

Human Values Classification in Social Network Using Machine Learning

Sérgio Ewerton Barbosa Correia¹, Yuri Malheiros¹,
Anderson Mesquita do Nascimento², Valdiney Veloso Gouveia²

¹Departamento de Ciências Exatas – Universidade Federal da Paraíba (UFPB)
Rio Tinto - PB - Brasil

²Departamento de Psicologia – Universidade Federal da Paraíba (UFPB)
João Pessoa - PB - Brasil

{sergio.ewerton, yuri}@dcx.ufpb.br, {andermnascimento, vvgouveia}@gmail.com

Abstract. *Twitter receives millions of messages daily from different users and regions of the world. These messages contain opinions, evaluations, and feelings, making Twitter a suitable platform for various types of study. Thus, we can use this social network to discover important information about people, for instance, emotional state, their opinions about products, political preference, etc. This work aims to create and evaluate models for classification of human values, a psychology concept that defines the guide principles of people. We tested three machine learning classifiers, trained using Twitter messages, to predict the answers of a questionnaire used by psychologists to determine human values. The best result reached 60,08% of accuracy.*

Resumo. *O Twitter recebe milhões de mensagens diariamente de diferentes usuários e regiões do mundo. Estas mensagens contêm opiniões, avaliações e sentimentos, tornando o Twitter uma plataforma apropriada para diversos tipos de estudos. Assim, pode-se usar esta rede social para descobrir informações importantes sobre as pessoas, por exemplo, estado emocional, suas opiniões sobre produtos, preferência política, etc. Este trabalho tem como objetivo criar e avaliar modelos para classificação de valores humanos, um conceito da psicologia que define os princípios guia das pessoas. Foram testados três classificadores de aprendizagem de máquina, treinados usando mensagens do Twitter, para prever respostas de um questionário usado por psicólogos para determinar valores humanos. O melhor resultado alcançou uma acurácia de 60,08%.*

1. Introdução

Atualmente, cada vez mais pessoas utilizam redes sociais para expor opiniões e sentimentos, sendo o Twitter uma das plataformas mais usadas para esse fim. Com mensagens de até 280 caracteres e cerca de 335 milhões de usuários ativos mensalmente, o Twitter recebe uma enorme quantidade de conteúdo todos os dias se tornando um ambiente apropriado para execução de diversos tipos de pesquisas [Twitter 2018]. Além disso, o Twitter disponibiliza uma plataforma para desenvolvedores que oferece APIs, ferramentas e recursos que possibilitam aproveitar essa rede social de uma forma automatizada.

Analisar textos compartilhados em redes sociais pode nos mostrar muitas informações sobre uma pessoa, por exemplo, seu estado emocional, suas opiniões sobre

produtos [Pak and Paroubek 2010], preferência política [Ceron et al. 2014] e, até mesmo, encontrar sinais de distúrbios mentais [Coppersmith et al. 2014]. Sabendo disso, a psicologia tem um papel fundamental nos estudos envolvendo pessoas em redes sociais digitais.

Especificamente na psicologia social, a teoria funcionalista dos valores humanos se propõe a investigar os princípios que guiam a vida de um indivíduo através da resposta a um questionário com 18 itens [Medeiros et al. 2011]. Esta teoria preconiza duas funções consensuais dos valores. A primeira é guiar o comportamento humano, que resulta em três tipos de orientações valorativas: social, central e pessoal. Já a segunda diz respeito a expressar cognitivamente as necessidades humanas, que resulta em dois tipos de motivadores: humanitário e materialista. Como pode ser visto na Figura 1, a junção dessas duas funções cria seis subfunções de valores (normativa, realização, existência, suprapessoal, interativa e experimentação) [Gouveia et al. 2009].

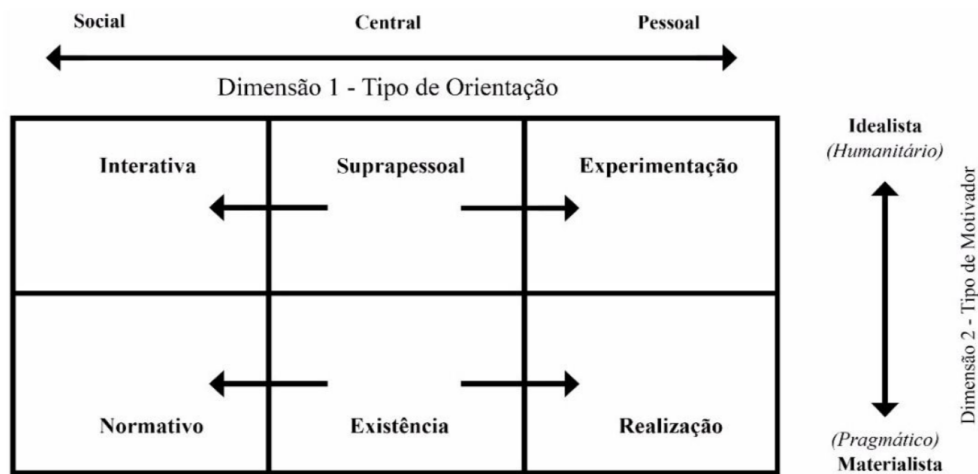


Figura 1. Matriz das duas funções consensuais dos valores [Gouveia et al. 2009]

