FUSÃO DE REDES CONVOLUCIONAIS HETEROGÊNEAS PARA DETECÇÃO DE NOVIDADES EM REDES SOCIAIS

MARTA AMORIM*, PATRICK MARQUES CIARELLI*

*UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
VITÓRIA, ES, 29075-910, BRASIL
E-MAILS: MARTA.AMORIM@IFES.EDU.BR, PATRICK.CIARELLI@UFES.BR

Abstract— Novelty detection is a classification task that attempt to identify data that differ in some aspects from the data available during training. It is assumed that the typical class is very well sampled, while the other classes are poorly sampled or have no samples. The challenge of automating novelty detection gets even higher when typical class data come in the form of continuous data stream. In that way, the amount of data is not only tremendous but also must be processed in a short time. When the data stream comes from social networks, there is still another aggravating factor: the poor quality of the data produced. To handle this problem, in this work is proposed the fusion of the results of two heterogeneous convolutional networks, one designed for image and other for text classification. In addition, a new labeled dataset is presented and available with images and texts from Facebook for the task of novelty detection. In experimental results obtained in this dataset, an average precision of 80.18% and an average recall of 88.00% were achieved, indicating that the presented method is promising.

Keywords— Novelty Detection, Convolutional Neural Network, Machine Learning, Social Network and Fusion Methods.

Resumo— Detecção de novidades é uma tarefa de classificação que tenta identificar dados que diferem em alguns aspectos em relação aos dados disponíveis durante o treinamento. É assumida que a classe normal é muito bem amostrada, enquanto outras classes são pouco amostradas ou não tem amostras. O desafio de automatizar a detecção de novidades fica ainda maior quando os dados da classe normal são provenientes na forma de um fluxo contínuo de dados. Dessa forma, a quantidade de dados não é apenas enorme, mas devem também ser processados em pouco tempo. Quando se trata de dados de fluxo contínuo das redes sociais, ainda se tem outro agravante: a baixa qualidade dos dados produzidos. Para tratar esse problema, neste trabalho é proposta a fusão dos resultados de duas redes convolucionais heterogêneas, uma projetada para classificação de imagens e outra para textos. Além disso, é apresentada e disponibilizada uma nova base de dados rotulada com imagens e textos provenientes da rede social Facebook para a tarefa de detecção de novidades. Em resultados experimentais foram alcançados uma precisão média de 80,18% e um recall médio de 88,00%, o que indica que o método apresentado é promissor.

Palavras-chave — Detecção de Novidades, Redes Neurais Convolucionais, Aprendizado de Máquina, Rede Social e Métodos de Fusão.

1 Introdução

Detecção de novidades pode ser definida como uma tarefa de reconhecimento dos dados de teste que se diferenciam em relação aos dados que estão disponíveis durante o treinamento. Métodos dessa natureza são normalmente aplicados em um grande número de exemplos normais (também conhecidos como exemplos positivos) e onde existem dados insuficientes para descrever as anormalidades (conhecidos como exemplos negativos) (Barnett et al., 1994). A natureza dos rótulos anormais ou não-observados é a razão pela qual métodos de detecção de novidades são referidos como não-supervisionados ou o*ne-class* enquanto métodos de classificação são referidos como supervisionados. Nos casos onde rótulos anômalos são observados, o problema se sintetiza para uma versão não-balanceada de detecção de novidades (Aggarwal, 2016).

Termos sinônimos intercambiáveis de detecção de novidades são frequentemente usados na literatura, tais como: detecção de anomalias e detecção de *outliers*. As diferentes terminologias são originárias de diferentes domínios de aplicações em que a classificação *one-class* pode ser aplicada, e não existe uma definição universalmente aceita. Merriam-Webster (Merriam-Webster, 2018) define "novidade" como algo "novo ou não parecido com algo anteriormente conhecido ou usado". Dessa forma, esse dado novo pode ser: normal, anormal ou ruidoso. Barnett e Lewis (Barnett et al., 1994) definem *outlier* como dados que são inconsistentes em relação ao conjunto de dados de treinamento. Portanto, detecção de *outlier* visa manipular estritamente observações anormais do conjunto de dados. Mas ambas definições, tanto detecção de novidades quanto detecção de anomalias, compartilham da mesma característica em relação a insuficiência dos dados anormais ou novos (Pimentel et al., 2014).

