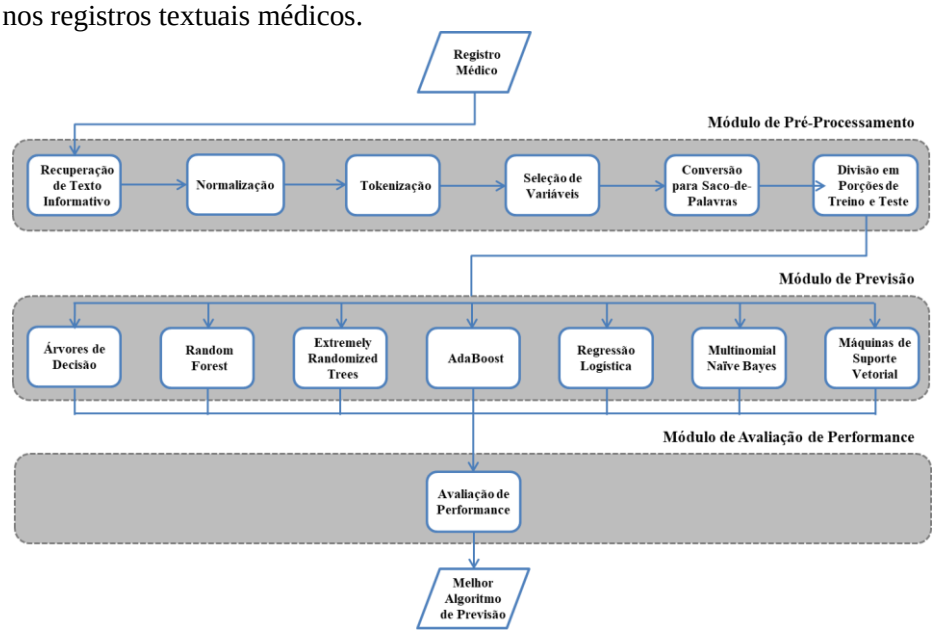


Figura 1: Visão geral do método proposto

Análises preliminares indicaram que todos os registros apresentavam conteúdo similar, o que incluía o número de registro da visita, tipo de urgência (ex.: clínica adulta, cirurgia adulta, clínica pediátrica ou cirurgia pediátrica), equipe de atendimento e anotações seguindo a estrutura SOIC. Além disso, todos os registros encerravam com o trecho “Preparado e assinado por (...)”. Boa parte do texto era gerado automaticamente pelo sistema da EH, não contendo informações discriminantes. Sendo assim, somente as anotações no formato SOIC foram consideradas

https://proceedings.science/p/85299?lang=pt-br

informativas e, portanto, foram utilizadas nas análises seguintes. As demais informações foram desconsideradas.

Em seguida, as anotações SOIC foram analisadas. Para aumentar a confiabilidade dos testes, foram excluídos os casos onde o status final do paciente estava explicitamente descrito no texto (ex.: textos informando a decisão médica de internar ou dar alta aos pacientes, pacientes que desistiram do atendimento, etc).

Uma vez que os registros foram reduzidos para as partes que apresentavam informações consideradas relevantes, foi realizada a normalização. Pontuação e palavras comuns (ex.: "e", "para", "no", etc) foram removidas, letras maiúsculas foram substituídas por minúsculas e as palavras reduzidas para os seus respectivos radicais, excluindo informação sobre gênero e número (normalização morfológica) [16].

A seguir, foi realizada a tokenização. Essa etapa consistiu em quebrar a sequência continua de caracteres de um texto em unidades linguísticas (*tokens*), como por exemplo palavras ou frases [16]. Tendo em vista que todos os registros foram escritos em português e que números e pontuações foram removidos previamente, os *tokens* foram delimitados pelos espaços em branco no texto contínuo. Depois dessa etapa, cada registro consistia em uma sequência de *tokens*.

Uma vez que os *tokens* foram identificados, as variáveis puderam ser definidas. Foram definidos três tipos de variáveis. A primeira consistia exclusivamente em unigramas (ex.: *tokens* sozinhos); a segunda incluía bigramas e a terceira incluía trigramas. Unigramas, bigramas e trigramas podem ser definidos como a sequência de uma, dois ou três *tokens* adjacentes de uma lista de *tokens* [16].