Este trabalho tem como objetivo desenvolver e avaliar modelos de classificação para valores humanos. Dessa forma, é analisada a viabilidade de usar aprendizagem de máquina para prever os resultados do questionário de valores humanos de acordo com mensagens do Twitter. Com um classificador dessa natureza, é possível definir os valores humanos de uma pessoa apenas analisando mensagens compartilhadas por ela, sem que seja necessário a adição de uma ou mais atividades extras. Assim, pode-se facilitar estudos para a compreensão e explicação de fenômenos psicológicos e sociais.

O restante do trabalho estrutura-se da seguinte forma. Na segunda seção a teoria dos valores humanos é explicada. Na terceira seção, apresenta-se a metodologia aplicada, descrevendo o processo da coleta dos dados, pré-processamento dos dados e treinamento do classificador. Na quarta, são apresentados os resultados alcançados. Na quinta seção, os resultados são discutidos. Na sexta seção, aborda-se os trabalhos relacionados. Na sétima e última seção, são apresentados a conclusão e os trabalhos futuros.

2. Valores Humanos

Os valores humanos podem ser definidos como características abstratas duradouras que servem como princípios que guiam a vida dos indivíduos. Eles são de suma importância

para a compreensão de uma gama de fenômenos sócio-psicológicos, tais como atitudes e comportamentos ambientais, religiosidade, preconceito, consumo de drogas, comportamentos antissociais, delinquência juvenil, atitudes frente à tatuagem, intenção de cometer suicídio, uso de água e atitudes pró-ambientais [Medeiros et al. 2011].

Sobretudo desde a segunda metade do século passado, vários teóricos se dedicaram a entender como os valores estão organizados na mente dos indivíduos, porque as pessoas dão diferentes níveis de importância a valores distintos e seu impacto na formação de atitudes e comportamentos. Dentre as teorias existentes, a teoria funcionalista dos valores humanos [Gouveia 2013] enfoca em duas funções principais dos valores para definir a maneira como estes se agrupam: (1) servir como critério de orientação que guia o comportamento das pessoas e (2) expressar cognitivamente as necessidades dos indivíduos.

De acordo com esta teoria, foi desenvolvido um questionário de autorrelato para mensurar o nível de importância que as pessoas atribuem a determinados valores que compõem a medida. O questionário dos Valores Básicos foi inicialmente desenvolvido em português e espanhol contendo 66 itens. Em seguida, foram criadas duas versões abreviadas em português. A primeira deixou o questionário com 24 itens e na segunda esse número caiu para 18, sendo esta a mais utilizada atualmente [Medeiros et al. 2011]. Cada um dos 18 itens do questionário pode ser respondido utilizando uma escala likert que varia de 1 (Totalmente não importante) a 7 (Totalmente importante).

A teoria funcionalista dos valores humanos conta com uma estrutura definida para as funções e subfunções dos valores, que pode ser observada na Figura 1. Tal estrutura é a combinação de duas dimensões principais, a primeira no eixo horizontal, correspondendo ao tipo de orientação (social, central ou pessoal) e a segunda representada no eixo vertical, correspondendo aos tipos de motivadores (materialista ou humanitário). As setas que saem do centro para as laterais, representam a referência da fonte principal das subfunções [Gouveia et al. 2009]. O cruzamento dessas duas funções gera seis subfunções de valores, configurando possíveis resultados para o questionário que estão detalhados na Tabela 1.

3. Metodologia

O presente trabalho tem como objetivo desenvolver e avaliar classificadores para valores humanos baseados em mensagens do Twitter. Para isso, inicialmente foram coletadas mensagens compartilhadas na rede social e respostas do questionário de valores humanos. Em seguida os dados foram pré-processados para serem usados no treinamento de classificadores de aprendizagem de máquina. Por fim, utilizou-se validação cruzada para obtenção dos resultados da classificação. As subseções a seguir explicam em detalhes cada um desses passos.

3.1. Coleta dos dados

O Twitter disponibiliza uma plataforma para desenvolvedores que fornece diversas APIs, ferramentas e recursos que possibilitam aproveitar as funcionalidades da rede de forma automatizada. Entre as funções disponibilizadas por essas APIs, existe uma que retorna uma coleção dos *tweets* mais recentes compartilhados por um determinado usuário.

