Inferência de idade utilizando o LIWC: identificando potenciais predadores sexuais

Rafael Guimarães Rodrigues [1], **Wladimir Wanderley Pereira**[1], **Eduardo Bezerra**[1] , **Gustavo Paiva Guedes**[1]

[1]CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
Av. Maracanã, 229 - Rio de Janeiro - RJ - Brasil.

rafael.rodrigues@cefet-rj.br, wladimir.pereira@gmail.com,
{eduardo.silva, gustavo.guedes}@cefet-rj.br

Abstract. *Social predators use the Internet to exploit children or adolescents with abusive or sexual purposes. More and more these predators use social networks to access their victims, often providing fake profiles, pretending to be adolescents. In this scenario, this study aims to analyze texts in Brazilian Portuguese to infer the age of the users. For this purpose, we used a tool named LIWC in its Brazilian Portuguese version. As a case of study, a Brazilian social network was used to perform experiments. This study focused on the analysis of texts written by adolescents and men between 25 and 45 years old, which represent the great majority of sexual predators. Results were relevant and open lacks for further studies.*

Resumo. *Predadores sociais utilizam a internet para explorar crianças ou adolescentes com propósitos abusivos ou sexuais. Cada vez mais esses predadores utilizam as redes sociais para ter acesso as suas vítimas, muitas vezes fornecendo perfis falsos para se passarem por adolescentes. Nesse cenário, o presente trabalho tem o objetivo de analisar os textos em português do Brasil para inferir a idade dos usuários. Para esse propósito, foi utilizada uma ferramenta denominada LIWC em sua versão do português do Brasil. Como estudo de caso, foi utilizada uma rede social brasileira para realizar os experimentos. O referido estudo concentrou-se na análise de textos de adolescentes e homens entre 25 e 45 anos, que representam a grande maioria dos predadores sexuais. Os resultados alcançados foram relevantes e abrem lacunas para trabalhos futuros.*

1. Introdução

Predadores sexuais online são definidos como adultos que utilizam a internet para explorar crianças ou adolescentes com propósitos abusivos ou sexuais [Potha et al. 2016]. Estudos indicam que a maioria dos predadores sexuais (*i.e.,* 75%) é constituída por indivíduos do sexo masculino [Kipane, A. 2014], de cor branca e com idade entre 25 e 45 anos [Plasencia 2000]. O Brasil ocupou o quarto lugar no ranking mundial da pornografia infantil em 2003. Esse número alarmante envolve redes internacionais de crime organizado, que financiam o sequestro de crianças com o intuito de utilizá-las em filmagens obscenas [Felipe 2006].

Os predadores sexuais de crianças (*i.e.,* pedófilos) utilizam ambientes como salas de bate-papo e redes sociais para acessar vítimas potenciais [Jackson 2008].

Muitas vezes os pedófilos fornecem perfis falsos, se fazendo passar por crianças ou adolescentes, o que facilita o acesso às suas vítimas [Peersman et al. 2011]. Entretanto, diversos estudos apontam que a expressão de certas emoções nos textos pode auxiliar na detecção do perfil desses predadores [Bogdanova et al. 2012, Clarke 2011]. É importante destacar que esses indivíduos estão, cada vez mais, utilizando a internet (*e.g.,* salas de bate-papo, redes sociais) para atrair e explorar sexualmente crianças e adolescentes [Dowdell et al. 2011].

Com a grande expansão das redes sociais online, surgiram diversos estudos a respeito do comportamento humano [Jin et al. 2013]. Tais estudos envolvem fatores de interesse no mundo inteiro, como: comunicação, segurança, comércio e privacidade [Campos et al.]. Nesse cenário também é possível destacar a relevância de estudos sobre o comportamento de pedófilos [Bogdanova et al. 2012]. Esses indivíduos possuem características distintas (*e.g.,* idade, sexo, nível de escolaridade) que podem influenciar o estilo da escrita dos textos publicados nessas redes. A identificação dessas influências mostrou-se possível em diversos trabalhos encontrados na literatura, dentre os quais podemos citar [Barbieri 2008, Nagarajan and Hearst 2009].

Grande parte dos estudos supracitados utiliza como base uma ferramenta denominada LIWC (*Linguistic Inquiry and Word Count*) [Pennebaker et al. 2001]. Essa ferramenta possibilita a extração de padrões psicolinguísticos de textos, o que pode auxiliar na busca por predadores sexuais [Parapar et al. 2012]. Dado que esses predadores podem fornecer perfis falsos, o estudo proposto em [Peersman et al. 2011] identifica esses perfis em redes sociais inferindo a idade e o sexo dos usuários a partir de seus textos. A literatura apresenta uma diversidade de estudos que inferem a idade a partir de textos utilizando o LIWC [Schwartz et al. 2013, Marquardt et al. 2014], entretanto não foram encontrados trabalhos utilizando o LIWC em português do Brasil[1]. Esse estudo se insere nesse panorama: inferir a idade dos usuários de redes sociais brasileiras utilizando o LIWC em português do Brasil, com o intuito de identificar potenciais predadores sexuais. Como objeto de estudo, foi utilizada a rede social online brasileira denominada *Meu Querido Diário*[2] (MQD). Essa rede social possui ampla quantidade de adolescentes[3] e funciona como um diário online.

