

Análise automatizada do discurso de aprendizes de programação: relações entre emoções e nível de experiência

Thiago Barcelos¹, Luiz Felipe Nascimento¹, Alexandra Souza^{1,2}, Leandro Silva², Roberto Muñoz³

¹ Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP
Av. Salgado Filho, 3501, Guarulhos, SP, Brasil – CEP 07115-000

² Universidade Presbiteriana Mackenzie
Rua da Consolação, 930, São Paulo, SP, Brasil – CEP 01302-907

³ Universidad de Valparaíso – Escuela de Ingeniería Civil Informática
General Cruz, 222, Valparaíso, Chile – CP 2362905

luizfelipesilvadonascimento@gmail.com; {tsbarcelos,
alexandra.souza}@ifsp.edu.br; leandroaugusto.silva@mackenzie.br;
roberto.munoz@uv.cl

Abstract. *Techniques related to Multimodal Learning Analytics may be useful to evaluate the previous level of expertise of students engaged in learning activities related to programming. In this article we present the preliminary results of an experiment with students who describe the strategies they applied to develop two algorithms. Five emotional states were detected through an analysis of speech recordings. The results were clustered using Self-Organizing Maps and indicated that speech excerpts associated to sadness and fear were related to participants with less experience in programming or with a lower self-evaluation.*

Resumo. *A utilização de técnicas de Análise Multimodal da Aprendizagem pode ser útil para avaliar o nível de experiência prévia do aluno em atividades de aprendizagem relacionadas à programação. Neste artigo são apresentados os resultados preliminares de um experimento de avaliação automatizada do discurso de aprendizes que descrevem suas estratégias para construção de dois algoritmos. A presença de cinco estados emocionais foi estimada pela análise da gravação da fala. O agrupamento dos resultados por meio de Mapas Auto-Organizáveis indicou que a presença de trechos do discurso relacionados à tristeza e medo estão relacionados a participantes com menor nível de experiência em programação ou a uma auto-avaliação mais baixa de suas habilidades.*

1. Introdução

Com a chegada da era digital começou a se pensar em quais habilidades seriam necessárias para os alunos que iriam fazer parte dessa sociedade futuramente. Dentre as habilidades relevantes para o século 21 destacam-se o pensamento crítico, capacidade de solução de problemas complexos, criatividade, comunicação eficiente e alta produtividade. Face a esse cenário diversas organizações começaram a destacar a

necessidade do desenvolvimento dessas habilidades por meio da educação, dando um espaço maior para metodologias centralizadas no aluno como a Aprendizagem Baseada em Projetos, entre outras (BARRON; DARLING-HAMMOND, 2010). Ao mesmo tempo, Worsley e Blikstein (2010a) destacam a exigência oriunda das políticas educacionais para a produção de evidências da evolução do aluno. Segundo os autores, essas exigências têm se tornado um problema para os educadores que pode ser solucionado com a criação de ferramentas de avaliação automatizada.

Situações em que o aluno produz soluções únicas, ou interagem em grupos também são difíceis de avaliar e gerar *feedback*. Porém, Blikstein (2013) destaca que com a análise de informações multimodais podem surgir novas maneiras de avaliar o que ocorre com aluno em situações que produzem soluções mais complexas e únicas. Essas técnicas têm sido denominadas de Análise Multimodal da Aprendizagem. Além de mostrarem o processo de aprendizagem do aluno, permitem também aos pesquisadores desenvolverem novas estruturas ao ensino. Segundo o autor, com a utilização da tecnologia vigente como, sensores portáteis, câmeras portáteis, sensores de gestos entre outras, é possível acompanhar detalhadamente a evolução do aprendizado do aluno. A utilização dessas diversas informações adquiridas a partir da atividade humana permite uma análise mais fiel das tarefas complexas.

