

Uma abordagem algorítmica para auxiliar precocemente ao diagnóstico de jovens em risco de TDAH

Yara de Lima Araújo¹, José Raimundo Macário Costa², Sérgio Manuel Serra da Cruz³

^{1,2,3} Instituto de Ciências Exatas - Departamento de Matemática – Universidade Federal Rural do Rio de Janeiro (UFRRJ)

BR-465, Km 7 - CEP 23.897-000 - Seropédica- RJ – Brasil

{yara,serra}@pet-si.ufrrj.br^{1,3}, mac_costa@yahoo.com²

Abstract. *Given the concern about academic, social and psychologic consequences in those diagnosed with ADHD, there is an opportunity to develop a tool for helping early diagnosis of ADHD. For this reason, computational techniques such as Neural Networks and Clustering algorithms are seen as useful alternatives by researchers. This paper presents results from tests with these algorithms. Within the study, it was possible to identify one possible candidate at risk of ADHD.*

Resumo. *Diante da preocupação com as consequências acadêmicas, sociais e psicológicas dos portadores do Transtorno de Deficit de Atenção e Hiperatividade, percebe-se que há oportunidades de desenvolvimento de ferramenta computacional capaz de auxiliar no diagnóstico precoce do TDAH. Para tal, técnicas como Redes Neurais e algoritmos de clusterização são vistas como boas alternativas pela comunidade de pesquisa. Este trabalho apresenta resultados da adoção destes algoritmos, e foi possível identificar um candidato em risco de TDAH.*

1. Introdução

Desde o estabelecimento das políticas de inclusão de portadores de necessidades especiais nas escolas regulares através da LDB 9394/96, artigo 58 (Brasil, 1996), profissionais da educação buscam orientações e informações acerca de como lidar com determinados transtornos de aprendizagem e ao mesmo tempo garantir o direito ao aprendizado a todos os alunos. Além disso, devido a grande ocorrência desses transtornos entre crianças e jovens, um grande número de pesquisadores de áreas interdisciplinares vem investigando o problema, buscando desenvolver aplicações computacionais para detectar precocemente os indivíduos com dificuldades de aprendizagem.

De acordo com o Manual Diagnóstico e Estatístico da Associação Americana de Psiquiatria (DSM-IV, 2014), um transtorno específico de aprendizagem, como o nome implica, é diagnosticado diante de déficits específicos na capacidade individual para perceber ou processar informações com eficiência e precisão. Dentre esses transtornos de aprendizagem podemos destacar o Transtorno de Deficit de Atenção e Hiperatividade (TDAH) manifesto inicialmente durante os primeiros anos de escolaridade formal; caracterizando-se por dificuldades persistentes nas habilidades básicas de leitura, escrita e/ou matemática (SAMPAIO e FREITAS, 2014, p. 131).

Essas características afetam o desenvolvimento acadêmico, os relacionamentos familiares e sociais e a vida laboral. O TDAH se manifesta na primeira infância e se não for corretamente diagnosticado e tratado precocemente pode apresentar sintomas na vida adulta, interferindo na vida acadêmica, profissional, afetiva e social. Estima-se que 70% das pessoas que tiveram TDAH diagnosticado na infância mantêm o transtorno na vida adulta (BASTOS et al. 2012).

Neste trabalho, investigamos a adoção de técnicas baseadas na utilização de algoritmos de clusterização que permitam a detecção precoce e automatizada de possíveis candidatos em risco de TDAH. Nossa pesquisa se justifica em estudos recentes que apontam de 3 a 5% de crianças atingidas por este transtorno (APA, 2002; ROHDE & KETZER, 1997). Além disso, ainda há desconhecimento sobre o TDAH mesmo por parte de pais e professores, o que pode resultar na identificação tardia do transtorno.

O objetivo desse trabalho é avaliar a eficácia de algoritmos de clusterização *K-means* e uma rede neural não supervisionada para a detecção de possíveis candidatos em risco de TDAH. Nesse trabalho foram extraídos dados de entrevistas realizadas com jovens brasileiros e espanhóis da faixa etária entre 8 a 18 anos através do questionário MTA-SNAP-IV.

Esse trabalho está organizado da seguinte forma: a seção 2 apresenta o referencial teórico e trabalhos relacionados. A seção 3 descreve os materiais e métodos utilizados e a amostra de dados coletados. Os resultados dos testes são apresentados na seção 4, seguida da discussão na seção 5. Finalmente, a seção 6 apresenta as considerações finais e possíveis trabalhos futuros.