O restante desse trabalho está organizado da seguinte forma: Na Seção 2, são apresentados os trabalhos relacionados. Na Seção 3, o processo de extração das características das entradas do MQD é descrito. Na Seção 4, os resultados experimentais são analisados. Por fim, na Seção 5, há uma discussão sobre as conclusões e cenários futuros de expansão do referido trabalho.

2. Trabalhos Relacionados

O trabalho realizado em [Nguyen et al. 2011] analisa o conteúdo das entradas de um blog que permite ao autor anexar, a cada uma dessas entradas, uma dentre 132 etiquetas de humor que podem evidenciar aspectos positivos (*e.g.,* alegre, feliz e grato) ou negativos (*e.g.,* descontente, triste e desconfortável). Desta forma, torna-se possível identificar

[1] A versão do LIWC em português do Brasil pode ser encontrada em [Filho 2013].

[2] http://www.meuqueridodiario.com.br

[3] Aqui são considerados como adolescentes os indivíduos com idade entre 12 anos completos e 18 anos, conforme estabelece o Estatuto da Criança e Adolescente no Brasil. O MQD possui 13.105 usuários nessa faixa de idade, o que corresponde a 22% dos usuários.

idade, estado de humor e conectividade social dos autores. Esse trabalho utiliza o LIWC para categorizar as palavras contidas nas entradas e considera que os fatores supracitados podem influenciar tanto no assunto escolhido quanto no estilo de escrita. O referido trabalho aponta, ainda, que enquanto a idade se mostra determinante para a escolha do assunto, o estado de humor influencia, de forma significativa, no estilo de escrita.

O estudo apresentado em [Peersman et al. 2011] evidencia a importância de se analisar a idade dos usuários de redes sociais para identificar perfis falsos e, consequentemente, pedófilos que se fazem passar por adolescentes. Os experimentos utilizam uma rede social belga para inferir a idade, adotando duas faixas etárias: menos de 16 anos e mais de 18 anos. Os resultados alcançaram precisão de 80.8%.

O estudo produzido em [Filho et al. 2014] visa inferir sexo e idade de usuários do Twitter, baseando-se na coleta e análise das 200 últimas entradas de um grupo de usuários. Foram separados os atributos considerados mais relevantes e, em seguida, foram aplicados alguns algoritmos de classificação. Foram obtidos bons resultados. Os autores consideram a inferência de faixa etária uma tarefa mais complexa e citam, ainda, que não foram encontrados trabalhos que visassem inferir idade para usuários que postam conteúdos em português.

É importante destacar que o presente trabalho se difere dos demais encontrados na literatura devido ao fato de utilizar o LIWC em português do Brasil para inferir a idade dos usuários, além de concentrar-se especificamente na identificação de potenciais predadores sexuais. Conforme mencionado anteriormente, o Brasil ocupou o quarto lugar no ranking mundial da pornografia infantil em 2003. Por essa razão foi utilizada uma rede social brasileira como objeto de estudo.

3. Inferência de faixa etária no uso da língua portuguesa em redes sociais utilizando o LIWC em Português do Brasil

O objetivo desta seção é descrever a extração das características das entradas dos usuários do MQD. Para isso, foi utilizado o conjunto de dados `MQD900`. Esse conjunto de dados contém 900 entradas de usuários distintos e 22649 atributos (*i.e.,* palavras diferentes). Das 900 entradas, 450 são de adultos do sexo masculino e 450 de adolescentes de ambos os sexos.

Nesse trabalho o `MQD900` foi utilizado para produzir outro conjunto de dados, denominado `MQD-AGE-2-LIWC-PT`. Para isso, em um primeiro passo as palavras do `MQD900` foram filtradas com a utilização do dicionário do LIWC em português do Brasil, ou seja, palavras não existentes no dicionário do LIWC foram descartadas. Em seguida, foi produzido um vetor $\vec{v}$ de $n = 64$ posições para representar cada entrada e proveniente do conjunto de dados `MQD900`. Cada posição de $\vec{v}$ representa uma das 64 categorias do LIWC. Assim, as categorias associadas a cada palavra p, utilizada em e, foram identificadas no dicionário do LIWC. Em seguida, para cada categoria identificada x_i em p, sua posição correspondente no vetor $\vec{v}$ foi incrementada. Dessa maneira, todas as entradas em `MQD900` foram representadas em `MQD-AGE-2-LIWC-PT`. Pode-se observar, na figura 1, a ilustração de um vetor de categorias de palavras que representa uma entrada do conjunto de dados `MQD-AGE-2-LIWC-PT` onde, por exemplo, 15 palavras se enquadraram na categoria x_2 e 22 palavras se enquadram na categoria x_5.