No passo de seleção de variáveis, as variáveis foram ranqueadas de acordo com um índice e aquelas acima de um determinado percentil (de maior valores) foram selecionadas. Os índices utilizados se basearam em χ^2 and F-values. Essas estatísticas foram escolhidas devido a sua boa performance em estudos anteriores [15,37–40]. Vinte e um percentis foram testados, assumindo valores entre 1 e 100.

O teste χ^2 é utilizado em estatística para verificar a independência entre um par de eventos, dentre outros usos [41]. Na seleção de variáveis, χ^2 foi utilizado para testar se a ocorrência das variáveis nas classes era independente. Ele foi baseado na seguinte equação:

$$\chi^2(t,C) = \sum_{t\in\{0,1\}}\sum_{C\in\{0,1\}}\frac{(N_{t,C}-E_{t,C})^2}{E_{t,C}} \quad (1)$$

onde *N* é a frequência observada de ocorrência, e *E* é a frequência de ocorrência esperada da variável *t* na classe *C*. Para a hipótese de independência (H_0) ser mantida, $N_{t,C}$ e $E_{t,C}$ devem convergir para 0,5, e a diferença para zero. Por outro lado, grandes valores de χ^2 indicam que a hipótese H_0 deve ser rejeitada. Nesse caso, a variável *t* deve ser selecionada [41].

Para determinado grupo de variáveis, o F-value de uma variável específica pode ser calculada por [37,42]:

$$F_value(i) = \frac{(X_i^{imp}-X'_i)^2+(X_i^{dis}-X_i)^2}{\frac{1}{n_{imp}-1}\sum_{k=1}^{n_{imp}}(X_{k,i}^{imp}-X_i^{imp})^2+\frac{1}{n_{dis}-1}\sum_{k=1}^{n_{dis}}(X_{k,i}^{dis}-X_i^{dis})^2} \quad (2)$$

onde X_i, X_i^{imp} e X_i^{dis} é a média da *i*-ésima variável nos conjuntos completo, da classe internado e da classe de alta, respectivamente; $X_{k,i}^{imp}$ é a *i*-ésima variável da *k*-ésima ocorrência de internação, e $X_{k,i}^{dis}$ é a *i*-ésima variável da *k*-ésima ocorrência de alta. O numerador indica a descriminação entre os conjuntos de internação e de alta, e o denominador indica aquele dentro de cada um dos dois conjuntos. Um valor limite é utilizado para selecionar as variáveis apropriadas de um conjunto. Se o F-value de uma determinada variável é maior que o valor limite, a variável é selecionada. Caso contrário, ela é excluída. O valor limite é obtido calculando a média de todos

https://proceedings.science/p/85299?lang=pt-br

os F-value de todas as variáveis. A equação (1) é usada para calcular o F-value da *i*-ésima variável, onde n_{imp} é o número de pacientes internados, e n_{dis} é o número de pacientes que receberam alta. X_k é um vetor, onde k = 1,…, n. Altos valores de F-values indicam variáveis discriminantes [37].

Uma vez que as variáveis foram selecionadas, é feita a conversão para saco-de-palavras. Nesse sentido, é criada uma matriz indicando a ocorrência das variáveis nos registros. As colunas da matriz correspondem às variáveis, vindo da lista das variáveis selecionadas. Já as linhas da matriz correspondem a cada um dos registros médicos. Nesse estudo, as células da matriz foram preenchidas utilizando três indicadores diferentes: representação binária (RP), frequência de termos (FT) e frequência de termos-inverso da frequência de documentos (FT-IFD). Na RB, sempre que uma variável ocorre em um registro, é atribuído o valor 1 para a célula onde a coluna e a linha se encontram. Caso contrário, é atribuído o valor 0. FT atribui a frequência com que a variável ocorre no registro. Já FT-IFD reflete a importância da variável em um registro, considerando um conjunto de registros. O valor FT-IFD aumenta proporcionalmente com a frequência da variável em um registro, porém é compensado pela frequência da variável no conjunto de registros. Matematicamente, é expressa como [16]:

$$FT_IFD(t,d,D) = \frac{f(t,d)}{\max\{f(t,d):t\in d\}} \times \log\frac{|D|}{|\{d\in D:t\in d\}|} \quad (3)$$

onde *t* indica a variável, *d* indica o registro, *D* é o número total de registros em um conjunto. $f(t,d)$ é o número de ocorrências da variável *t* no registro *d* [16].