A escassez de exemplos negativos pode ser devido ao alto custo em obtê-los ou a baixa frequência que os eventos anormais ocorrem. Por exemplo, uma máquina pode operar por um longo período em operação normal, e sempre que um comportamento anormal for detectado, um alarme é disparado para evitar qualquer avaria. Os dados da máquina durante sua operação normal usualmente são acessíveis e de fácil medição. Por outro lado, a obtenção das falhas de uma máquina requer a operação inadequada em todas as possíveis formas o que pode provocar a sua avaria. Portanto, é difícil, senão impossível, obter uma classe negativa bem amostrada. O problema pode se tornar mais difícil se existe uma significante variabilidade

entre os modos anormais, alguns dos quais podem não ser conhecidos a priori, o que torna esquemas de classificação multiclasse inadequados para essas aplicações. A solução para este problema é oferecida pela detecção de novidades, em que a descrição da normalidade é aprendida pela construção de um modelo com numerosos exemplos representando instâncias positivas (Pimentel et al., 2014). A detecção automática de novidades pode ser empregada para tornar vários processos mais confiáveis, tais como: na identificação de um padrão de tráfico anômalo em uma rede de computadores, que pode significar um computador invadido por hackers; na detecção de uma imagem anômala de ressonância magnética, que pode indicar a presença de tumores malignos; na descoberta de anormalidades em dados de transação de cartão de crédito, que pode indicar furto do cartão; na detecção de leituras anômalas de um sensor, que pode significar alguma falha em algum componente mecânico; na revelação de fluxo temporal anormal de mensagens nas redes sociais, que pode caracterizar algum evento importante ou inesperado, etc. (Chandola et al., 2009).

Além da falta dos exemplos negativos, outra questão que tem emergido na tarefa de detecção de novidades é a heterogeneidade dos dados, sendo uma característica inerente das redes sociais. Os dados produzidos nas redes sociais podem conter não apenas textos, mas também imagens, vídeos, informações temporais, conexões sociais, preferências dos usuários e outros metadados. Entretanto, a maioria das detecções de novidades na *Web* tratadas na literatura consideram apenas o conteúdo textual e as conexões sociais, enquanto que o conteúdo visual é ignorado. Gao et al. (2015), em recente investigação, mostraram que 30% dos blogs postam conteúdos com imagens e o número continua aumentando. Portanto, o conteúdo visual torna-se importante na análise de detecção de novidades (Gao et al., 2015). A proliferação de aplicações que compartilham fotos na *web*, tais como Flickr, Instagram, Panoramio e Google Picasa, têm grandes quantidades de fotos disponíveis para acesso público. Hoje, o Flickr contém mais de 5 bilhões de fotos, das quais 100 milhões são geolocalizadas[1] (Ruocco et al., 2014). Atualmente, o Facebook é a rede social mais popular em compartilhamento de fotos, com mais de 100 milhões de fotos compartilhadas diariamente (LATimes, 2016). Outra peculiaridade importante das redes sociais é que os dados são produzidos na forma de *data stream (*fluxo contínuo de dados). A quantidade de dados nas redes sociais não é apenas enorme, mas devem também ser processados em pouco tempo, pois a informação cresce a cada minuto. Isso torna-se um desafio na tarefa detecção de novidades (Gao et al., 2015).

Tendo em vista todos esses desafios, o objetivo deste trabalho é identificar novidades em dados de redes sociais combinando a classificação de textos e imagens. Para esse propósito foi criada uma base de dados rotulada com dados provenientes do Facebook. Além disso, pretende-se explorar o desempenho da fusão de duas redes convolucionais treinadas com dados heterogêneos (imagens e textos) para identificação de novidades, o que não é muito comum na literatura. Normalmente, encontra-se em abundância arquiteturas próprias para dados balanceados e homogêneos.

Este artigo está organizado nas seguintes seções. Na Seção 2 são apresentados os trabalhos da literatura relacionados a detecção de novidades usando redes convolucionais. O método proposto está descrito na Seção 3. Na Seção 4 são detalhados os parâmetros da arquitetura das redes convolucionais. As informações sobre a base de dados estão incluídas na Seção 5. Na Seção 6 são discutidos os experimentos e resultados. Por final, na Seção 7 são relatados as conclusões e trabalhos futuros.

2 Trabalhos Relacionados

Existe uma vasta quantidade de distintos métodos empregados para detecção de novidades e anomalias. Por isto, aqui serão mencionados somente aqueles que utilizam redes neurais convolucionais.

O trabalho de detecção de anomalias de Zhang et al. (2017) trata de uma rede neural convolucional (CNN - *Convolutional Neural Network*) para identificar defeitos em imagens de peças (imagem de componentes eletrônicos). A proposta é criar uma rede CNN o*ne-class* baseada na AlexNet (Krizhevsky et al., 2012), sendo que a saída possui 256 neurônios, levando a imagem original a ser mapeada em um espaço de características de 256 dimensões. A estratégia adotada pelo autor é usar CNN para mapear exemplos não-defeituosos dentro de uma hiperesfera de alta-dimensão, enquanto que exemplos defeituosos ficariam fora da hiperesfera. Nos experimentos foram utilizadas 400 imagens não defeituosas e 300 imagens defeituosas para treinar a rede. A métrica empregada foi a área sob a curva (*Area Under Curve* - AUC) ROC (taxa de falsos positivos em relação a taxa de verdadeiros positivos) e obteve o resultado de 98% de assertividade na detecção das anomalias.