No entanto, é necessário destacar que essas técnicas ainda não estão muito populares no ramo de análise do aprendizado. No contexto do desenvolvimento do Pensamento Computacional, atividades didáticas que envolvem a colaboração e a construção de artefatos são comuns (RODE et al., 2015); ainda, muitas vezes as atividades propostas demandam dos alunos o desenvolvimento de competências e habilidades relacionadas à programação. A análise automatizada de dados provenientes da fala de alunos que descrevem suas estratégias cognitivas para a resolução de problemas tem se mostrado como uma estratégia promissora, dado que trabalhos anteriores (HAIDER et al., 2016; WORSLEY; BLIKSTEIN, 2010b) mostraram ser possível inferir o estado emocional do entrevistado bem como suas dúvidas e incertezas. Dessa forma, este trabalho tem como objetivo apresentar, dentro da perspectiva de uma Análise Multimodal da Aprendizagem, os resultados preliminares de um experimento com alunos de cursos na área de Computação com diferentes níveis de experiência em programação onde as ocorrências de emoções na fala foram correlacionadas com o nível de experiência prévio dos participantes.

2. Trabalhos relacionados

A análise automatizada da fala tem sido utilizada dentro do processo de análise da aprendizagem com o objetivo de inferir o desempenho, as dúvidas ou mesmo o estado emocional do aprendiz. Worsley e Blikstein (2010b) descrevem um experimento no qual alunos foram convidados a descrever verbalmente suas soluções para problemas de engenharia; características prosódicas (frequência, entonação e formantes) foram utilizados juntamente com uma análise do texto obtido a partir do reconhecimento da fala para inferir o nível de experiência dos participantes naquele domínio. Mais recentemente, Haider *et al.* (2016) analisam as características prosódicas juntamente com análise do gestual para inferir a qualidade intrínseca de apresentações verbais.

Considerando especificamente o desenvolvimento do Pensamento Computacional e de competências relacionadas à programação, as emoções identificadas durante o processo de aprendizagem podem trazer indícios de uma maior ou menor efetividade desse processo. Bosch e D’Mello (2017) reportam um estudo com 99 estudantes que tiveram sua primeira experiência com programação de computadores registrada em vídeo e áudio; suas conclusões indicam que a transição entre estados de engajamento e confusão, e entre estados de tédio e engajamento estavam, respectivamente, relacionados a eventos inesperados (“bugs” no código) e orientação do instrutor para sanar dúvidas. No entanto, no estudo os estados emocionais foram identificados *a posteriori* por meio da rotulação de momentos das gravações pelos próprios participantes.

É importante destacar que Bosch e D’Mello (2017) postulam que o estado emocional de engajamento relacionado ao “fluxo” (*flow*) (CSIKSZENTMIHALYI, 1990) seria o ideal em um processo de aprendizagem de programação. No tocante ao desenvolvimento do Pensamento Computacional, Basawapatna *et al.* (2013) também trabalham com essa mesma hipótese na proposta de um modelo para aprendizagem escalonada. Por outro lado, Brennan e Resnick (2012) indicaram que a análise do discurso de aprendizes relacionado às estratégias de construção de artefatos computacionais pode trazer evidências do desenvolvimento do Pensamento Computacional. Dessa forma, esta pesquisa se situa na confluência entre os trabalhos anteriores que buscaram a análise automatizada do discurso, porém com o enfoque na identificação automatizada das emoções expressadas durante uma atividade de programação de computadores.

3. Métodos e técnicas

O experimento consistiu em recrutar participantes que tinham como pré-requisito possuir ao menos noções básicas de programação de computadores. Solicitou-se que cada participante resolvesse dois exercícios de programação, sendo o primeiro exercício de controle, o qual se esperava que todos os participantes pudessem resolver, e o segundo foi utilizado para análise comparativa. No primeiro exercício o participante deveria construir um programa que imprimisse o maior valor entre três valores lidos via teclado. No segundo exercício o participante precisava construir um programa que imprimisse os *N* primeiros números da sequência de Fibonacci. Após a realização dos exercícios cada participante foi convidado a explicar, da forma mais completa possível, como resolveu os problemas por meio de uma entrevista gravada. Durante a entrevista, a fala de cada entrevistado sofreu o mínimo de intervenção possível do pesquisador com, por exemplo, perguntas de reforço ou interrupções.