2. Trabalhos Relacionados

Atualmente, existem métodos acessíveis para identificação do TDAH para pais e professores ou ao próprio portador do transtorno. Dentre eles, estão os questionários de escala que podem ser respondidos por pais, cuidadores e/ou professores. Destacamos o *ADHD Rating Scale*, o questionário de Conners e o SNAP-III e IV (MENDONÇA DE ANDRADE et al, 2011). Ressaltamos que os questionários fazem parte da primeira etapa do processo de diagnóstico do TDAH e que deve ser seguida de análise clínica por profissional de saúde especializado, ou seja, o caráter do questionário é de indicação e não de diagnóstico.

O TDAH, diferentemente do transtorno da dislexia, ainda carece de ferramentas computacionais voltadas para investigação precoce automatizada (COSTA *et al*, 2014). Nas investigações sobre dislexia, os autores utilizaram redes neurais para identificar possíveis portadores de dislexia. Nestes estudos, as redes neurais tiveram acertabilidade de aproximadamente 80% dos casos, o que foi considerado satisfatório.

Outro trabalho relacionado foi apresentado por (MIRANDA *et al*, 2011), onde os autores aplicaram o questionário MTA-SNAP-IV em um grupo de crianças com o objetivo de testar a eficácia do questionário. Foi possível indicar os sintomas presentes na amostra de crianças que poderiam levar ao diagnóstico de TDAH.

Tenev et al (2013) utilizaram técnicas computacionais não-supervisionada do tipo SVM para classificar sub-tipos de TDAH em adultos, através do eletroencefalograma em

diferentes situações. Os dados procedem de 67 pessoas previamente diagnosticadas com TDAH e outras 50 não-portadoras de qualquer transtorno neurológico. Os padrões foram definidos através da SVM por intermédio da lógica de Karnot produzindo a saída dos classificadores. Os resultados mostraram que mais de 80% das instâncias foram classificadas corretamente.

Neste sentido, nossa contribuição no presente trabalho é acelerar o processo de diagnóstico precoce de TDAH em crianças e jovens em idade escolar, fazendo uso de diferentes técnicas computacionais e algoritmos que se mostram interessantes métodos para classificação de dados.

3. Material e Métodos

O questionário utilizado nessa pesquisa foi o MTA-SNAP-IV, livremente disponibilizado na rede com tradução em Português (Mattos *et al.*, 2006). O questionário é composto por 26 perguntas das quais 18 relacionadas ao TDAH e 8 relacionadas ao Transtorno de Oposição. Das 18 perguntas, 9 são a respeito da hiperatividade e impulsividade, e as outras 9 sobre desatenção. Além destas, foram adicionadas 9 perguntas relacionadas ao comportamento de indivíduos de forma geral para fins de teste.

O questionário é aplicado a pais e professores, o que é importante, pois para diagnóstico de TDAH o indivíduo precisa ser analisado em relação a mais de um ambiente no qual está inserido. Além disso, como o desempenho acadêmico da criança portadora de TDAH é geralmente afetado, os professores tendem a perceber certos sintomas da criança no seu dia-a-dia.

3.1. Amostra de dados

O público selecionado foi composto por jovens da faixa etária entre 8 a 18 anos residentes na cidade do Rio de Janeiro, Brasil e em Salamanca, Espanha. Foram avaliados um total de 52 crianças/adolescentes. A tabela a seguir ilustra a distribuição das faixas etárias dos entrevistados.

Tabela 1. Idade da população alvo, adaptada de COSTA, 2011.

Faixa etária	Total por faixa etária	%
9-12	16	30
13-16	15	30
16 ou mais	21	40
Total	52	100

3.2. Tipo de pesquisa

Esta pesquisa tem caráter experimental qualitativo. De acordo com Wazlawick (2014, p. 23), este tipo de pesquisa implica ter uma ou mais variáveis experimentais que podem ser controladas pelo pesquisador, e uma ou mais variáveis observadas, cuja medição poderá levar à conclusão de que existe algum tipo de dependência com a variável experimental. Ou seja, dados provenientes dos entrevistados foram manipulados e analisados por ambos o algoritmo de clusterização *K-means* e a rede neural não supervisionada do tipo SOM, e a partir deste processo, é possível extrair informações como a influência de determinadas variáveis atuam no resultado obtido.