Na última etapa do módulo de pré-processamento, o saco-de-palavras é dividido em porções de treino e teste, de modo a permitir uma validação cruzada *k-fold*. O método consiste em dividir o banco de dados em *k* subconjuntos mutuamente exclusivos de tamanho igual, de modo que um subconjunto é utilizado para teste e os demais são utilizados para treino dos parâmetros. Esse processo se repte *k* vezes, sempre alternando o subconjunto que é utilizado para teste. Ao final, estatísticas de performance são calculadas com os resultados [43]. Nesse estudo, utilizou-se k igual a 10.

2.2 Módulo de Previsão

Métodos diferentes de classificação foram testados nesse módulo. Nesse estudo, a necessidade de leitos está relacionada com a classe internou. Por exemplo, se um registro foi classificado como internou, um novo leito foi necessário; por outro lado, se um registro foi classificado como alta, não foi necessária a internação.

Os métodos de classificação testados nesse estudo foram selecionados com base em performances reportadas em estudos anteriores de classificação de textos, os quais envolviam um grande número de variáveis. Dentre os métodos, cita-se árvores de decisão [37,44,45], Random Forest [46,47], Extremely Randomized Tree [48,49], AdaBoost [50–53], regressão logística [37,54], Multinomial Naïve Bayes [37,45,54] and máquinas de suporte vetorial. [20,25,29,45,55].

2.3 Avaliação de Performance

Baseado nos resultados do módulo de previsão, estatísticas de performance foram calculadas para as diferentes combinações de estratégias de pré-processamento e métodos de previsão. Valores médios, desvios-padrão e intervalos de confiança foram calculados para três indicadores (Precisão, Sensibilidade e F_1-score), os quais são frequentemente utilizados em estudos similares [20,21,25,39,54,56,57].

Precisão fornece o percentual de registros classificados corretamente (ex.: razão entre os verdadeiros positivos e a soma dos verdadeiros positivos e falso positivos). Sensibilidade fornece o percentual de registros corretos na classificação (ex.: razão entre os verdadeiros positivos e a soma dos verdadeiros positivos e falso negativos). Nesse estudo, foi estimado o F_1-score, o qual representa a média harmônica entre as estimativas de precisão e sensibilidade [43]. O melhor método de previsão que combina os módulos de pré-processamento e de previsão é aquele que

https://proceedings.science/p/85299?lang=pt-br

maximiza o valor de F_1-score. Se diferentes combinações de pré-processamento e de previsão levarem a F_1-score que não sejam estatisticamente diferentes, a melhor combinação será aquela que tiver o maior valor médio de F_1-score.

3. Resultados

Os dados analisados incluíram registros médicos de 16.703 pacientes que visitaram a EH entre janeiro e dezembro de 2014. Após excluir os registros que o status final do paciente estava descrito explicitamente no texto, o número de registros foi reduzido para 11.705, dos quais 5.584 foram classificados como alta e 5.592 como internou.

A Tabela 1 apresenta a performance de cada um dos classificadores utilizados. Comparando os indicadores de performance com respeito à classe internou, o método Nu-SVC foi aquele que apresentou o melhor desempenho de previsão, utilizando as seguintes opções de pré-processamento: seleção de variáveis utilizando o F-value como índice, percentil de 65 e saco-de-palavras preenchido com FT-IFD e trigramas. Essa combinação obteve um F_1-score médio de 77.70% (DP = 0.66%), Sensibilidade média de 82.48% (DP = 1.05%), e Precisão média de 73.47% (DP = 1.14%). Também foi analisado se o melhor F_1-score obtido poderia ser considerado estatisticamente igual aos obtidos pelas outras combinações (Tabela 2).

A pior performance media foi obtida pelo método de Árvores de Decisão com os seguintes pré-processamentos: seleção de variáveis utilizando o χ^2 como índice, percentil de 20 e saco-de-palavras preenchido com RB e bigramas. Essa combinação obteve um F_1-score médio de 67.86% (DP = 1.29%), Sensibilidade média de 68.00% (DP = 1.70%), e Precisão média de 67.72% (DP = 1.18%).