Sabendo disso, foi desenvolvida uma aplicação *web* utilizando a linguagem de programação Python e o *framework* para desenvolvimento *web* Django¹ na qual o usuário

¹www.djangoproject.com

Tabela 1. Seis subfunções valorativas

Subfunções Valorativas	Motivadores e Orientações	Valores Básicos
Experimentação	Motivador humanitário e orientação pessoal	EMOÇÃO: Desfrutar desafiando o perigo; buscar aventuras. PRAZER: Desfrutar a vida; satisfazer todos os seus desejos. SEXUALIDADE: Ter relações sexuais; obter prazer sexual.
Realização	Motivador materialístico e orientação pessoal	PODER: Ter poder para influenciar os outros e controlar decisões; ser o chefe de uma equipe. PRESTÍGIO: Saber que muita gente o conhece e admira; quando velho, receber uma homenagem por suas contribuições. ÊXITO: Obter o que se propõe; ser eficiente em tudo que faz.
Existência	Motivador materialístico e orientação central	SAÚDE: Preocupar-se com sua saúde antes mesmo de ficar doente; não estar enfermo. ESTABILIDADE PESSOAL: Ter certeza de que amanhã terá tudo o que tem hoje; ter uma vida organizada e planejada. SOBREVIVÊNCIA: Ter água e comida, e poder dormir bem todos os dias; viver em um lugar com abundância de alimentos.
Suprapessoal	Motivador humanitário e orientação central	BELEZA: Ser capaz de apreciar o melhor da arte, música e literatura; ir a museus ou exposições onde possa ver coisas belas. CONHECIMENTO: Procurar notícias atualizadas sobre assuntos pouco conhecidos; tentar descobrir coisas novas sobre o mundo. MATURIDADE: Sentir que conseguiu alcançar seus objetivos na vida; desenvolver todas as suas capacidades.
Interativa	Motivador humanitário e orientação social	AFETIVIDADE: Ter uma relação de afeto profunda e duradoura; ter alguém para compartilhar seus êxitos e fracassos. CONVIVÊNCIA: Conviver diariamente com os vizinhos; fazer parte de algum grupo, como: social, religioso, esportivo, entre outros. APOIO SOCIAL: Obter ajuda quando necessitar; sentir que não está só no mundo.
Normativa	Motivador materialístico e orientação social	OBEDIÊNCIA: Cumprir seus deveres e obrigações do dia a dia; respeitar seus pais, os superiores e os mais velhos. RELIGIOSIDADE: Crer em Deus como o salvador da humanidade; cumprir a vontade de Deus. TRADIÇÃO: Seguir as normas sociais de seu país; respeitar as tradições de sua sociedade.

Tabela 2. Quantidade de usuários e quantidade de *tweets* após cada filtro

	Nº usuários	Nº Tweets
Realizaram login no sistema	2.496	2.277.832
Responderam o questionário completo	2.048	1.937.953
Tinham pelos menos 1 <i>tweet</i>	1.975	1.937.953

pode realizar *login* através de sua conta do Twitter para autorizar a captura dos seus 3.200 *tweets* mais recentes, sendo esse número a quantidade máxima estabelecida pela API para cada usuário. Após a autorização, o usuário é redirecionado para o questionário com as 18 perguntas sobre valores humanos e de forma assíncrona é realizada a captura dos *tweets*. Ao finalizar o questionário a aplicação persiste as respostas e os *tweets* coletados em um banco de dados. Dessa forma, para cada usuário que finalizou o questionário, tem-se diversos dados textuais representados pelos *tweets* capturados e uma classe associada representada pela subfunção valorativa determinada pelo questionário. Este formato permite o treinamento e uso de classificadores de aprendizagem de máquina. Na Figura 2 podemos ver a página inicial da aplicação e a tela da primeira pergunta do questionário.