Cada participante também foi classificado com base no seu tempo de experiência com programação, relatado por meio do preenchimento de um questionário prévio. Foi classificado como iniciante o aluno com menos de um ano de experiência; como intermediário, o aluno com mais de um ano e menos de dois anos de experiência; avançado, o aluno com mais de dois anos de experiência. O questionário ainda solicitava que o participante informasse uma auto-avaliação do seu nível de experiência em programação como ruim, regular ou bom.

As entrevistas foram gravadas em áudio e editadas para eliminar quaisquer intervenções do entrevistador, mantendo unicamente a fala dos entrevistados. A seguir,

cada arquivo de áudio passou por uma normalização em lote. As edições foram realizadas utilizando o software Audacity. A detecção automatizada de emoções foi realizada com o OpenVokaturi SDK, uma biblioteca para processamento de áudio que utiliza uma rede neural treinada com dois bancos de dados de gravações de voz previamente rotulados para cinco classes de emoções: alegria, tristeza, raiva, medo e neutralidade (VOKATURI DEVELOPERS, 2017). Os arquivos de áudio foram processados pela biblioteca utilizando-se um script que segmenta a análise em intervalos de um segundo. Dessa forma, a cada segundo de análise foram geradas probabilidades de ocorrência de cada uma das cinco emoções listadas anteriormente.

Após a análise, o conjunto de dados composto pela identificação do participante, exercício de programação ao qual se refere o trecho da fala (1 ou 2) e as probabilidades de ocorrência dos cinco estados emocionais no trecho em questão foram utilizados para o treinamento de uma rede neural baseada em Mapas Auto-Organizáveis (KOHONEN, 2013) com o objetivo de permitir a visualização das relações entre as ocorrências de estados emocionais e o nível de experiência dos participantes. O treinamento da rede consiste em submeter iterativamente o conjunto de dados de entrada a um modelo de aprendizagem competitiva, na qual os neurônios disputam entre si a representatividade dos objetos do conjunto de treinamento, num processo que envolve a adaptação dos pesos da rede a partir dos estímulos fornecidos pelos dados de entrada. Ao final do treinamento cada neurônio representa um subconjunto de objetos usado no treinamento, que são semelhantes entre si, mas que também possuem características similares aos neurônios mais próximos.

A análise foi realizada por intermédio da linguagem R com o pacote Kohonen. A base de treinamento foi composta por 1055 registros e o treinamento foi executado em 2000 iterações (épocas). A topologia do mapa foi definida experimentalmente buscando a menor distância média entre nós após o treinamento, e dessa forma optou-se pela dimensão de 5x5, com vizinhança hexagonal.

4. Resultados e discussão

Participaram do experimento 8 alunos do Instituto Federal de São Paulo, sendo 2 alunos do sexo feminino e 6 do sexo masculino. A amostra também se segmenta por curso e nível, sendo que 2 alunos cursam o 4º semestre do Curso Técnico em Informática para Internet e os demais participantes estão matriculados no Curso Superior em Análise e Desenvolvimento de Sistemas, onde 4 alunos são do 1º semestre e 2 do 6º semestre. O nível de experiência obtido a partir do tempo de experiência informado pelos participantes, bem como sua auto-avaliação são apresentados na Tabela 1.

Tabela 1. Escolaridade e nível de experiência dos participantes.