A Figura 2 apresenta os intervalos de confiança dos F_1-scores para cada combinação de pré-processamento e previsão. Análises visuais indicam que os intervalos associados com os seis primeiros métodos na Figura 2 estão sobrepostos. Entretanto, analizando os *p*-values correspondentes (Tabela 2), a hipótese de F_1-scores iguais não foi rejeitada somente para os métodos NuSVC, SVC Linear, e Extremely Randomized Tree. Importante salientar que a sobreposição de intervalos de confiança não significa necessariamente que as médias não são estatisticamente diferentes. Esse é o caso, visto que a distância da média é calculada de formas diferentes para a estatística *t* em relação à média do intervalo de confiança; veja [58].

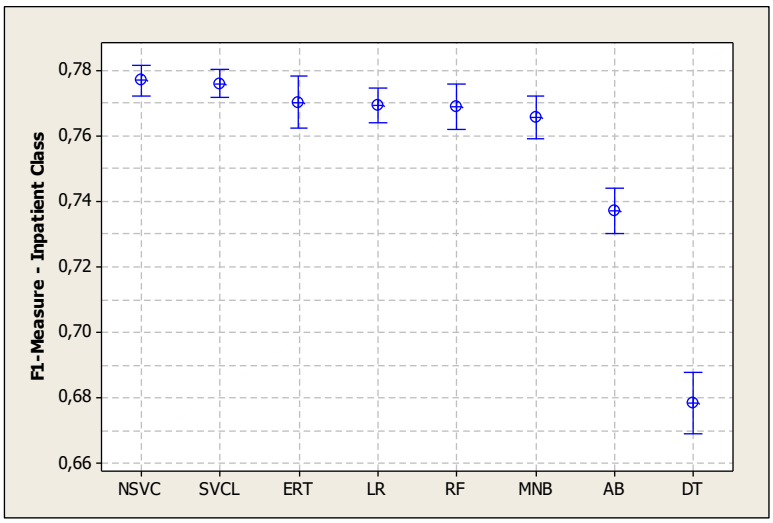


Figura 2. Comparação de *F*1 95% intervalos de confiança para métodos diferentes de previsão

NSVC: NuSVC, SVCL: SVC Liner, ERT: Extremely Randomized Tree, LR: Regressão Logística, RF: Random Forest, MNB: Multinomial Naïve Bayes, AB: AdaBoost, DT: Árvores de Decisão.

O método NuSVC com as opções de pré-processamento previamente descritas foi o melhor (baseado no valor F_1-score) para prever hospitalizações utilizando registros da EH.

https://proceedings.science/p/85299?lang=pt-br

Entretanto, diversas combinações alternativas de pré-processamento e previsão obtiveram resultados de F_1-scores que não são estatisticamente diferentes da nossa escolha.

4. Discussão

No nosso conhecimento, esse é o primeiro estudo que usa MT para estimar a demanda de leitos baseado em registros médicos gerados no primeiro encontro entre médicos e pacientes em uma EH, antes que informações de resultados de exames estivessem disponíveis. Nu-SVC obteve um F_1-score médio de 77.70% e, portanto, representa um bom estimador. Complementarmente, esse é um dos poucos estudos que compara diferente estratégias de pré-processamento e de previsão utilizando registros textuais de uma EH [15,33,59].

Mowery *et al*. [15] utilizaram um classificador automático de SOIC, o qual permitiu mapear registros médicos em uma das classes da estrutura SOIC. Para tanto, treinou um classificador com registros que haviam sido previamente preparados e classificados por humanos. Em uma validação cruzada, eles obtiveram um F_1-score médio de 85.30% nas suas classificações. Considerando o estudo de Mowery *et al*. [15], nossos resultados mostram potenciais benefícios em adotar uma abordagem de MT em EH cronicamente lotadas. Apesar da adoção da estrutura SOIC não ser mandatória na EH estudada, a abordagem estava presente um grande quantidade de registros médicos e, portanto, pode ter contribuído com o desempenho do modelo de MT.

Adicionalmente à contribuição metodológica do presente estudo, foi apresentada uma ferramenta poderosa para utilizar mais efetivamente a informação abundante registrada em texto, que está disponível nos registros médicos. Essa abordagem pode levar a melhor gestão das operações no sistema de saúde.

Esse estudo apresenta limitações que podem ser analisadas no futuro. Esse estudo foi realizado em somente um hospital e, portanto, os resultados não podem ser generalizados para outras instituições. Além disso, poucos estudos avaliaram a estrutura SOIC em EH, portanto pode ser difícil de comparar os resultados com outros estudos. Por outro lado, o estudo analisou uma amostra grande e, portanto, minimizou os riscos associados a análises em bancos de dados pequenos.