A forma de coleta de dados se deu através de questionário de escala, então foi necessário realizar um processo de quantificação para que os qualificadores tivessem compatibilidade com os requisitos do algoritmo. O questionário possui os qualificadores “*nem um pouco*”, “*pouco*”, “*bastante*”, “*demais*.” Estes foram mapeados em números correspondentes (0,1,2,3) em um banco de dados para então servirem de input para o *k-means* e a rede neural em questão (SOM).

3.3. Algoritmo de clusterização *k-means* e rede neural não-supervisionada

O *k-means* é um algoritmo clássico relativamente simples, trabalha agrupando objetos em k números de grupos, com $k > 1$. O número de grupos k que se deseja encontrar nos dados selecionados deve ser definido previamente pelo usuário (JAIN, MURTY e FLYNN, 1999; KANUNGO, et al, 2002). O algoritmo calcula os centros de um grupo (centróides). Neste trabalho, optamos por 3 clusters ($k=3$), pois foi o número mínimo necessário para assegurar a separação das variáveis em clusters distintos.

Nos algoritmos de aprendizado não-supervisionado não existe o papel do especialista treinando a rede. Neste tipo de rede, são disponíveis apenas padrões de entrada de treinamento de classificação desconhecida. O algoritmo avalia os conjuntos de dados apresentados, determina algumas propriedades dos conjuntos de dados e aprende estas propriedades na sua saída. O método de agrupamento de cada padrão segue um critério de similaridade e depende do algoritmo empregado, dos dados utilizados e da medida de similaridade adotada (CARVALHO, 2005).

Neste trabalho, optamos por utilizar algoritmo de rede neural não-supervisionada do tipo SOM (*Mapas Auto-Organizáveis*), também conhecidas como *Mapas de Kohonen*. As redes SOM utilizam o algoritmo “vencedor-leva-tudo”, o que significa que apenas um neurônio com maior número de ativação irá fornecer a saída da rede, e somente seus pesos são alterados. Durante o chamado treinamento competitivo, o neurônio cujo peso for mais próximo do vetor de entrada será o vencedor, e esta proximidade é calculada como *distância euclidiana*. Para tal, a seguinte fórmula é utilizada:

$$d_i = \sum_{j=1}^n (W_{ij} - X_j)^2$$

onde n é o número de neurônios de entrada (elementos vetor de entrada). A rede repetirá o cálculo de distâncias até que os pesos sejam bem próximos aos do vetor de entrada. O objetivo deste processo é agrupar dados em diversos grupos próximos ou clusters (COPIN, 2013). Como não definimos previamente a quantidade de classes a serem

identificadas pela rede, os Mapas Auto-Organizáveis se mostraram uma opção apropriada.

4. Resultados

As amostras de dados, como citado anteriormente, provém de questionário MTA-SNAP-IV com os quatro quantificadores. Realizamos dois conjuntos de experimentos (algoritmo de clusterização e rede neural não supervisionada SOM).

O algoritmo de clusterização foi aplicado através do software MatLab (MATLAB, 2016). Como citado anteriormente, é necessário definir previamente o número de agrupamentos. Definimos inicialmente como $k=2$, e a grande maioria das variáveis foi atraída apenas para um cluster, o que não permitiu uma classificação clara das variáveis. A partir de três clusters ($k=3$), pudemos classificar diversas variáveis em diversos clusters. Verificamos experimentalmente que com apenas 3 já seria suficiente, pois nosso objetivo era ter as classes: “*possível candidato a TDAH*”, “*não é candidato*”, “*possível candidato a outro transtorno*”.

Na figura 1, percebe-se que o algoritmo produz como resultado uma imagem em 3D, contendo todas as variáveis (círculos menores) e os três clusteres representados pelos círculos maiores (centróides com cores azul, verde e vermelho). As cores das variáveis indicam que elas estão ligadas ao cluster da cor correspondente. Na figura 1, os pequenos pontos da imagem representam todos os indivíduos avaliados. Pode-se observar uma grande proximidade de um indivíduo ao cluster verde.

