

Análise de dados de exercícios físicos combinados com dados climáticos como sistema de apoio à prescrição de treinamento

Rafael Macedo Sulino¹, Carlos Norberto Fischer¹, Wellington Roberto Gomes de Carvalho²

¹Instituto de Biociências – Universidade Estadual Paulista “Julio de Mesquita Filho”
Avenida 24A, 1515 – CEP 13506-900 – Rio Claro – São Paulo – Brasil

²Faculdade de Educação Física – Universidade Federal de Uberlândia
Rua Benjamin Constant, 1286 – CEP 38400-678 – Uberlândia – Minas Gerais – Brasil

{rmsulino,carlos}@rc.unesp.br, wrgcarvalho@ufu.br

Abstract. *Technological evolution increasingly enables the production of digital data worldwide, especially with the advancement of the "Internet of Things", through various sensors and mobile devices. The practice of physical exercise is undoubtedly important for the prevention of chronic diseases, such as cardiovascular diseases, diabetes, among others, so that mobile devices and applications capable of recording physical activity data become of special interest to health. This paper presents a proposal for collecting and analyzing physical exercise data combined with climatic data, obtained from applications and mobile devices available for the monitoring of physical activity, in order to produce information that can support the professional in decision making for the prescription of physical exercises.*

Resumo. *A evolução tecnológica possibilita cada vez mais a produção de dados digitais em todo o mundo, especialmente com o avanço da “Internet das Coisas”, por meio de diversos sensores e dispositivos móveis. A prática de exercícios físicos é indiscutivelmente importante para a prevenção de doenças crônicas não transmissíveis, como as doenças cardiovasculares, diabetes, entre outras, de modo que se torna de especial interesse para a área da saúde os dispositivos móveis e aplicativos capazes de registrar dados de atividade física. Este artigo apresenta uma proposta de coleta e análise de dados de exercícios físicos combinados com dados climáticos, obtidos a partir de aplicativos e dispositivos móveis disponíveis para o monitoramento da atividade física, de forma a produzir informações que possam apoiar o profissional na tomada de decisões para a prescrição de exercícios físicos.*

1. Introdução

Vivemos atualmente na era da chamada “Internet das Coisas”, em que diversas formas de interação entre pessoas e dispositivos ou sensores resultam em um grande volume de dados digitais. Observamos o desenvolvimento crescente de sensores “vestíveis” (*wearables*), *smartphones* e uma série de outros dispositivos capazes de coletar dados e se comunicar com outros dispositivos. Estima-se que a quantidade desse tipo de dispositivo já supera a população existente no planeta, o que possibilita uma coleta de dados expressiva e relevante para a aplicação de técnicas de mineração de dados [Aggarwal 2015]. Com a crescente preocupação por uma vida mais saudável, os dispositivos móveis ganham notável importância ao possibilitarem a produção de dados sobre saúde, prática de atividade física e aspectos nutricionais em tempo real. Dessa forma, temos à disposição diversos aplicativos e dispositivos que oferecem dietas nutricionais e programas de treinamento físico, baseados em dados do indivíduo, como gênero, idade, estatura e massa corporal, além de seus objetivos, como o emagrecimento

[Oikawa 2013].

De modo geral, os dados de saúde e atividades físicas coletados a partir de aplicativos e dispositivos móveis não recebem um tratamento e análise adequados, por meio de uma abordagem científica. A análise adequada dos dados possibilitaria uma melhor tomada de decisão no sentido de planejar as atividades físicas de forma mais eficiente, de acordo com o objetivo desejado (emagrecimento, desempenho, controle de diabetes, entre outros). Serrano et al. (2016) apontam que mais da metade dos aplicativos instalados em *smartphones* envolvem algo relacionado a controle de massa corporal, dieta ou exercício. Os dados coletados por meio destes aplicativos e dispositivos abrangem milhões de usuários, e ao serem tratados com um método científico, como a mineração de dados, oferecem potencial para importantes descobertas de conhecimento.

A mineração de dados pode ser definida como um processo de análise e extração de informações úteis, padrões e tendências, muitas vezes desconhecidos, a partir de grande quantidade de registros armazenados em bancos de dados [Thuraisingham 1999]. Segundo Côrtes, Porcaro e Lifschitz (2002), o conceito de mineração de dados se populariza cada vez mais como uma ferramenta de descoberta de informações, capazes de revelar estruturas de conhecimento e guiar decisões em condições de certeza limitada. Trata-se de um processo cooperativo entre homem e máquina, visando explorar grandes volumes de dados com o objetivo de extrair conhecimento por meio de reconhecimento de padrões e relacionamento entre variáveis, utilizando-se técnicas comprovadamente confiáveis e validadas.