Ding et al. (2015) definem exclusivamente que as novidades detectadas são as intenções dos usuários das redes sociais em consumir ou adquirir algo. Para isso foi construído um modelo denominado *Consumption Intention Mining Model* (CIMM) baseado em uma rede neural convolucional de texto. Para demonstrar a efetividade do modelo, foram conduzidos experimentos em dois domínios: crianças e filmes. O *corpus* textual contém 20 milhões de *posts*, 76 milhões de sentenças e 1,3 bilhões de palavras. Os experimentos obtiveram 94,52% de acurácia para o domínio de crianças e 85,28% para o domínio de filmes.

[1]http://blog.flickr.net/en/2009/02/05/100000000-geotagged-photos-plus/

Em (Zhou et al., 2016) foi proposto um método para localizar atividades anômalas em sequências de vídeos contendo aglomerados de pessoas. O método utiliza uma rede neural convolucional temporal. Essa arquitetura permite capturar ambas dimensões, temporal e espacial, pela realização de uma convolução espaço-temporal, desse modo, informações de movimento e visuais são codificadas. A arquitetura da rede proposta pelos autores é formada por 8 camadas, sendo 4 camadas convolucionais, 2 camadas de subamostragem e 2 camadas totalmente conectadas. Sendo que das 4 camadas convolucionais, 2 realizam convolução 3-D e 2 realizam convolução 2-D. Os experimentos foram executados em três bases de dados: UMN, Subway, and U-turn. A métrica de área sob a curva ROC foi empregada e obteve os seguintes resultados: UMN atingiu 99,63%, Subway 92,7% e U-turn 95,2%.

Ribeiro et al. (2017) apresentaram a partir da perspectiva detecção de novidades, uma rede convolucional autoencoder (CAE), para capturar informações espaciais 2-D da imagem. Além disso, a rede permitiu reconstruir erros de cada *frame* do vídeo. Na etapa dos experimentos, foram utilizadas as bases de dados UCSD pedestrian que atingiu 89% de AUC e Avenue com 75,4% de AUC.

3 Método Proposto

Nesta seção é descrita a metodologia proposta neste trabalho.

3.1 Contextualização da novidade

Como discutido nas seções anteriores, um dado anômalo ou novo tem como característica principal a raridade na ocorrência. Em virtude disso, no cenário de dados provenientes das redes sociais, este trabalho é focado essencialmente em dois tipos: comum e novidade.

No contexto desta pesquisa, dados comuns são definidos a partir da análise do site de notícias *Bussiness Insider* (BusinessInsider, 2013), na qual indica os assuntos mais discutidos nas redes sociais: 28% tecnologia e mídias sociais; 25% comida, bebida e viagem; 18% filme e televisão; 10% saúde e esportes; 11% música e artes e 9% moda e shopping. Ao contrário, as novidades são assuntos com pouca vinculação ou ocorrem de forma rara, por exemplo: desastres naturais, ataques hackers, guerra, etc. Percebe-se intuitivamente mais um desafio para o algoritmo ao analisar as classes, além da falta de amostras e desbalanceamento das classes, pois alguns assuntos possuem uma fronteira limítrofe não muito clara. Por exemplo, ataques hackers estão muito relacionados à classe tecnologia, assim como desastres naturais estão relacionados à classe viagem, ou seja, não é uma tarefa trivial determinar a qual classe deve pertencer um dado apenas analisando a imagem ou o texto. Essa afirmação será verificada na Seção 6 de experimentos e resultados.

3.2 Rede Convolucional de Imagem

O propósito inicial de uma rede convolucional é extrair características. A convolução é uma operação linear que preserva o relacionamento dos pixels da imagem pela aprendizagem das características a partir da janela que desliza sobre a imagem de entrada. Na terminologia CNN, essas pequenas janelas são chamadas de filtros ou *kernels*, e a matriz formada pelo deslizamento deste filtro sobre a imagem é chamada de mapa de ativação ou mapa de características. Matematicamente, dado um filtro g de tamanho $(2k_1+1)$ x $(2k_2+1)$ cuja coordenada central é $g(0,0)$, e uma imagem f, então o mapa de características é a matriz resultante h da operação de convolução apresentada na Equação 1 (Gonzalez et al., 2007) (Amorim et al., 2018).

$$h(x,y) = f(x,y) * g(x,y) = \sum_{n_1=-k_1}^{k_1} \sum_{n_2=-k_2}^{k_2} f(x-n_1, y-n_2).g(n_1, n_2) \quad (1)$$

A maior parte das redes convolucionais utilizadas no campo de pesquisa da visão computacional tratam da classificação com classes balanceadas. Tal qual então o objetivo é verificar se essas arquiteturas podem ser empregadas em detecção de novidades. Para isso, neste trabalho são usadas as arquiteturas AlexNet e a Lppreconv1, apresentada por Amorim et al. (2018). Na Seção 4 serão detalhadas as configurações das arquiteturas das redes propostas.