Estudo futuros poderão também avaliar as contribuições da estrutura SOIC em uma progressão temporal, especialmente no contexto de estimativa de recursos para pacientes com doenças crônicas que se apresentam em EH lotadas com pacientes com demandas de baixa acuidade.

5. Referências

[1] P.C. Sprivulis, J.-A. Da Silva, I.G. Jacobs, A.R.L. Frazer, G.A. Jelinek, The association between hospital overcrowding and mortality among patients admitted via western australian emergency departments, Medical Journal of Australia. 184 (2006) 208–212.

[2] N.R. Hoot, D. Aronsky, Systematic review of emergency department crowding: causes, effects, and solutions., Annals of Emergency Medicine. 52 (2008) 126–36. doi:10.1016/j.annemergmed.2008.03.014.

[3] A. Alijani, G.B. Hanna, D. Ziyaie, S.L. Burns, K.L. Campbell, M.E.T. Mcmurdo, A. Cuschieri, S. Mason, M. Hussain-gambles, B. Leese, K. Atkin, J. Brown, Instrument for objective assessment of appropriateness of surgical bed occupancy : validation study Representation of South Asian people in randomised clinical trials : analysis of trials ' data, BMJ. 326 (2003) 1243–1244.

[4] M. Mackay, Choice of Models for the Analysis and Forecasting of Hospital Beds, Health Care Management Science. 8 (2005) 221–230.

[5] OECD, Hospital Beds, in: OECD Publishing (Ed.), Health at a Glance 2011: OECD Indicators, 2011. http://dx.doi.org/10.1787/health_glance-2011-en.

[6] Brasil. Tribunal de Contas da União, Relatório Sistêmico de Fiscalização da Saúde - FISC Saúde, (2013).

[7] R. Lucas, H. Farley, J. Twanmoh, A. Urumov, N. Olsen, B. Evans, H. Kabiri, Emergency department patient flow: the influence of hospital census variables on emergency

https://proceedings.science/p/85299?lang=pt-br

department length of stay., Academic Emergency Medicine. 16 (2009) 597–602. doi:10.1111/j.1553-2712.2009.00397.x.
[8] M.I. Schmidt, B.B. Duncan, G. Azevedo e Silva, A.M. Menezes, C.A. Monteiro, S.M. Barreto, D. Chor, P.R. Menezes, Chronic non-communicable diseases in Brazil: burden and current challenges, Lancet (London, England). 377 (2011) 1949–61. doi:10.1016/S0140-6736(11)60135-9.
[9] J.L. Wiler, D. a Handel, A. a Ginde, D. Aronsky, N.G. Genes, J.L. Hackman, J. a Hilton, U. Hwang, M. Kamali, J.M. Pines, E. Powell, M. Sattarian, R. Fu, Predictors of patient length of stay in 9 emergency departments., The American Journal of Emergency Medicine. 30 (2012) 1860–4. doi:10.1016/j.ajem.2012.03.028.
[10] National Association of Public Hospitals and Health Systems, Perfecting Patient Flow, (2005).
[11] R.D.T. Farmer, J. Emami, Models for forecasting hospital bed requirements in the acute sector, Journal of Epidemiology and Community Health. 44 (1990) 307–312.
[12] T.K. Sweeney, S.A. Ashley, Forecasting hospital bed needs, British Medical Journal. 283 (1981) 331–334.
[13] C. García-Gubern, L. Colón-Rolón, M.M. Aponte, Workforce projections for emergency medicine in Puerto Rico: a five-year follow-up of an evident demand, The Journal of Emergency Medicine. 29 (2005) 107–10. doi:10.1016/j.jemermed.2005.01.013.
[14] S.R. Bird, W. Kurowski, G.K. Dickman, I. Kronborg, Integrated care facilitation for older patients with complex health care needs reduces hospital demand, Australian Health Review. 31 (2007) 451–461.
[15] D. Mowery, J. Wiebe, S. Visweswaran, H. Harkema, W.W. Chapman, Building an automated SOIC classifier for emergency department reports., Journal of Biomedical Informatics. 45 (2012) 71–81. doi:10.1016/j.jbi.2011.08.020.
[16] R. Feldmann, J. Sanger, The text mining handbook: Advanced approaches in analyzing unstructured data, Cambridge Press, New York, 2006.
[17] I. Spasić, J. Livsey, J. a Keane, G. Nenadić, Text mining of cancer-related information: review of current status and future directions, International Journal of Medical Informatics. 83 (2014) 605–623. doi:10.1016/j.ijmedinf.2014.06.009.
[18] G. Karystianis, I. Buchan, G. Nenadic, Mining characteristics of epidemiological studies from Medline: a case study in obesity., Journal of Biomedical Semantics. 5 (2014) 1–11. doi:10.1186/2041-1480-5-22.
[19] I. Solt, D. Tikk, V. Gál, Z.T. Kardkovács, Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier, Journal of the American Medical Informatics Association : JAMIA. 16 (2009) 580–584. doi:10.1197/jamia.M3087.
[20] H. Yang, I. Spasic, J. a Keane, G. Nenadic, A text mining approach to the prediction of disease status from clinical discharge summaries, Journal of the American Medical Informatics Association : JAMIA. 16 (2009) 596–600. doi:10.1197/jamia.M3096.
[21] K.H. Ambert, A.M. Cohen, A system for classifying disease comorbidity status from medical discharge summaries using automated hotspot and negated concept detection, Journal of the American Medical Informatics Association : JAMIA. 16 (2009) 590–595. doi:10.1197/jamia.M3095.
[22] C.F. Elizabeth S. Chen, George Hripcsak, Hua Xu, Marianthi Markatou, Automated Acquisition of Disease – Drug Knowledge from Biomedical and Clinical Documents : An Initial Study, Journal of the American Medical Informatics Association : JAMIA. 15 (2008) 87–98. doi:10.1197/jamia.M2401.Introduction.
[23] X. Wang, G. Hripcsak, M. Markatou, C. Friedman, Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study, Journal of the American Medical Informatics Association : JAMIA. 16 (2009) 328–337. doi:10.1197/jamia.M3028.