Diante do exposto, justificam-se os esforços no sentido de empregar uma abordagem científica em dados coletados por aplicativos e dispositivos de monitoramento de atividade física e de saúde, objetivando a descoberta de conhecimento a partir do uso da mineração de dados. Nesse sentido, visando estabelecer tal abordagem científica no tratamento destes dados, este estudo propõe o desenvolvimento de um sistema que possibilite a coleta e armazenamento de dados de aplicativos e dispositivos populares, além da análise desses dados por meio da mineração de dados, de forma a oferecer aos profissionais, especialmente da área de Educação Física, ferramentas de apoio na tomada de decisão para a prescrição de treinamento de exercícios físicos.

2. Material e Método

A amostra foi extraída de um conjunto de dados contendo 401 registros de atividade física, coletados a partir do serviço *Strava* (www.strava.com), dos quais foram selecionados 88 registros que continham dados referentes à frequência cardíaca, provenientes de 10 participantes adultos (7 homens e 3 mulheres), com idade entre 25 e 55 anos. O conjunto de dados recebeu o incremento de dados climáticos provenientes do serviço *Weather Underground* (www.wunderground.com), de acordo com a localização geográfica, data e hora de cada exercício.

Após receber o convite, cada participante se inscreveu por meio de acesso ao sistema disponível em “<https://rafaelsulino.pro.br/research>”, autorizando o compartilhamento de seus dados armazenados no *Strava*. Os dados foram coletados de forma automática, por meio de uma aplicação computacional desenvolvida em linguagem de programação *Python*, e armazenados em um banco de dados *MongoDB*, executado em um Servidor Virtual Privado com sistema operacional *Linux Ubuntu* 16.04.

Os dados foram previamente processados, excluindo-se registros inconsistentes. Algumas variáveis foram transformadas do tipo “*string*” para o tipo “nominal”, com atributos pré-definidos, de forma a se tornarem compatíveis com a análise de dados empregada. Para o presente estudo, foram utilizadas três das variáveis coletadas: zona

de intensidade do exercício (faixas de frequência cardíaca), ganho de elevação (em metros) e velocidade do vento (em quilômetros por hora). A intensidade do exercício foi classificada em cinco zonas de frequência cardíaca (Z1 a Z5).

A análise dos dados foi realizada por meio do método de aprendizagem supervisionada, utilizando-se o algoritmo de aprendizagem C4.5 [Quinlan 1993], o qual é capaz de produzir classificadores e gerar árvores de decisão, buscando a influência de determinadas variáveis na intensidade do exercício (zonas de frequência cardíaca), sendo o modelo treinado testado na mesma amostra (“*full training set*”). Foi utilizado o *software Weka* versão 3.8.1 [Frank, Hall e Witten 2016].

3. Resultados e Discussão

O algoritmo de aprendizagem C4.5 foi utilizado para produzir classificadores em duas situações distintas, obtendo 52% de sucesso ao verificar a influência da velocidade do vento na intensidade do exercício (zona de frequência cardíaca) e 64% de sucesso ao verificar a influência do ganho de elevação (subidas acumuladas) na intensidade do exercício. Um modelo treinado com base em uma amostra com maior número de instâncias deverá obter um percentual de sucesso mais elevado.

A Figura 1 apresenta a árvore de decisão produzida pelo algoritmo C4.5, na qual é possível observar que, para a amostra estudada, uma maior intensidade do exercício (zona 3 de frequência cardíaca) ocorreu quando a velocidade do vento foi mais elevada (maior que 27,8 km/h).

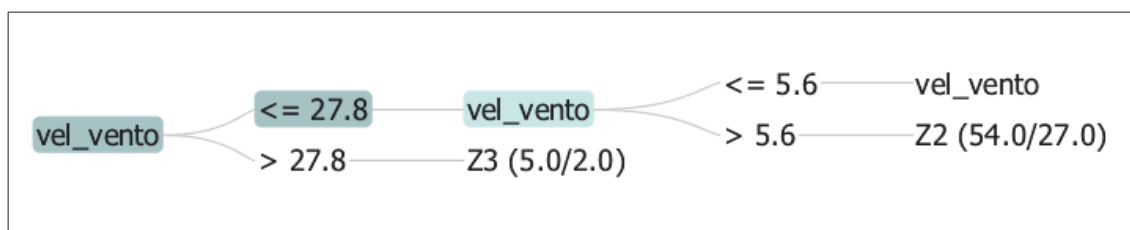


Figura 1. Árvore de decisão demonstrando a influência da velocidade do vento na intensidade do exercício.