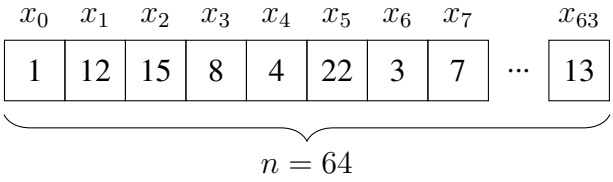

Figura 1. Vetor e_1 representando uma entrada utilizando as categorias do LIWC.

Posteriormente, os vetores foram normalizados. Para isso, foi utilizada a norma L^1, ou seja, a soma das componentes em cada vetor normalizado resulta no valor 1: $x_0 + x_1 + ... + x_{63} = 1$. O vetor normalizado da entrada e_1, representado na Figura 1, é apresentado Figura 2. Considerando os valores entre x_8 e x_{62} como zeros, podemos notar que 17% (0.17) das palavras da entrada e_1 se enquadram na categoria x_2. Analogamente, 25% das palavras se encaixam na categoria x_5. Foi produzido um novo conjunto de dados `MQD-AGE-2-LIWC-PT-N` a partir da normalização supracitada. Ambos os conjuntos de dados foram empregados para produzir modelos de classificação.

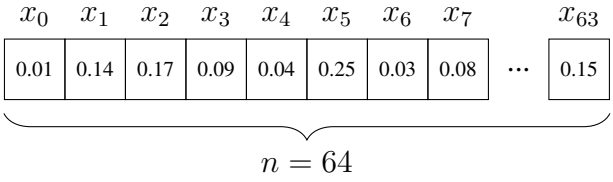

Figura 2. Vetor normalizado representando uma entrada utilizando as categorias do LIWC. Corresponde ao vetor não-normalizado ilustrado na Figura 1.

4. Resultados experimentais

Nesta seção são apresentados os resultados experimentais encontrados na inferência da faixa etária de usuários da rede social brasileira MQD utilizando o LIWC em português do Brasil. Para geração desses resultados, foram utilizados alguns algoritmos de classificação conhecidos na literatura: ZeroR, Random Forest (RF), Naive Bayes (NB), NB Multinomial (NBM), SMO e LMT. Os 5 primeiros algoritmos também foram utilizados em [Filho et al. 2014]. Para produção dos experimentos foi utilizado a ferramenta Weka [Hall et al. 2009]. Cada um dos algoritmos foi utilizado com sua configuração padrão. Os experimentos foram executados utilizando a técnica de validação cruzada denominada *k-fold validation* com dez partições. A medida *F1-score* foi utilizada para a avaliação dos resultados obtidos.

A Tabela 1 descreve os resultados obtidos. Os valores em negrito indicam os melhores F1 para os algoritmos apresentados. O algoritmo ZeroR foi utilizado como *baseline*. Os algoritmos RF e LMT apresentaram melhor F1 para o conjunto de dados `MQD-AGE-2-LIWC-PT`, com o LMT sendo um pouco superior ao RF. Para o conjunto `MQD-AGE-2-LIWC-PT-N`, os algoritmos SMO e LMT apresentaram os melhores resultados, com o LMT apresentando uma pequena melhoria com relação ao SMO.

Vale destacar que todos os algoritmos apresentaram resultados superiores ao *baseline*, indicando que a utilização do LIWC em Português do Brasil pode auxiliar na inferência da idade de usuários de redes sociais brasileiras.

Tabela 1. Classificação de faixa etária - Média F1

	ZeroR	RF	NB	NBM	SMO	LMT
MQD-AGE-2-LIWC-PT	0.333	0.718	0.519	0.710	0.716	**0,729**
MQD-AGE-2-LIWC-PT-N	0.333	0.711	0.519	0.693	0.715	**0,729**

5. Conclusões e trabalhos futuros

A contribuição desse trabalho consiste em inferir a idade de usuários de redes sociais brasileiras, considerando as diferenças linguísticas e psicológicas observadas nos textos, objetivando a identificação de potenciais predadores sexuais. Para isso, utilizamos o LIWC em português do Brasil. A principal motivação consiste no auxílio à identificação de falsos perfis, muitas vezes cadastrados por predadores sexuais. Neste trabalho, os grupos estudados foram divididos em adolescentes (*i.e.,* publico alvo) e adultos do sexo masculino com idade entre 25 e 45 anos (*i.e.,* potenciais predadores sexuais).