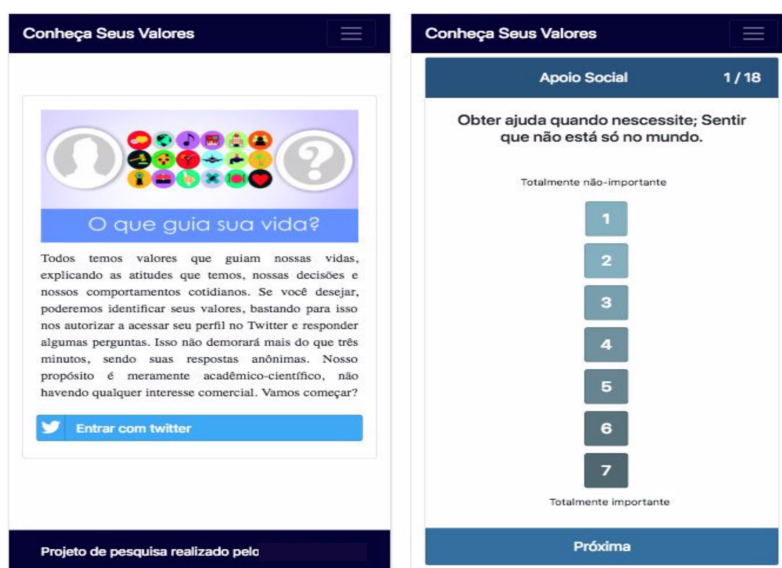


Figura 2. Página inicial e primeira pergunta do questionário

Como podemos observar na Tabela 2, foram coletados dados de 2.496 usuários, entretanto foi necessário excluir os usuários que não responderam o questionário completamente, uma vez que só é possível identificar seus valores humanos com o questionário totalmente respondido, e também os usuários que não possuíam *tweets* em sua timeline. Assim, no conjunto de dados tem-se 1.937.953 *tweets* de 1.975 usuários.

3.2. Pré-processamento dos dados

A etapa de pré-processamento dos dados tem como objetivo preparar os dados para serem utilizados no treinamento e avaliação dos classificadores. Primeiramente foram removidos dos *tweets* URLs, nomes de usuário, que no Twitter começam com o caracter @, hashtags, que são palavras iniciadas com o caracter #, dígitos, pontuações, quebras de

Tabela 3. Exemplos de tweets antes e depois do pré-processamento

Antes	Depois
Um espetinho de coração nunca pedi nada	espet coraca nunc ped nad
achei interessante. eh pro meu tcc https://t.co/ShuUOBbxVs	ach interest eh pro tcc
O único que tem estilo la na sala é o Zidane	unic estil la sal zidan

Tabela 4. Quantidade de usuários e tweets por classe

Classe	Nº Usuários	Nº Tweets
Existência	481	289.375
Experimentação	148	82.908
Interativa	293	186.814
Normativa	171	94.432
Realização	281	109.213
Suprapessoal	169	352.788
Total	1.885	1.115.530

linha e espaços em branco repetidos. Em seguida converteu-se todo o texto para letras minúsculas e os caracteres acentuados foram trocadas pelos seus correspondentes sem acento. *Retweets* foram removidos, pois são mensagens de outros usuários, além dos *tweets* em inglês que divergem da maioria dos *tweets* em português coletados. Por fim, foram retiradas *stopwords* em português e aplicou-se *stemming* através da biblioteca NLTK [Bird et al. 2009]. Após o pré-processamento, alguns *tweets* ficaram vazios, por isso eles também foram removidos. A Tabela 3 apresenta três exemplos de *tweets* antes e depois do pré-processamento.

A Tabela 4 traz a quantidade de usuários e *tweets* por cada uma das seis classes depois do pré-processamento. Note que existe um desbalanceamento em relação a quantidade de *tweets* por classe. Por exemplo, nos dados coletados, tem-se 82.908 *tweets* de usuários que obtiveram a classe experimentação como resultado do questionário, por outro lado, tem-se 352.788 *tweets* de usuários que obtiveram a classe suprapessoal. O total de usuários e *tweets* após o pré-processamento é menor que o inicialmente coletado, pois vários *tweets* foram removidos no processo.

3.3. Treinamento

Com o desbalanceamento dos dados identificado anteriormente, os algoritmos de aprendizagem de máquina podem se tornar tendenciosos para um determinado resultado, prejudicando assim o desempenho final [Mani and Zhang 2003]. Sabendo disso, os classificadores foram testados usando duas abordagens diferentes, uma com os dados completos e outra utilizando a técnica *undersampling* de balanceamento. Esta técnica consiste na identificação da menor classe e a exclusão de uma parcela dos dados das classes majoritárias, com a intenção de realizar o balanceamento tomando como referência a menor classe [Japkowicz and Stephen 2002]. Assim, como a menor classe possui 82.908 *tweets*, também foram utilizados para as outras classes 82.908 exemplos, estes escolhidos aleatoriamente do conjunto total de *tweets* para cada classe.

Para treinamento, cada *tweet* foi usado como um exemplo dos dados, sendo sua classe o resultado do questionário dos valores humanos respondido pelo usuário que enviou o *tweet*. Por exemplo, se foram capturados 1.000 *tweets* de um usuário que obteve a classe suprapessoal como resposta do questionário, cada um dos seus *tweets* foi usado como dado de treinamento com a classificação suprapessoal.