#	Escolaridade	Tempo de experiência com programação	Experiência (classificação)	Experiência (Auto-avaliação)
1	Superior (6º semestre)	Mais de 3 anos	Experiente	Bom
2	Superior (6º semestre)	Mais de 3 anos	Experiente	Ruim
3	Técnico (4º semestre)	Entre 1 e 2 anos	Intermediário	Médio
4	Técnico (4º semestre)	Entre 1 e 2 anos	Intermediário	Médio
5	Superior (1º semestre)	Entre 2 e 3 anos	Intermediário	Médio
6	Superior (1º semestre)	Menos de 6 meses	Novato	Ruim
7	Superior (1º semestre)	Entre 1 e 2 anos	Intermediário	Médio
8	Superior (1º semestre)	Entre 6 meses e 1 ano	Novato	Médio

O resultado do treinamento do Mapa Auto-Organizável é apresentado a seguir na Figura 1. Pode-se observar uma tendência de agrupamento dos trechos de discurso com detecção da mesma classe de emoções: raiva (em rosa) na parte superior esquerda, medo (em branco) na parte superior direita, neutralidade (verde) na parte inferior esquerda e tristeza (bege) na parte inferior direita.

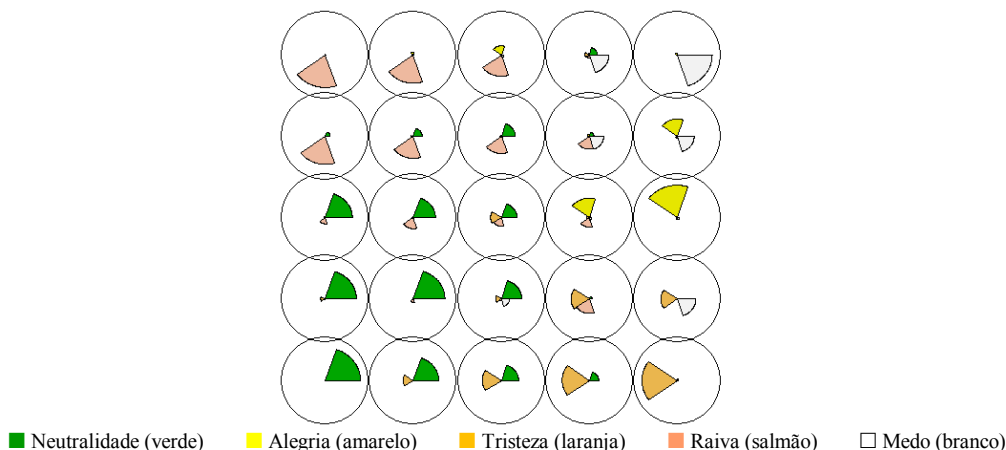


Figura 1. Mapa Auto-Organizável resultante do processo de treinamento

O próximo passo da análise foi segmentar os trechos de discurso relacionados a cada participante e à descrição de cada um dos exercícios. Na Figura 2 são apresentados os mapas com os trechos de discurso gerados por cada participante. Os números em preto referem-se a um trecho de descrição do exercício 1 e os números em vermelho, a um trecho de descrição do exercício 2.