Para realizar esse trabalho, foram produzidos dois conjuntos de dados a partir de textos de usuários de uma rede social brasileira. Esses conjuntos de dados foram avaliados com seis algoritmos de classificação: ZeroR, RF, NB, NBM, SMO e LMT. O melhor resultado para ambos os conjuntos foi alcançado pelo algoritmo LMT. Nesse cenário, observa-se que esses resultados podem servir como base para trabalhos que infiram a idade de usuários em português do Brasil e, principalmente, para trabalhos com foco na identificação de predadores sexuais do sexo masculino. Os resultados preliminares foram considerados satisfatórios.

Durante o desenvolvimento desse trabalho, emergiram algumas ideias para trabalhos futuros, dentre elas, considerar outras faixas etárias na inferência da idade de usuários. Também seria possível investigar e propor outros modelos que melhorem os resultados. Conforme mencionado, já existem trabalhos similares utilizando o dicionário do LIWC em outras línguas. No entanto, não foram encontrados trabalhos que utilizam a versão do LIWC em português do Brasil e com o objetivo de identificação de potenciais predadores sexuais. Vale ressaltar que, como trabalho futuro, pretende-se realizar a comparação com outras abordagens da literatura.

Referências

[Barbieri 2008] Barbieri, F. (2008). Patterns of age-based linguistic variation in american english. *Journal of Sociolinguistics*, 12(1):58–88.

[Bogdanova et al. 2012] Bogdanova, D., Rosso, P., and Solorio, T. (2012). On the impact of sentiment and emotion based features in detecting online sexual predators. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 110–118. Association for Computational Linguistics.

[Campos et al.] Campos, G. E., Costa, H., and Monlevade-MG-Brasil, J. Caracterização dos perfis comerciais na rede social instagram.

[Clarke 2011] Clarke, R. (2011). *Antisocial Behavior: Causes, Correlations and Treatments*. Psychology of emotions, motivations, and actions series. Nova Science Publishers.

[Dowdell et al. 2011] Dowdell, E. B., Burgess, A. W., and Flores, J. R. (2011). Original research: online social networking patterns among adolescents, young adults, and sexual offenders. *AJN The American Journal of Nursing*, 111(7):28–36.

[Felipe 2006] Felipe, J. (2006). Afinal, quem é mesmo pedófilo. *Cadernos Pagu*, 26:201–223.

[Filho et al. 2014] Filho, R. M., Carvalho, A. I., and Pappa, G. L. (2014). Inferência de sexo e idade de usuários no twitter.

[Filho 2013] Filho, Pedro P. Balage;Pardo, T. A. S. R. M. A. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis.

[Hall et al. 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

[Jackson 2008] Jackson, L. A. (2008). Adolescents and the internet. *The changing portrayal of American youth in popular media*, pages 377–410.

[Jin et al. 2013] Jin, L., Chen, Y., Wang, T., Hui, P., and Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine*, 51(9):144–150.

[Kipane, A. 2014] Kipane, A. (2014). Woman as a sexual offender – reality or myths? *SHS Web of Conferences*, 10:00016.

[Marquardt et al. 2014] Marquardt, J., Farnadi, G., Vasudevan, G., Moens, M.-F., Davalos, S., Teredesai, A., and De Cock, M. (2014). Age and gender identification in social media. In *Proceedings of CLEF 2014 Evaluation Labs*, pages 1129–1136.

[Nagarajan and Hearst 2009] Nagarajan, M. and Hearst, M. A. (2009). An examination of language use in online dating profiles. In *ICWSM*.

[Nguyen et al. 2011] Nguyen, T., Phung, D., Adams, B., and Venkatesh, S. (2011). Prediction of age, sentiment, and connectivity from social media text. In *International Conference on Web Information Systems Engineering*, pages 227–240. Springer.

[Parapar et al. 2012] Parapar, J., Losada, D., and Barreiro, A. (2012). A learning-based approach for the identification of sexual predators in chat logs. In *Conference and Labs of the Evaluation Forum: PAN 2012 Lab Uncovering Plagiarism, Authorship, and Social Software Misuse*.

[Peersman et al. 2011] Peersman, C., Daelemans, W., and Van Vaerenbergh, L. (2011). Predicting age and gender in online social networks. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents*, pages 37–44. ACM.

[Pennebaker et al. 2001] Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). *Linguistic Inquiry and Word Count*. Lawerence Erlbaum Associates, Mahwah, NJ.

[Plasencia 2000] Plasencia, M. M. (2000). Internet sexual predators: Protecting children in the global community. *J. Gender Race & Just.*, 4:15.

[Potha et al. 2016] Potha, N., Maragoudakis, M., and Lyras, D. (2016). A biology-inspired, data mining framework for extracting patterns in sexual cyberbullying data. *Knowledge-Based Systems*, 96:134–155.

[Schwartz et al. 2013] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.