Entretanto, sabendo que cada *tweet* tem no máximo 280 caracteres e que após passar pelo pré-processamento ele apresenta uma perda significativa de seu tamanho, podemos ter poucos dados em cada exemplo, o que pode prejudicar o desempenho dos classificadores. Por isso, além de usar os *tweets* individualmente, utilizou-se a estratégia de agrupar *tweets* de uma mesma classe de resultado, com o objetivo de investigar como esses agrupamentos podem alterar o resultado. Foram utilizados os seguintes tamanhos para os grupos: 1, 10, 20, 40, 60, 80, 100, 120, 140, 160 e 180 *tweets*. Assim, cada exemplo do treinamento tem o conteúdo de um número de *tweets* de acordo com o tamanho do grupo e a classificação de acordo com o resultado do questionário.

Para cada grupo foram treinados e avaliados três classificadores de aprendizagem de máquina supervisionada: Regressão Logística, Florestas Aleatórias e SVC Linear, todos implementados pela biblioteca Scikit-Learn [Pedregosa et al. 2011]. Para avaliar os classificadores foi utilizada a validação cruzada com 5 *folds*. Por fim, a média dos resultados dos *folds* foi calculada para cada classificador com a finalidade de medir o desempenho dos algoritmos.

4. Resultados

Na Tabela 5 são apresentados os resultados da validação cruzada para cada classificador de acordo com os diferentes agrupamentos de *tweets*. Além disso, tem-se os diferentes resultados obtidos para as classificações com *undersampling* e sem *undersampling* (normal). Na primeira coluna observa-se a quantidade de *tweets* combinados em cada uma das execuções. Ou seja, o valor 1 representa que um *tweet* foi usado para cada exemplo, já o valor 10 significa que 10 *tweets* foram agrupados para formar um exemplo, e assim sucessivamente. A partir da segunda coluna até o fim tem-se a média dos resultados da validação cruzada para cada um dos classificadores usando ou não *undersampling*.

A Figura 3 traz dois gráficos com os resultados de cada classificador em relação ao número de *tweets* agrupados. O gráfico da esquerda expõe os resultados para o treinamento sem *undersampling* e o da direita os resultados com *undersampling*.

5. Discussão

Como podemos observar na Tabela 5, obtivemos os valores das acurácias para três classificadores distintos e para 11 agrupamentos diferentes de *tweets*. Os testes foram realizados com o objetivo de verificar a viabilidade de classificar os valores humanos usando as mensagens do Twitter e quais classificadores têm um bom desempenho nessa tarefa. Além disso, também foi avaliada a influência do tamanho dos exemplos nos resultados.

Analisando a Figura 3, nota-se que o SVC Linear e a Regressão Logística obtiveram os melhores resultados entre os classificadores testados. O SVC Linear alcançou uma acurácia máxima de 0,6008 quando foram usados 140 *tweets* por exemplo e não foi utilizado balanceamento. Já a Regressão Logística obteve a acurácia máxima de 0,5214

Tabela 5. Acurácia dos classificadores

Tweets por exemplo	Regressão Logística		Florestas Aleatórias		SVC Linear	
	Normal	Undersampling	Normal	Undersampling	Normal	Undersampling
1	0.3543	0.2631	0.3185	0.2304	0.3560	0.2644
10	0.4391	0.3780	0.3229	0.2285	0.4597	0.3954
20	0.4723	0.4206	0.3221	0.2387	0.5088	0.4348
40	0.5027	0.4638	0.3291	0.2376	0.5470	0.4805
60	0.5134	0.4731	0.3244	0.2401	0.5716	0.5113
80	0.5163	0.4942	0.3287	0.2418	0.5848	0.5341
100	0.5215	0.4945	0.3284	0.2430	0.5944	0.5358
120	0.5115	0.4921	0.3310	0.2401	0.5956	0.5334
140	0.5090	0.4912	0.3209	0.2315	0.6008	0.5269
160	0.5043	0.4835	0.3222	0.2309	0.5986	0.5325
180	0.4948	0.4773	0.3161	0.2322	0.5961	0.5341