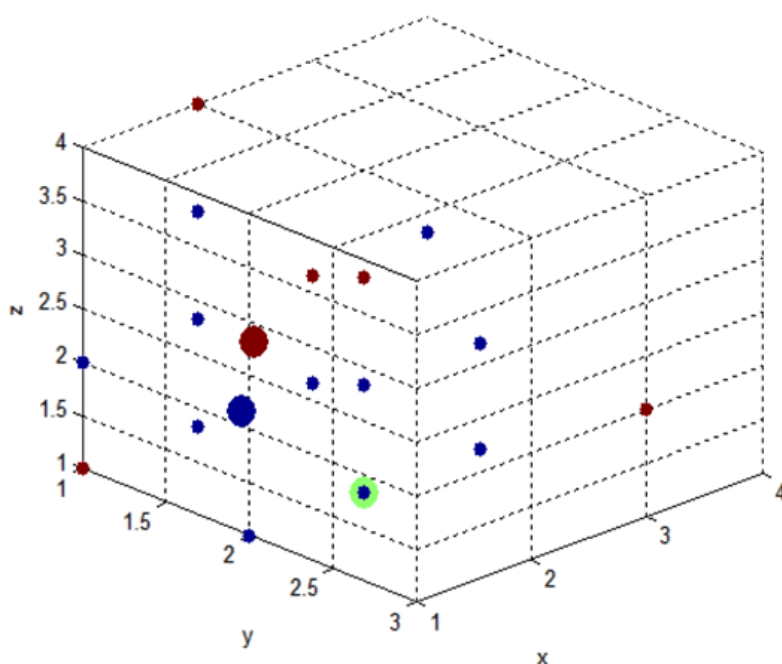


Figura 1. Agrupamento em três clusters (verde, azul e vermelho) do algoritmo *k-means*

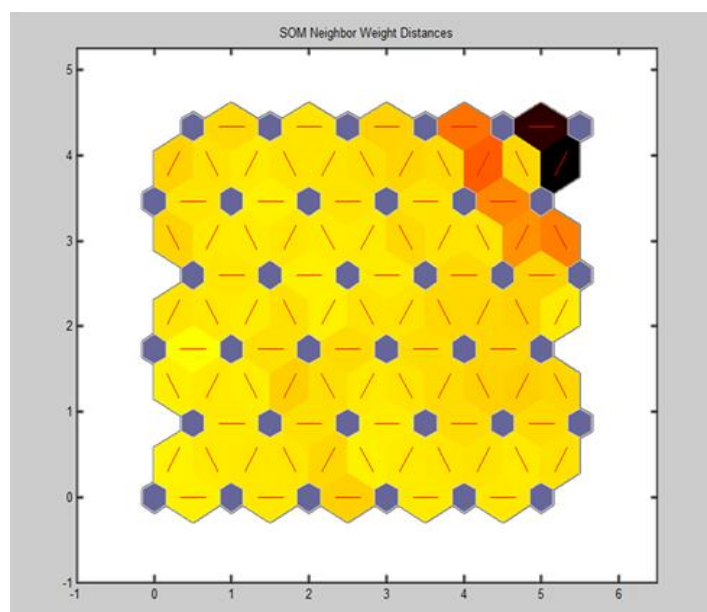


Figura 2. Mapa de vizinhança das variáveis

A rede neural utilizada neste experimento do tipo não supervisionada (Mapas Auto Organizáveis-SOM) foi processada através do software MatLab (MATLAB, 2016). De acordo com a Figura 2, que ilustra o mapa de vizinhanças (Neighbor Weight Distances) como saída da rede SOM, observa-se que foi possível separar em classes registros com características bem definidas. Percebe-se a presença marcante de três registros bem próximos situados no canto superior direito do mapa (hexágonos roxo). Porém existe apenas um candidato que se apresenta de maneira acentuada destacando-se dos demais. Este candidato (hexágono) encontra-se situado na extremidade do canto superior direito da Figura 2. Dessa forma, podemos inferir que é possível que estes três indivíduos sejam fortes candidatos em risco de TDAH.

Nestes experimentos estamos interessados em identificar possíveis candidatos em risco de TDAH. Percebe-se que apenas um indivíduo foi identificado como muito próximo a um cluster (Figura 1, círculo verde maior). Na figura 2, o mesmo indivíduo foi classificado pela rede SOM como um forte candidato em risco de TDAH.

Após inúmeras execuções do *k-means* ($n > 20$), o algoritmo permaneceu exibindo o mesmo resultado. A rede SOM foi executada com 200 épocas, também apresentando o mesmo resultado. Constatou-se que o mesmo o candidato destacou-se nos diferentes algoritmos. Estes resultados permitem inferir que se trata de um forte candidato ao risco de TDAH. Porém, isso só poderá ser confirmado após o diagnóstico clínico do especialista.