Na Figura 2 observa-se que, para a amostra estudada, ocorreu uma maior intensidade do exercício (zona 3 de frequência cardíaca) quando houve um maior ganho de elevação (maior que 765 metros), ou seja, uma maior quantidade de subidas acumuladas durante o percurso.

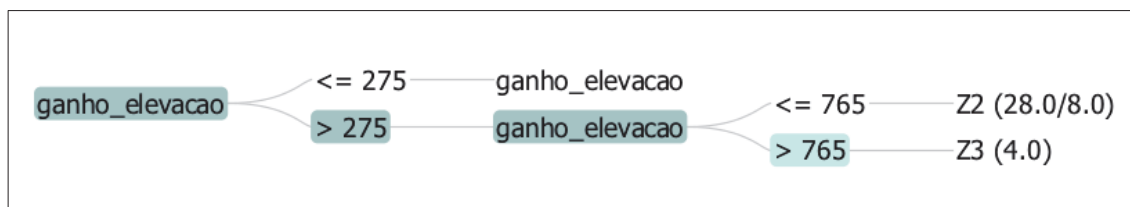


Figura 2. Árvore de decisão demonstrando a influência do ganho de elevação na intensidade do exercício.

Os resultados apresentados mostram a viabilidade da utilização da mineração de dados no sentido de extrair informações relevantes de dados coletados por meio de dispositivos móveis e aplicativos utilizados no monitoramento de atividades físicas, especialmente quando combinados com dados climáticos.

Exemplificando a aplicação prática, com base nos achados da análise preliminar,

o profissional, ao prescrever um ciclo de treinamento, poderia levar em conta a influência da velocidade do vento na intensidade do exercício e propor como alternativa uma atividade em ambiente fechado em dias de vento forte, ou levar em conta a influência do ganho de elevação na intensidade do exercício e propor como alternativa um percurso mais plano quando o objetivo for um treino mais leve ou de recuperação.

Devido à limitação da amostra utilizada no presente estudo, em razão do tamanho e número reduzido de variáveis, os padrões encontrados não apresentam novidades no que diz respeito a conhecimentos pertinentes à prescrição de exercícios físicos. No entanto, eles mostram a viabilidade de se utilizar técnicas como mineração de dados e aprendizagem de máquina como sistemas de apoio à tomada de decisões na prescrição daqueles exercícios. Quanto maior a quantidade de dados disponível e maior o número de variáveis consideradas, maiores são as possibilidades de identificação de padrões e descoberta de conhecimento relevante.

4. Considerações Finais

Os dados aqui apresentados são preliminares, como projeto-piloto para um estudo maior, o qual contará com a aquisição de dados de diversas fontes, tais como aplicativos de monitoramento de exercícios físicos e dispositivos móveis, além de uma quantidade maior de variáveis. Quanto a estas novas variáveis, serão incluídos dados de histórico de lesões, histórico de patologias (como diabetes e hipertensão arterial), uso de medicamentos (como betabloqueadores que atuam de forma direta no comportamento da frequência cardíaca), informações sobre qualidade de vida, dentre outras, que, associadas a um volume maior de dados, deverá trazer uma melhor compreensão desses dados e a aquisição de conhecimento relevante para auxiliar o profissional no planejamento de exercícios físicos.

Referências

- Aggarwal, C. C. (2015). *Data mining: the textbook*, Springer.
- Côrtes, S. C.; Porcaro, R. M.; Lifschitz, S. (2002). *Mineração de Dados: funcionalidades, técnicas e abordagens*. PUC.
- 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.
- Oikawa, E. (2013). *Dinâmicas relacionais contemporâneas: visibilidade, performances e interações nas redes sociais da internet*. Interações em rede.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Serrano, K. J. et al. (2016). *Mining Health App Data to Find More and Less Successful Weight Loss Subgroups*. *J Med Internet Res*, v. 18, n. 6, p. e154.
- Thuraisingham, B. M. (1999). *Data mining: technologies, techniques, tools, and trends*. Boca Raton: CRC Press.