3.3 Rede Convolucional de Texto

Diferentemente da rede convolucional de imagem, os dados de texto não podem ser enviados em sua forma original para a rede, ou seja, o texto deve ser codificado antes. Portanto, deve-se escolher um modelo que represente o texto em termos computacionais. No campo da linguística computacional, a codificação do texto é conhecida rigorosamente como modelo de linguagem. Existem vários modelos de linguagens, tais como: modelos n-gram, modelos neurais, modelos estatísticos, etc.

Os modelos que vêm se destacando recentemente na literatura por bons resultados são os modelos neurais, os quais geram os conhecidos vetores embutidos ou vetores densos. Esses modelos caracterizam cada palavra como um vetor de pesos de números reais, e que são representações internas das redes neurais (Levy et al., 2015).

Representando matematicamente o conceito, seja um vocabulário de palavras de tamanho W, onde cada palavra é identificada pelo seu índice. Inspirada na hipótese distribucional, as representações das palavras são treinadas para predizer quais palavras aparecem em seu contexto. Mais formalmente, dado um grande corpus de treinamento representado como sequência de palavras $w_1 ... w_T$, o objetivo do modelo é

maximizar o log da probabilidade descrita na Equação 2:

$$\sum_{t=1}^{T} \sum_{c \in C_t} \log p(w_c | w_t) \quad (2)$$

onde C_t indica o conjunto dos índices de palavras do contexto da palavra w_t, e T é o total de palavras em w_t. Esse modelo é conhecido como Skip-gram, pois dada a palavra w_t, as palavras do seu contexto C_t são usadas para obter o vetor denso de w_t (Bojanowski et al., 2016). A Figura 1 ilustra a arquitetura do modelo de linguagem neural. Neste trabalho foi utilizado um algoritmo variante do Skip-gram, conhecido como GloVe[2] (*Global Vectors for Word Representation*) (Pennington et al., 2014).

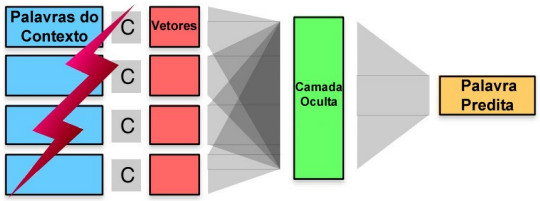

Figura 1. Arquitetura do modelo neural (Koehn, 2017)

A entrada da rede da Figura 1 são vetores de alta-dimensão conhecidos como vetores *one-hot*. Eles são do tamanho do vocabulário de palavras T e completamente preenchidos com zeros, exceto no índice da palavra que eles representam, como no caso da palavra "dog" = (0, 0, ..., 0, 1, 0, ..., 0, 0)^T. As palavras preditas na saída são aquelas que aparecem no mesmo contexto das palavras de entrada. Ela também é representada na forma *one-hot*. Ao final do treinamento da rede, os pesos de conexão entre cada neurônio da camada de saída e os da camada oculta representam um vetor de baixa dimensão que relaciona cada palavra semanticamente com as palavras do seu contexto. Estes vetores reduzidos são representações enriquecidas das palavras, e os mesmos alimentarão a CNN de texto.

Após a representação do texto de forma adequada, pode-se utilizar qualquer classificador. Levando em consideração uma das maiores competições da área de linguística computacional, a SemEval-2017[3], as redes CNN vêm apresentando os melhores resultados. Portanto, por esta razão ela foi selecionada como classificador.

A rede convolucional textual recebe uma entrada unidimensional. As camadas podem ser as mesmas da rede de imagem, tais como: convolução, maxpooling, função de ativação, etc. A arquitetura será detalhada na Seção 4.

3.4 Fusão das Redes Convolucionais

Vários classificadores são construídos com base no tipo de conjunto de dados e na distribuição. As fusões desses distintos modelos podem prover alto desempenho na classificação de várias aplicações, tais como: reconhecimento de padrões, análise de documento, mineração de dados, etc.