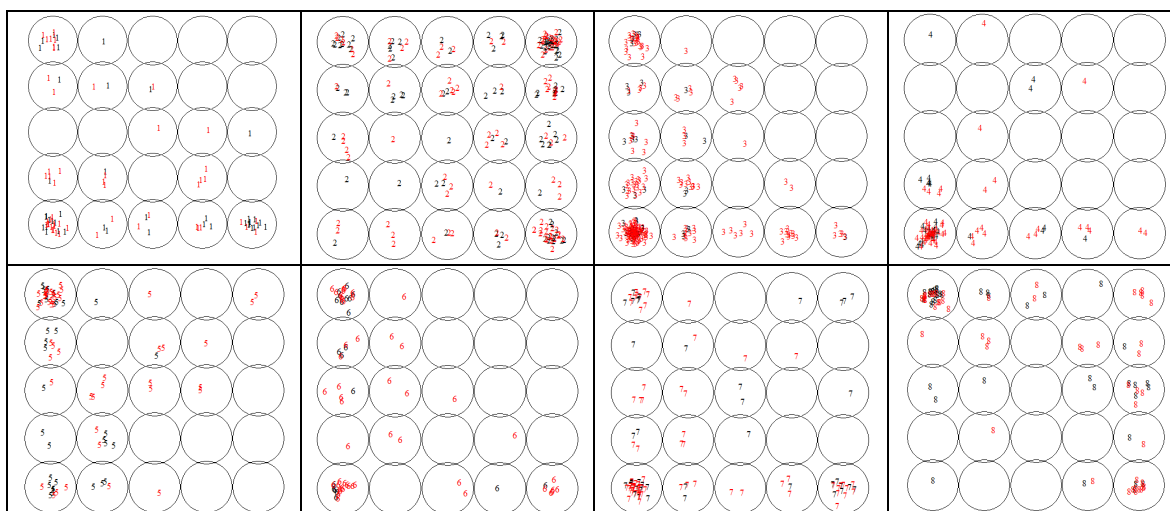


Figura 2. Trechos do discurso de cada participante relacionados ao exercício 1 (em preto) e ao exercício 2 (em vermelho)

A presença frequente de trechos de discurso com detecção de neutralidade era esperada, considerando que cada participante teve liberdade de estruturar sua explicação e que ela, muito possivelmente, incluiria a descrição objetiva de aspectos dos algoritmos que não teriam nenhuma emoção específica associada a sua verbalização. Apesar nos participantes 2 e 8 houve pouca presença de neutralidade na fala (ocupando,

respectivamente, 12,93% e 11,3% do tempo total do discurso). Em ambos os casos tratam-se de aprendizes com experiência relativamente baixa em programação – apesar do participante 2 declarar ter experiência suficiente para sua classificação como “experiente”, a sua auto-avaliação foi “ruim”. Já o participante 8 afirma ter menos de um ano de experiência com programação, apesar de ter avaliado sua própria experiência como “boa”. Esses dois participantes ainda foram aqueles que apresentaram de forma mais destacada a presença de trechos de discurso na classe de emoções de “medo”, o que pode denotar a insegurança em afirmações realizadas sobre a implementação dos algoritmos.

A classe de emoções de “tristeza” teve alguma correlação com o exercício descrito no discurso dos participantes. Em seis dos oito participantes, a frequência de trechos relacionados à descrição do exercício 2 e rotulados com “tristeza” foi maior do que a frequência de trechos relacionados ao exercício 1. Isso ocorreu com os participantes 2, 3, 4, 6 e 8. Todos esses participantes tiveram o nível de experiência classificado como “Intermediário” ou “Novato”, com exceção do participante 2 cuja auto-avaliação foi “ruim”, como mencionado anteriormente. Novamente levanta-se a hipótese da maior frequência dessa classe de emoções estar relacionada a incertezas na fala dos participantes.

Considerando a segmentação do áudio em trechos de 1 segundo, uma maior ocorrência de trechos na figura acima representa um discurso mais longo por parte do participante. Nesse ponto, observou-se a tendência de participantes com maior nível de experiência gastarem mais tempo para descrever sua solução. Os participantes classificados como experientes gastaram em média 74,5 segundos para descrever cada solução (min: 43s; max: 101s); aqueles classificados como intermediários gastaram 73,25 segundos (min: 36s; max: 211s) e os classificados como iniciantes, 42,75 segundos (min: 23s; max: 55s). Esse resultado contrasta com o experimento de Worsley e Blikstein (2010b), no qual verificou-se onde os participantes com mais experiência respondiam de forma mais concisa. Se por um lado o domínio daquele experimento era a construção de artefatos de engenharia, a alta variabilidade nos tempos de explicação na amostra obtida neste estudo exige a confirmação dessa tendência em estudos futuros.