https://proceedings.science/p/85299?lang=pt-br

[24] R.L. Deendayal Dinakarpadian, Yugyng Lee, Kartik Vishwanath, MachineProse : An Ontological Framework for Scientific Assertions, Journal of the American Medical Informatics Association : JAMIA. 13 (2006) 220–232. doi:10.1197/jamia.M1910.This.
[25] X. Lu, B. Zheng, A. Velivelli, C. Zhai, Enhancing text categorization with semantic-enriched representation and training data augmentation., Journal of the American Medical Informatics Association : JAMIA. 13 (2006) 526–35. doi:10.1197/jamia.M2051.
[26] J.D. Michelson, J.S. Pariseau, W.C. Paganelli, Assessing surgical site infection risk factors using electronic medical records and text mining., American Journal of Infection Control. 42 (2014) 333–336. doi:10.1016/j.ajic.2013.09.007.
[27] K. Vallmuur, Machine learning approaches to analysing textual injury surveillance data: a systematic review., Accident Analysis and Prevention. 79 (2015) 41–49. doi:10.1016/j.aap.2015.03.018.
[28] L. Pereira, R. Rijo, C. Silva, M. Agostinho, ICD9-based Text Mining Approach to Children Epilepsy Classification, Procedia Technology. 9 (2013) 1351–1360. doi:10.1016/j.protcy.2013.12.152.
[29] M. Yang, M. Kiang, W. Shang, Filtering big data from social media--Building an early warning system for adverse drug reactions., Journal of Biomedical Informatics. 54 (2015) 230–240. doi:10.1016/j.jbi.2015.01.011.
[30] S. Ayvaz, J. Horn, O. Hassanzadeh, Q. Zhu, J. Stan, N.P. Tatonetti, S. Vilar, M. Brochhausen, M. Samwald, M. Rastegar-Mojarad, M. Dumontier, R.D. Boyce, Toward a complete dataset of drug-drug interaction information from publicly available sources., Journal of Biomedical iIformatics. 55 (2015) 206–217. doi:10.1016/j.jbi.2015.04.006.
[31] Y.L. Yongjing Lin, Wenyuan Li, Keke Chen, A Document Clustering and Ranking System for Exploring MEDLINE Citations, Journal of the American Medical Informatics Association : JAMIA. 14 (2007) 651–661. doi:10.1197/jamia.M2215.Introduction.
[32] N. Genes, D. Chandra, S. Ellis, K. Baumlin, I. Technology, M. Sinai, N. York, Validating Emergency Department Vital Signs Using a Data Quality Engine for Data Warehouse, The Open Medical Informatics Journal. 7 (2013) 34–39.
[33] J. St-maurice, M.H. Kuo, Analyzing Primary Care Data to Characterize Inappropriate Emergency Room Use, Quality of Life through Quality of Information. (2012) 990–994. doi:10.3233/978-1-61499-101-4-990.
[34] Python Software Foundation, The Python Language Reference, (n.d.). https://www.python.org/.
[35] S. Bird, E. Klein, E. Loper, Natural Language Processing with Python, 1º edition, O'Reilly Media, Sebastopol, CA, 2009.
[36] G. Hackeling, Mastering Machine Learning With scikit-learn, Packt Publishing Ltd., Birmingham, UK, 2004.
[37] S. Bashir, U. Qamar, F.H. Khan, L. Naseem, HMV: A medical decision support framework using multi-layer classifiers for disease prediction, Journal of Computational Science. 13 (2016) 10–25. doi:10.1016/j.jocs.2016.01.001.
[38] N.R. B Sarojini Ilango, A Hybrid Prediction Model with F-score Feature Selection for Type II Diabetes Databases, in: A2CWiC Amrita ACM-W Celebration on Women in Computing in India, 2010.
[39] A.M.C. Ohen, W.R.H. Ersh, K.P. Eterson, P.O.I.N.Y. En, Reducing Workload in Systematic Review Preparation Using Automated Citation Classification, Journal of the American Medical Informatics Association : JAMIA. 13 (2006) 206–219. doi:10.1197/jamia.M1929.The.
[40] P. Meesad, P. Boonrawd, V. Nuipian, A Chi-Square-Test for Word Importance Differentiation in Text Classification, in: 2011 International Conference on Information and Electronics Engineering, IACSIT Press, Singapore, 2011: pp. 110–114.
[41] A.K. Uysal, S. Gunal, A novel probabilistic feature selection method for text classification, Knowledge-Based Systems. 36 (2012) 226–235. doi:10.1016/j.knosys.2012.06.005.