5. Discussão

Os resultados obtidos através de nossos experimentos podem ser considerados satisfatórios uma vez que os dois algoritmos convergiram após diversas execuções. Através do algoritmo *k-means* percebemos que um dos indivíduos se destacou sendo o único de uma das três classes; pudemos concluir que este é um candidato forte ao risco

de TDAH. A rede não supervisionada apontou uma variável como sendo a mais destacada, como visto na figura 2. Assim, percebemos que os resultados dos dois algoritmos apontam para a existência de um mesmo candidato ao risco de TDAH.

Neste trabalho nenhum dos entrevistados passou *a priori* pelo processo de diagnóstico clínico com os especialistas. Portanto, será necessário que os possíveis candidatos apontados pelos experimentos sejam avaliados clinicamente.

Ressaltamos algumas dificuldades encontradas na realização desta pesquisa. Como citado anteriormente, não foram encontrados padrões de indivíduos diagnosticados, assim como os entrevistados não possuem diagnóstico confirmado de TDAH feito por especialista. Na busca por padrões, foi necessária uma análise mais detalhada dos resultados algorítmicos.

6. Conclusões

O TDAH pode atingir até 5% de crianças em idade escolar e ocasionar perdas acadêmicas e sociais. Desta forma, surge a oportunidade de pesquisar estratégias computacionais voltadas para esse público. Este trabalho investigou o uso de técnicas de Inteligência Artificial baseadas em algoritmo de clusterização *k-means* e em rede neural não-supervisionada do tipo SOM como uma forma de indicar indivíduos em risco de TDAH, auxiliando seu diagnóstico precoce.

Esse trabalho explorou por meio de dois experimentos um dataset composto de dados obtidos de pais respondentes. Os resultados obtidos se mostraram consistentes pois um indivíduo foi indicado como possível candidato em risco de TDAH pelos dois algoritmos avaliados, mesmo após diversas execuções ($n > 20$).

Como trabalhos futuros, pretendemos desenvolver um sistema que incorpore as técnicas de Inteligência Artificial para processar dados de novos possíveis candidatos a partir de padrões de diagnosticados clinicamente. O sistema poderá ser utilizado no ambiente escolar pela equipe pedagógica para auxiliar na detecção precoce de crianças e adolescentes em risco de TDAH.

Referências

- APA DSM IV TR – Manual Diagnóstico e Estatístico de Transtornos Mentais, 4ª ed. revisada. Porto Alegre: ArtMed, 2002.
- BRASIL. Lei de Diretrizes e Bases da Educação Nacional n. 9394/96, artigo 58, de 20 de dezembro de 1996.
- Carvalho, L. A. V. (2005) Datamining – A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração. Rio de Janeiro: Editora Ciência Moderna Ltda.
- COPPIN, Ben. Inteligência Artificial. Tradução e revisão técnica Jorge Duarte Pires Valério. – [Reimpr.] – Rio de Janeiro: LTC, 2013.
- COSTA, M. Uma estratégia computacional na detecção de dislexia. UFRJ, 2011.

- COSTA et al. Desafios e Oportunidades em Neurociência Computacional na Educação Brasileira. 3º Seminário Grandes Desafios da Computação. SBC, 2014. Disponível em < <https://goo.gl/tCBbh8>>. Acesso em 20 de Jun 2016.
- Mattos et al. Apresentação de uma versão em Português para uso no Brasil do instrumento MTA-SNAP-IV de avaliação de sintomas de Transtorno de déficit de atenção/hiperatividade e sintomas de transtorno desafiador e de oposição. Revista de Psiquiatria RS, 2006.
- MIRANDA, C. T. ; SANTOS JUNIOR, G. ; PINHEIRO, N. A. M. ; STADLER, Rita de Cassia L. . Questionário SNAP-IV: a utilização de um instrumento para identificar alunos hiperativos. In: VIII ENPEC - I CIEC, 2011, Campinas. VIII ENPEC - I CIEC, 2011.
- Rohde, L. A., & Ketzer C. R. (1997). Transtorno de déficit de atenção e hiperatividade. In N. Fichtner (Org.), Transtornos mentais da infância e adolescência (pp. 232-243). Porto Alegre: Artmed.
- Wazlawick, Raul Sidnei. Metodologia de pesquisa para ciência da computação. Campus, 2ª ed., 2014.
- KANUNGO, T. et al., 2002, An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, nº 7, July.
- JAIN, A. K., MURTY, M. N. e FLYNN, P., 1999, Data clustering: a review. ACM Computing Surveys 3 (31): 264–323.
- MatLab – The Language of Technical Computing. R2016a, MatWorks. Mais informações em www.mathworks.com – Acesso em 2016.