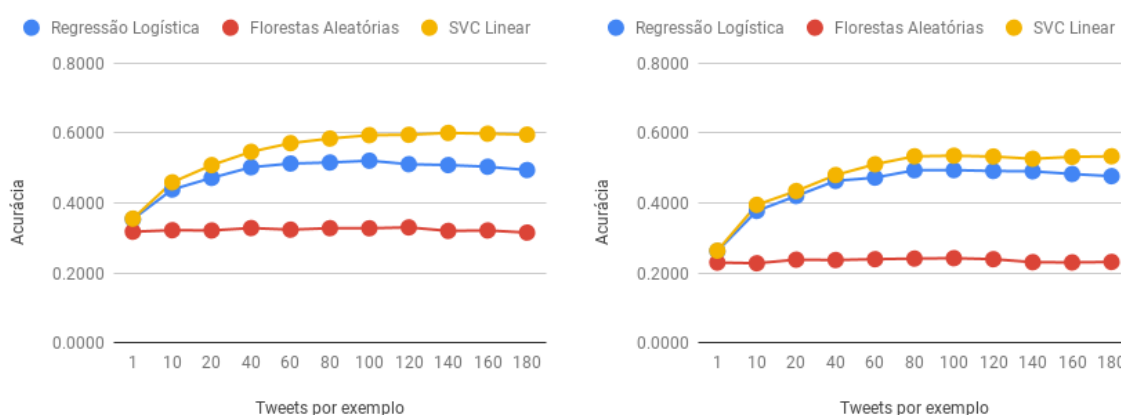


Figura 3. Acurácia dos classificadores em relação aos agrupamentos dos tweets sem undersampling (esquerda) e com undersampling (direita)

usando 100 tweets por agrupamento também sem balanceamento. Para esses algoritmos, ampliar o número de tweets por agrupamento tende a fazer a acurácia das classificações aumentar. O crescimento da acurácia foi acentuado com o aumento do uso de 1 tweet por exemplo para 10 tweets e em seguida para 20 tweets. A acurácia continuou crescendo para os agrupamentos de 40, 60 e 80 tweets, mas a partir de 100 os resultados começam a ter variações de menos de 1%, sendo que em alguns casos a acurácia diminuiu.

A Figura 4 apresenta duas matrizes de confusão para o algoritmo SVC Linear usando grupos de 140 tweets, que foi o melhor resultado alcançado nos testes. A matriz da direita traz os resultados sem balanceamento e a da esquerda os resultados com balanceamento. Estes resultados foram obtidos utilizando 80% do conjunto de dados para treinamento e 20% para teste com os exemplos escolhidos aleatoriamente para que a ordem de captura não influenciasse. Na matriz de confusão sem balanceamento, como era de se esperar, as classes majoritárias recebem a maioria das predições, mas isso não resulta em um desempenho ruim para acurácia. Na matriz de confusão com balanceamento,

não temos mais a predominância das classes majoritárias, mas em todos os testes, essa abordagem teve resultados piores que os testes sem balanceamento.

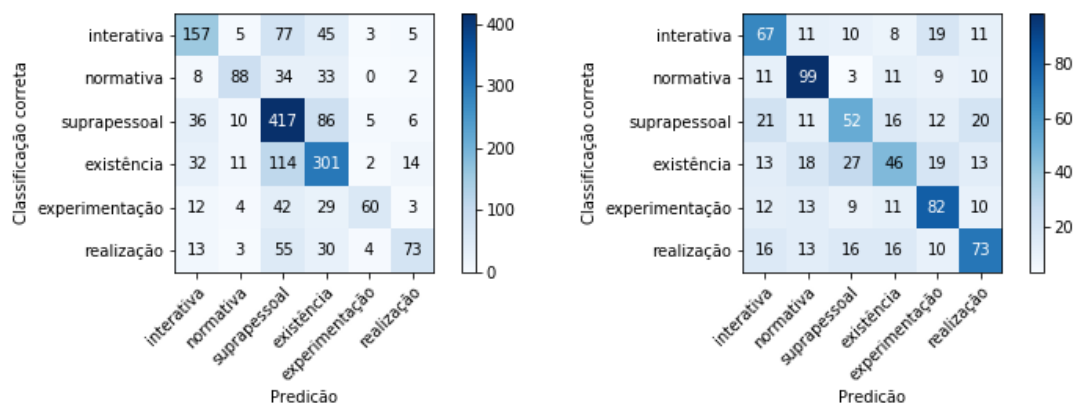


Figura 4. Matriz de confusão para o algoritmo SVC Linear utilizando 140 tweets por exemplo sem undersampling (esquerda) e com undersampling (direita)

Para o algoritmo Florestas Aleatórias, as acurácias foram menores em relação ao SVC Linear e a Regressão Logística. Nele, aumentar o número de tweets agrupados teve pouco efeito nos resultados, assim as acurácias nos diferentes testes tiveram poucas variações. Com undersampling, a acurácia variou entre 0,2285 e 0,2430, sem undersampling, a acurácia variou entre 0,3161 e 0,3310.

Com esses dados, temos a indicação de que usar exemplos maiores, com mais tweets agrupados, tende a melhorar o desempenho para os classificadores, principalmente para a Regressão Logística e para o SVC Linear. Para eles, os resultados mostram que a partir de 100 tweets agrupados a taxa de acerto se aproxima de um plateau. Além disso, é importante levar em consideração que a quantidade de tweets agrupados é inversamente proporcional a quantidade de exemplos no conjunto de dados. Assim, o número de exemplos para treinamento e para testes diminui, podendo causar problemas de underfitting.