5. Conclusões

O desenvolvimento de competências e habilidades relacionadas ao Pensamento Computacional e à programação de computadores envolve cada vez mais atividades colaborativas e complexas, o que cria novos desafios à avaliação. Técnicas relacionadas a um paradigma de Analítica Multimodal da Aprendizagem, como a análise automatizada do discurso, podem permitir a obtenção e análise de mais fatos relevantes relacionados ao processo de aprendizagem.

Neste artigo foram apresentados os resultados preliminares de um experimento relacionado à análise automatizada do discurso de oito alunos de cursos técnicos e superiores em Computação que realizaram a implementação de dois algoritmos diferentes. Os resultados indicam que a presença de trechos do discurso relacionados à tristeza e medo estão correlacionados a participantes com menor nível de experiência em programação, bem como a uma auto-avaliação mais baixa de suas habilidades. Ainda, em contraste com trabalhos anteriores, um maior tempo de fala para descrever a

implementação produzida se relacionou com um maior nível de experiência dos participantes.

Algumas limitações à validade do estudo devem ser destacadas, como o tamanho da amostra e a ausência de uma análise da semântica do discurso e da variação dos estados emocionais ao longo do tempo. No entanto, os resultados obtidos até o momento indicam a viabilidade da técnica para identificação e avaliação do nível de experiência dos aprendizes. Pretende-se que as próximas etapas da pesquisa explorem as limitações identificadas, visando a implementação da análise automatizada do discurso em um sistema de Analítica Multimodal da Aprendizagem atualmente em desenvolvimento.

Referências

- BARRON, B.; DARLING-HAMMOND, L. Prospects and challenges for inquiry-based approaches to learning. In: **The nature of learning: Using research to inspire practice**. [s.l.] OECD, 2010. p. 199–225.
- BASAWAPATNA, A. R. et al. **The Zones of Proximal Flow: Guiding Students Through a Space of Computational Thinking Skills and Challenges**. Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research. **Anais...: ICER '13**. New York: ACM, 2013 Disponível em: <<http://doi.acm.org/10.1145/2493394.2493404>>
- BLIKSTEIN, P. **Multimodal Learning Analytics**. Proceedings of the Third International conference on Learning Analytics and Knowledge (LAK '13). **Anais... In: LAK '13**. Leuven: ACM, 2013
- BOSCH, N.; D'MELLO, S. The Affective Experience of Novice Computer Programmers. **International Journal of Artificial Intelligence in Education**, v. 27, n. 1, p. 181–206, 1 mar. 2017.
- BRENNAN, K.; RESNICK, M. **New frameworks for studying and assessing the development of computational thinking**. Proceedings of the 2012 annual meeting of the American Educational Research Association. **Anais... In: AERA 2012**. Vancouver: American Educational Research Association, 2012
- CSIKSZENTMIHALYI, M. **Flow: The psychology of optimal experience**. New York: Harper and Row, 1990.
- HAIDER, F. et al. **Presentation quality assessment using acoustic information and hand movements**. IEEE, mar. 2016 Disponível em: <<http://ieeexplore.ieee.org/document/7472190/>>. Acesso em: 15 mar. 2017
- KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, v. 37, p. 52–65, jan. 2013.
- RODE, J. A. et al. **From Computational Thinking to Computational Making**. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. **Anais... In: UBIComp '15**. New York: ACM, 2015 Disponível em: <<http://doi.acm.org/10.1145/2750858.2804261>>

VOKATURI DEVELOPERS. **Measuring emotions**, 2017. Disponível em: <<https://developers.vokaturi.com/algorithms/emotions>>. Acesso em: 20 ago. 2017

WORSLEY, M.; BLIKSTEIN, P. **What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis**. Educational Data Mining 2011. **Anais...** In: 4TH INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING (EDM 2011). Eindhoven: 2010a

WORSLEY, M.; BLIKSTEIN, P. **Towards the development of learning analytics: Student speech as an automatic and natural form of assessment**. . In: 2010 ANNUAL MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION (AERA). Denver: AERA, 2010b