Seja um classificador qualquer C que realiza o mapeamento $C: X \rightarrow Y$ a partir do espaço de características X em um conjunto de classes $Y= \{l_1,l_2,...,l_j\}$. A fusão de classificadores $\{C_1,C_2,...,C_M\}$ é baseado no treinamento $T=\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$. Então, as predições dos classificadores são combinadas em um $\hat{C}$* usando uma função de combinação F descrita na Equação 3 (Preisach et al., 2008):

$$\hat{C}* = F(\hat{C}_1, \hat{C}_{2,...}, \hat{C}_M) \quad (3)$$

Várias regras combinatórias foram propostas na literatura para a função F. Para este trabalho é desejado identificar o maior número possível de novidades (ou seja, aumentar o valor de recall definido na Seção 6), portanto, a função F usada neste trabalho retorna a classe novidade se ao menos um classificador, de imagem ou texto, retornar a classe novidade, caso contrário o dado é classificado como normal. A Equação 4 mostra isso, onde S_1 é a classe de saída do classificador C_1 e S_2 é a saída de C_2, e l_1 e l_2 são as classes normal e novidade, respectivamente.

$$\hat{C}* = \begin{cases} l_2, S_1 = l_2 \vee S_2 = l_2 \\ l_1, caso\ contrário \end{cases} \quad (4)$$

O método tenta levar vantagem na força dos dois classificadores estritamente na recuperação das novidades, não importando se é um falso-positivo. Isso pode significar uma melhora na métrica *recall* de novidades em detrimento da precisão. Outra função de combinação F desenvolvida neste trabalho foi baseada na função *Border Count* (*BC*)(Ruta et al., 2000), apresentada na Equação 5. A diferença entre as Equações 4 e 5 é que a última equação reflete a magnitude do nível de concordância pela soma das probabilidades.

$$\hat{C}_1* = \begin{cases} \text{argmax}_{(i=l_1,l_2)}(P(\hat{C}_1=i)+P(\hat{C}_2=i)), S_1 = l_1 \vee \\ l_2, caso\ contrário \end{cases}$$

(5)

4 Arquitetura Proposta

Nesta seção são detalhadas as configurações das arquiteturas das redes convolucionais de texto e imagem.

[2]https://nlp.stanford.edu/projects/glove/
[3]http://alt.qcri.org/semeval2017/

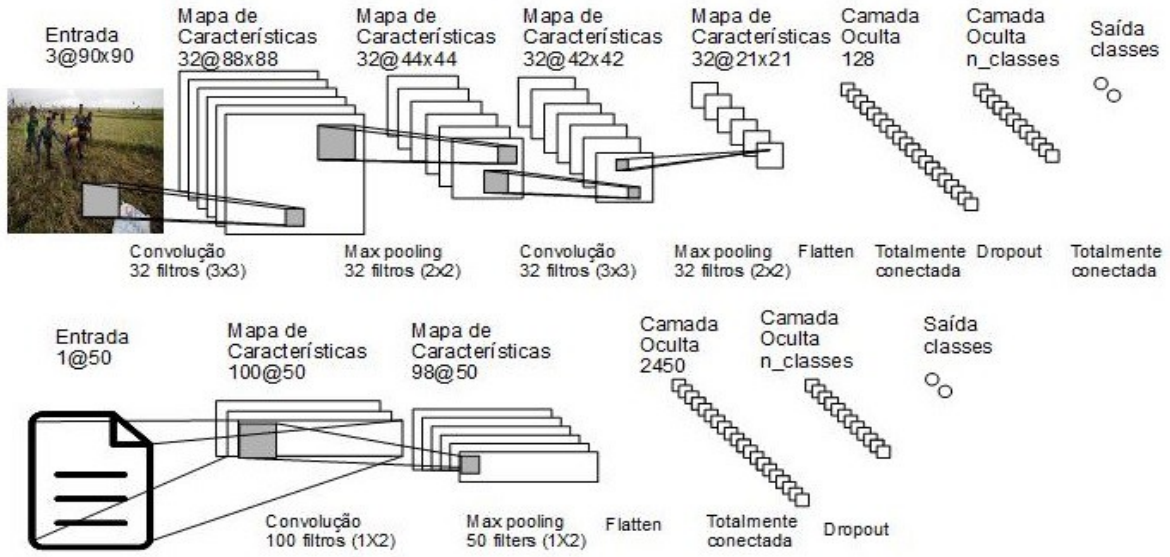

Figura 2. Na parte superior, a arquitetura da rede convolucional de imagem (Arquitetura Lppreconv1) (a) Na parte inferior, a arquitetura da rede convolucional de texto (b)

4.1 Rede Convolucional de Imagem

Foram definidas duas arquiteturas de redes convolucionais para imagens: a arquitetura Lppreconv1, proposta por Amorim et al. (2018) (ilustrada na Figura 2a), e a arquitetura AlexNet.