Poderíamos ter resultados diferentes com o agrupamento de mais tweets, mas, devido ao desbalanceamento, o número de exemplos por classes ficaria muito reduzido. Com grupos de 180 tweets, cada classe tinha 552 exemplos e como foi usada validação cruzada com 5 folds, 80% dos exemplos (aproximadamente 442) eram usados para treinamento. Observamos também, que a porcentagem de acertos na Regressão Logística e no SVC Linear foram as melhores e se mantiveram bem próximas em alguns casos. Em uma possível escolha entre eles, poderíamos utilizar para o desempate o tempo de execução, no qual o algoritmo SVC Linear tem um tempo de treinamento em uma máquina de configuração (sistema operacional Mac OS X, processador i5 e memória de 8gb) cerca de três vezes menor que a Regressão Logística.

Como podemos observar nos resultados, o classificador conseguiu encontrar relações entre o que os usuários escrevem no Twitter e os seus valores humanos retornados pelo questionário. Em diversos testes realizados, conseguiu-se taxas de mais de 50% de acerto para as 6 classes. Para se ter ideia da qualidade desse resultado, uma escolha aleatória, na qual nenhum padrão fosse aprendido pelos algoritmos, teria a probabilidade de 16,666% de acerto. Portanto, taxas de acerto de mais de 50% representam

um ganho significativo em relação a uma escolha aleatória.

Com isso, tem-se um forte indício que podemos usar mensagens textuais compartilhadas no Twitter por pessoas para inferir os valores humanos sem a necessidade da resposta do questionário.

Vale salientar que para o conjunto de dados desbalanceados, a classe com a maior quantidade de *tweets* corresponde a 31,6% dos dados. Consequentemente, uma classificação pela classe majoritária atingiria essa taxa de acerto. Assim, tomando este dado como linha de base ao invés da escolha aleatória, temos que o ganho na acurácia para os dados desbalanceados foi menor que para os dados balanceados.

6. Trabalhos Relacionados

As redes sociais e em especial o Twitter têm sido usadas em diversos trabalhos ao longo dos últimos anos. Em seu artigo, [Qiu et al. 2012] realizou um estudo para medir os Cinco Grandes traços de personalidade baseados em análise linguística: abertura para experiência, conscienciosidade, extroversão, neuroticismo e agradabilidade [Norman 1963]. Para descobrir se existe relação entre *tweets* compartilhados e a personalidade das pessoas, 142 participantes foram recrutados para responder o questionário de identificação dos Cinco Grandes traços de personalidade [John et al. 1991] e em seguida seus *tweets* foram coletados. Para analisar os *tweets*, oito avaliadores receberam as mensagens coletadas e atribuíram a impressão que eles tiveram dos 142 participantes de acordo com o Cinco Grandes traços de personalidade. Como resultado, observou-se a maior correlação entre a avaliação e o resultado do questionário para os traços agradabilidade e neuroticismo.

Para automatizar a avaliação dos Cinco Grandes traços de personalidade, [Golbeck et al. 2011] propôs o uso de aprendizagem de máquina para prever os resultados do questionário. Esta abordagem é semelhante a do presente artigo, tendo como principal diferença que em nosso artigo foi trabalhado valores humanos ao invés dos Cinco Grandes traços de personalidade. Foram coletadas respostas do questionário de identificação dos Cinco Grandes traços de personalidade e 2.000 *tweets* de cada um dos 279 participantes dessa pesquisa. Esses dados foram usados para treinar os algoritmos Gaussian Process e ZeroR ambos implementados no software WEKA [Hall et al. 2009]. Como resultado, a classe abertura para experiência obteve a menor taxa de erro, 0,1192, e neuroticismo a maior, 0,1821.

O trabalho de [Sumner et al. 2012] vai além dos Cinco Grandes traços de personalidade e explora traços antissociais: narcisismo, maquiavelismo e psicopatia, comumente chamados da Tríade Sombria [Paulhus and Williams 2002]. 2.927 usuários do Twitter participaram desse estudo, eles responderam um questionário para identificação de traços da Tríade Sombria e em seguida tiveram os seus *tweets* coletados. Assim como o trabalho anterior, o WEKA foi utilizado para classificação, entretanto diferentes algoritmos foram avaliados: SVM, Florestas Aleatórias, Árvore de Decisão e Naive Bayes. Além disso, também foi usado o Kaggle² para que outros pesquisadores e desenvolvedores pudessem usar os dados coletados para criar modelos e fazer previsões. O artigo conclui que as técnicas de aprendizagem de máquina foram incapazes de prever traços da Tríade Sombria individualmente. Por outro lado, para grupos maiores os modelos podem ser úteis para prever tendências de aumento ou diminuição de traços antissociais na sociedade.