https://proceedings.science/p/85299?lang=pt-br

[42] Y. Chen, C. Lin, Combining SVMs with Various Feature Selection Strategies, in: Studies in Fuzziness and Soft Computing, Springer-Verlag, Heildelberg, 2006: pp. 315–324.
[43] T. Hastie, R. Tibshirani, Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2º Edition, Springer, 2008.
[44] D. Delen, A. Oztekin, L. Tomak, An analytic approach to better understanding and management of coronary surgeries, Decision Support Systems. 52 (2012) 698–705. doi:10.1016/j.dss.2011.11.004.
[45] N. Esfandiari, M.R. Babavalian, A.-M.E. Moghadam, V.K. Tabar, Knowledge discovery in medicine: Current issue and future trend, Expert Systems with Applications. 41 (2014) 4434–4463. doi:10.1016/j.eswa.2014.01.011.
[46] L.E.O. Breiman, Random Forests, Machine Learning. 45 (2001) 5–32.
[47] X. Pu, K. Fan, X. Chen, L. Ji, Z. Zhou, Facial expression recognition from image sequences using twofold random forest classifier, Neurocomputing. 168 (2015) 1173–1180. doi:10.1016/j.neucom.2015.05.005.
[48] F. Scalzo, R. Hamilton, S. Asgari, S. Kim, X. Hu, Intracranial hypertension prediction using extremely randomized decision trees., Medical Engineering & Physics. 34 (2012) 1058–65. doi:10.1016/j.medengphy.2011.11.010.
[49] P. Geurts, D. Ernst, L. Wehenkel, Extremely Randomized Trees, Machine Learning. 63 (2006) 3–42. doi:10.1007/s10994-006-6226-1.
[50] D.R. Nayak, R. Dash, B. Majhi, Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests, Neurocomputing. 177 (2016) 188–197. doi:10.1016/j.neucom.2015.11.034.
[51] M.V. Reddy, R. Sodhi, A rule-based S-Transform and AdaBoost based approach for power quality assessment, Electric Power Systems Research. 134 (2016) 66–79. doi:10.1016/j.epsr.2016.01.003.
[52] T. Gaber, A. Tharwat, A.E. Hassanien, V. Snasel, Biometric cattle identification approach based on Weber's Local Descriptor and AdaBoost classifier, Computers and Electronics in Agriculture. 122 (2016) 55–66. doi:10.1016/j.compag.2015.12.022.
[53] Q. Nie, L. Jin, S. Fei, Probability estimation for multi-class classification using AdaBoost, Pattern Recognition. 47 (2014) 3931–3940. doi:10.1016/j.patcog.2014.06.008.
[54] M.A. Morid, M. Fiszman, K. Raja, S.R. Jonnalagadda, G. Del Fiol, Classification of Clinically Useful Sentences in Clinical Evidence Resources., Journal of Biomedical Informatics. 60 (2016) 14–22. doi:10.1016/j.jbi.2016.01.003.
[55] A.M. Cohen, K. Ambert, M. McDonagh, Cross-topic learning for work prioritization in systematic review creation and update., Journal of the American Medical Informatics Association : JAMIA. 16 (2009) 690–704. doi:10.1197/jamia.M3162.
[56] L.H. Ben Wellmer, Matt Huyck, Scott Mardis, John Aberdeen, Alex Morgan, Leonid Peshkin, Alex Yeh, Janet Hitzeman, Rapidly Retargetable Approaches to De-identification in Medical Records, Journal of the American Medical Informatics Association : JAMIA. 14 (2007) 564–573. doi:10.1197/jamia.M2435.Introduction.
[57] Y. Huang, H.J. Lowe, D. Klein, R.J. Cucina, Improved Identification of Noun Phrases in Clinical Radiology Reports Using a High-Performance Statistical Natural Language Parser Augmented with the UMLS Specialist Lexicon, Journal of the American Medical Informatics Association : JAMIA. 12 (2005) 275–285. doi:10.1197/jamia.M1695.The.
[58] A. Knezevic, Overlapping Confidence Intervals and Statistical Significance, 2008. https://www.cscu.cornell.edu/news/statnews/stnews73.pdf.
[59] D. Zmiri, Y. Shahar, M. Taieb-Maimon, Classification of patients by severity grades during triage in the emergency department using data mining methods., Journal of Evaluation in Clinical Practice. 18 (2012) 378–388. doi:10.1111/j.1365-2753.2010.01592.x.