A Lppreconv1 foi projetada em (Amorim et al., 2018) com 6 camadas, sendo que a primeira camada realiza uma operação log-polar antes da convolução. A proposta da arquitetura é explorar características de invariância a rotação por meio da transformação log-polar, que é inspirado no sistema visual humano.

A rede Lppreconv1 foi customizada neste trabalho com os seguintes parâmetros: duas camadas de convolução 2-D, com 32 filtros de tamanho 3x3, com função de ativação ReLU, maxpooling com janela de 2x2, seguido por *dropout* de 0,25, uma camada totalmente conectada com 128 neurônios, função de ativação ReLU, e um *dropout* de 0,5, e por final uma camada totalmente conectada de 2 neurônios com função de ativação softmax. Foi utilizado o otimizador Adam (Kingma et al., 2015). A imagem de entrada da rede Lppreconv1 possui uma dimensão de 90x90 e a saída da rede convolucional possui duas classes categóricas: normal e novidade. No treinamento foram utilizadas 10 épocas com mini *batches* de 32. A rede está ilustrada na Figura 2a.

Para a CNN AlexNet, foram utilizadas imagens de entrada de 224x224, 20 épocas de treinamento e mini *batches* de 32.

4.2 Rede Convolucional de Texto

A rede convolucional de texto apresenta os seguintes parâmetros: duas camadas de convolução 1-D, a primeira convolução com 100 filtros e a segunda com 50 filtros, todos com dimensão 1x2, com função de ativação ReLU, maxpooling com janela de 1x2, seguido por um *dropout* de 0,4, por último uma camada totalmente conectada com função de ativação softmax. Para o otimizador foi utilizado gradiente descendente estocástico com taxa de aprendizagem 0,01.

A CNN possui como entrada vetores embutidos com dimensão de 1x50 e tem como saída duas classes categóricas: normal e novidade. Para o treinamento da CNN foram utilizadas 5 épocas com mini *batches* de 32. A rede está ilustrada na Figura 2b. A rede neural que gerou os vetores embutidos a partir dos dados de treinamento, produziu um dicionário com 20.365 *tokens* singulares. Além disso, foi considerado que o texto de entrada deve ter no mínimo 3 palavras. A rede tem como entrada texto compreendendo no máximo 1000 palavras e a saída com dimensão de 1x50. A rede foi treinada com 5 épocas e com mini *batches* de 32.

5 Base de Dados

Para a realização dos experimentos foram coletados dados postados no Facebook entre 2008 e 2017 de diversas fontes de notícias em inglês: Reuters, BCC, CNN, etc. Em termos quantitativos, a base de dados possui 11.538 itens coletados, sendo que a classe novidades contém cerca de 6% dos registros, o que corresponde a 742 itens. A classe normal contém cerca de 94% o que corresponde a 10.796 itens. Todos os itens possuem texto e imagem associados. A Tabela 1 detalha a quantidade de itens (imagem + texto) por classe. As restrições para filtragens das notícias ou dados foram estabelecidas por meio das palavras chaves próprias de cada classe.

Neste trabalho, a subclasse *Disasters* (desastres naturais) foi considerada como sendo novidade por ser considerado um evento inesperado, que pode causar prejuízos e mortes, cuja ocorrência é em menor quantidade se considerada as notícias como um todo dos demais assuntos. A base de dados, o código fon-

te, a lista de fontes e as palavras-chaves estão disponíveis para livre download[4].

Tabela 1. Detalhamento da base de dados

Classes	Subclasses	Quantidade de itens
Novidades	Disasters	742
Normal	Crime	1090
	Culture	1066
	Food	1118
	Gossip	1132
	Health	1148
	Movie	1022
	Nature	1032
	Politics	1074
	Sports	1040
	Technology	1074
Total		11538

6 Experimentos e Resultados

A arquitetura proposta neste trabalho foi desenvolvida usando as tecnologias Python 3.5 e o Keras 2.1.3. Todos os experimentos foram realizados em um computador equipado com uma placa de vídeo NVIDIA 1080 GPU 8G, processador Intel i7-2600K CPU 3.4 GHz, e 16GB de RAM, executando Ubuntu 14.04.3 LTS.

Os dados foram divididos em proporção de 80/20% para as amostras de treinamento/teste.

As métricas utilizadas foram: precisão, *recall* e acurácia. A precisão descrita na Equação 6 é definida como o percentual de novidades reportadas que verdadeiramente mostram ser novidades.

$$Precisão = \frac{(|S \cap G|)}{S} \times 100 \quad (6)$$

Detalhando elementos da fórmula, o *S* é o conjunto declarado como novidades após a classificação, e o *G* é o conjunto das novidades verdadeiras ou conhecidos como *Ground-Truth*. Portanto, a operação de intersecção $|S \cap G|$ pode ser reconhecida como o conjunto das novidades detectadas verdadeira-positivas. Logo, precisão avalia o quão preciso é o sistema ao afirmar que uma notícia é novidade.