²www.kaggle.com

7. Conclusão e trabalhos futuros

Este trabalho apresentou a criação e avaliação de modelos para classificação de valores humanos, com a finalidade de analisar a viabilidade de usar aprendizagem de máquina para investigar os valores humanos das pessoas de acordo com mensagens textuais compartilhadas no Twitter. Para isso, foram coletados 2.277.832 *tweets* enviados por 2.496 usuários, que posteriormente foram pré-processados. Dada a quantidade diferente de *tweets* por cada classe também foram feitos testes usando a técnica de *undersampling* para balanceamento, assim reduzindo a quantidade de *tweets* por classe de acordo com a classe minoritária. Como um *tweet* pode ter pouco texto, foi testada a estratégia de agrupar *tweets* para formar os exemplos dos dados de treinamento. Por fim, para cada agrupamento, foram treinados e avaliados três algoritmos de aprendizagem de máquina supervisionada: Regressão Logística, Florestas Aleatórias e SVC Linear.

Com todos os algoritmos, foram obtidos resultados muito superiores se compararmos com uma escolha aleatória. Em especial, o algoritmo SVC Linear obteve a maior acurácia, chegando a 0,6008 usando agrupamentos de 140 *tweets* por cada exemplo. Com esse resultado, verificamos um forte indício da viabilidade de uso de modelos textuais criados por algoritmos de aprendizagem de máquina como instrumento para identificar os valores humanos das pessoas sem a necessidade da aplicação de um questionário. Durante a coleta dos dados para este trabalho, 448 pessoas (17,9% do total) começaram a responder o questionário, mas não terminaram. Ou seja, em um questionário que pode ser respondido em aproximadamente três ou quatro minutos, tivemos essa porcentagem considerável de abandono, problema que não aconteceria com o uso exclusivo do classificador.

Como trabalhos futuros, pretende-se coletar mais dados para aumentar os exemplos disponíveis para treinamento e teste para tentar melhorar o desempenho dos classificadores. Além disso, para aumentar a acurácia, planeja-se a execução de outros algoritmos de aprendizagem de máquina, por exemplo, redes neurais recorrentes que atualmente são o estado da arte para criação de modelos de linguagem. Também pretende-se replicar a mesma ideia deste trabalho para a rede social Facebook, tendo em vista um número ainda maior de usuários mensalmente ativos, aproximadamente 2,32 bilhões de usuários [Facebook 2018].

Referências

- Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc."
- Ceron, A., Curini, L., Iacus, S. M., and Porro, G. (2014). Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to italy and france. *New Media & Society*, 16(2):340–358.
- Coppersmith, G., Dredze, M., and Harman, C. (2014). Quantifying mental health signals in twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 51–60.
- Facebook (2018). Facebook reports fourth quarter and full year 2018 results. Acessado: 25 de Março de 2019.

- Golbeck, J., Robles, C., Edmondson, M., and Turner, K. (2011). Predicting personality from twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 149–156. IEEE.
- Gouveia, V. V. (2013). Teoria funcionalista dos valores humanos: fundamentos, aplicações e perspectivas [functional theory of human values: Fundamentals, applications, and perspectives]. *São Paulo: Casa do Psicólogo*.
- Gouveia, V. V., Milfont, T. L., Fischer, R., and de Miranda Coelho, J. A. P. (2009). Teoria funcionalista dos valores humanos: Aplicações para organizações. *Revista de Administração Mackenzie (Mackenzie Management Review)*, 10(3).
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- Japkowicz, N. and Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449.
- John, O. P., Donahue, E. M., and Kentle, R. L. (1991). The big five inventory—versions 4a and 54.
- Mani, I. and Zhang, I. (2003). knn approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of workshop on learning from imbalanced datasets*, volume 126.
- Medeiros, E. D. d. et al. (2011). Teoria funcionalista dos valores humanos: Testando sua adequação intra e interculturalmente.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, 66(6):574.
- Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326.
- Paulhus, D. L. and Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of research in personality*, 36(6):556–563.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Qiu, L., Lin, H., Ramsay, J., and Yang, F. (2012). You are what you tweet: Personality expression and perception on twitter. *Journal of Research in Personality*, 46(6):710–718.
- Sumner, C., Byers, A., Boochever, R., and Park, G. J. (2012). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *2012 11th International Conference on Machine Learning and Applications*, volume 2, pages 386–393. IEEE.
- Twitter (2018). Quarterly results — about. <https://investor.twitterinc.com/results.cfm>.