https://proceedings.science/p/85299?lang=pt-br

Tabela 2: Melhor Performance de Cada Método de Previsão

Método	FS	P	SdP	NG	F-Measure					Sensibilidade				Precisão			
					Média	SD	LB-CI	UB-CI	PV	Média	SD	LB-CI	UB-CI	Média	SD	LB-CI	UB-CI
NuSVC	F-score	65	TF-IDF	3	77.70%	0.66%	77.23%	78.18%	1.000	82.48%	1.05%	81.73%	83.23%	73.47%	1.14%	72.65%	74.28%
SVC Linear	F-score	45	TF-IDF	2	77.59%	0.60%	77.16%	78.02%	0.693	79.70%	1.25%	78.80%	80.59%	75.62%	1.21%	74.75%	76.49%
Extra Tree	F-score	30	TF	3	77.03%	1.10%	76.24%	77.82%	0.117	82.04%	1.83%	80.73%	83.35%	72.62%	0.98%	71.91%	73.32%
Logistic Regression	χ^2	65	BR	3	76.95%	0.74%	76.42%	77.48%	0.028	80.16%	1.50%	79.09%	81.23%	74.02%	1.34%	73.06%	74.98%
Random Forest	χ^2	35	BR	3	76.90%	0.97%	76.21%	77.59%	0.046	84.82%	1.39%	83.83%	85.81%	70.35%	1.24%	69.46%	71.23%
Multinomial NB	χ^2	60	BR	2	76.55%	0.91%	75.90%	77.20%	0.005	85.30%	1.71%	84.08%	86.53%	69.44%	0.86%	68.83%	70.05%
AdaBoost	χ^2	1	BR	1	73.72%	0.97%	73.03%	74.41%	0.000	70.86%	1.47%	69.81%	71.90%	76.84%	1.22%	75.97%	77.71%
Decision Tree	χ^2	20	BR	2	67.86%	1.29%	66.93%	68.78%	0.000	68.00%	1.70%	66.79%	69.21%	67.73%	1.18%	66.88%	68.57%

FS: Seleção de Variaveis, P: Percentil, SdP: saco-de-palavras, NG: n-gramas, SD: Desvio-Padrão, LB-CI: Limite Inferior – Intervalor de Confiança, UB-CI: Limite Superior – Intervalo de Confiança, PV: p-value

https://proceedings.science/p/85299?lang=pt-br