O *recall* descrito na Equação 7 é o percentual de novidades G*round-Truth* que foram reportados como novidades. Assim, o *recall* avalia a capacidade do sistema de retornar o maior número possível de notícias que são novidades.

$$Recall = \frac{(|S \cap G|)}{G} \times 100 \quad (7)$$

Há um compromisso entre precisão e *recall*, ou seja, o aumento do *recall* faz a precisão reduzir e vice-versa. Para detecção de novidades é mais satisfatório um *recall* maior que a precisão, pois é melhor selecionar um falso-positivo, do que perder novidades verdadeiras (falsos-negativos) (Aggarwal et al., 2016). Além dessas métricas é usada a acurácia, descrita na Equação 8. No entanto, a acurácia não está tão ligada a medição das novidades, e sim em medir o total de dados normais e novidades classificadas corretamente. O algoritmo pode ter uma boa acurácia, contudo apresenta baixa precisão e *recall* para as novidades, já que essas são em menor proporção na base de dados.

$$Acc = \frac{(predições\ corretas)}{(total\ de\ predições)} \times 100 \quad (8)$$

Os resultados apresentados nas Tabelas 2 e 3 são as médias de 10 execuções de treinamento/teste. A Tabela 2 contém os resultados apenas da arquitetura da CNN Lppreconv1 de imagem proposta na Figura 2a e a Tabela 3 contém os resultados da arquitetura da AlexNet. Outrossim, foram incluídos os resultados dos algoritmos de fusão descritos nas Equações 4 e 5, respectivamente denominados fusão Max e Mag (Magnitude). Com o propósito de testar a robustez do método apresentado neste artigo, foi reduzida aleatoriamente a proporção de amostras da classe novidades dos conjuntos de teste e treinamento de 6% para 3%. Dessa forma, foram realizados testes sob o desbalanceamento de classes contendo 6% e 3% do total dos dados de novidades.

Tabela 2. Resultados com a CNN Lppreconv1

CNN	Precisão	*Recall*	Acurácia
6% de Novidades			
Fusão Max	80.18%	**88.00%**	**98.50%**
Fusão Mag	11,00%	78,00%	98,30%
Imagem	66.66%	23.00%	94.10%
Texto	**95.27%**	86.16%	98.00%
3% de Novidades			
Fusão Max	84.00%	**73.00%**	**99.10%**
Fusão Mag	12.00%	73.00%	99.00%
Imagem	50.00%	11.00%	96.68%
Texto	**96.00%**	72.00%	99.00%

Tabela 3. Resultados com a AlexNet

CNN	Precisão	*Recall*	Acurácia
6% de Novidades			
Fusão Max	78.05%	**86.00%**	98.00%
Imagem	15.00%	15.30%	94.02%
3% de Novidades			
Fusão Max	70.00%	**76.00%**	98.00%
Imagem	21.00%	10.00%	95.00%

Os testes realizados mostram que o algoritmo de fusão Max conseguiu melhorar o recall, tornando as CNN heterogêneas ligeiramente mais eficazes na identificação das novidades, mesmo quando os dados

[4] https://github.com/cba2018proj/proj

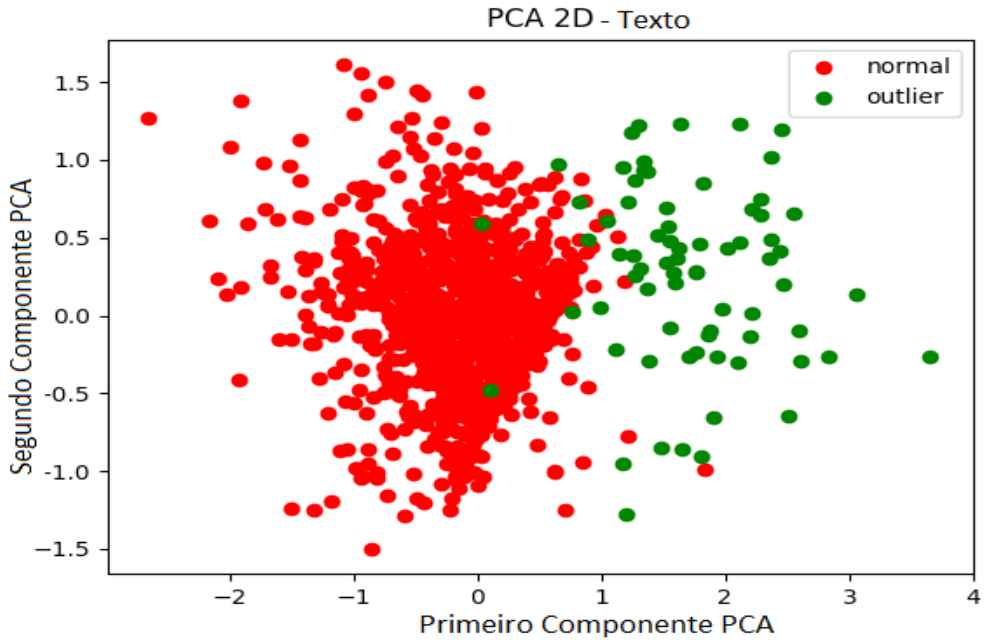

Figura 3. Primeiro e segundo componente principal da camada totalmente conectada da CNN de texto

de novidades foram reduzidos de 6% para 3% . No entanto, o algoritmo de fusão Mag não obteve sucesso tal como o algoritmo de fusão Max. Por esse motivo, obtém-se interesse apenas na fusão Max com a arquitetura AlexNet. Além disso, a arquitetura Lppreconv1 obteve melhor *recall* em relação a AlexNet.

Como mencionado na Seção 1, há desafios em classificar imagem e texto da internet por causa da qualidade dos dados disponibilizados nas redes sociais. De acordo com as métricas, a rede CNN de texto apresentou resultados relevantes, mas as CNN de imagem não produziram bons resultados, embora a arquitetura Lppreconv1 apresentou melhores resultados do que a AlexNet. À vista disto, foi examinado a distribuição dos dados escolhendo a camada totalmente conectada da rede CNN por meio da seleção dos 2 principais componentes. As Figuras 3 e 4 ilustram a distribuição dos dados por meio do PCA das camadas totalmente conectadas. Pode-se observar que a CNN de texto consegue projetar os dados em um espaço de características quase que totalmente separável, enquanto a CNN Lppreconv1 de imagem, apesar de agrupar as novidades, não conseguem separabilidade dos dados de novidade e normais.

A partir disso, foi identificado quais dessas classes e imagens normais se confundem com as novidades. As regiões no gráfico foram ampliadas para destacar onde as novidades se uniram com dados normais. Ao localizar a imagem 100 da classe nature e a 4 da classe novidades nota-se semelhança de cores, texturas e padrões tornando mais difícil a classificação tanto pela máquina quanto pelo próprio ser humano. Outros exemplos são encontrados semelhantes a esses, como das Figuras 5 e 6. É importante frisar a baixa qualidade de algumas imagens, como a 100 e 106 acentuando a dificuldade da classificação.

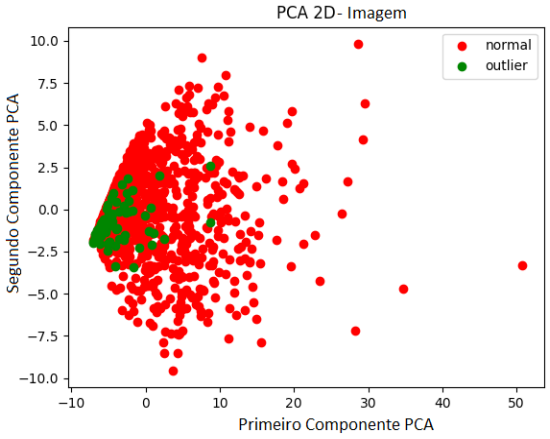

Figura 4. Primeiro e segundo componente principal da camada totalmente conectada da CNN da imagem

Figura 5. Do lado esquerdo a imagem 4 da classe novidades e do lado direito a imagem 100 da classe nature.

Figura 6. Do lado esquerdo a imagem 51 da classe novidades e do lado direito a imagem 106 da classe nature.

8 Conclusões e Trabalhos Futuros

A fusão de CNNs de imagem e texto proposta neste trabalho apresentou um *recall* superior as redes CNNs isoladas, principalmente em relação a CNN de imagem. No entanto, é necessário um aprofundamento na pesquisa no que tange às CNNs para tratar novidades em imagens, principalmente aquelas com baixa resolução.Como a tarefa de detecção de novidades tem forte ligação a classificadores o*ne-class*, tende-se a migrar a pesquisa para CNNs o*ne-class*.

Referências Bibliográficas

Aggarwal, C. C. (2016). Outliers Analysis, 2rd Edition, Springer Press.

Amorim, M.; Bortoloti, F.; Ciarelli, P. M.; De Souza, A. F. and Oliveira, E. (2018). Analysing rotation-invariance of a log-polar transformation in convolutional neural networks, The International Joint Conference on Neural Networks (IJCNN). in press.

Barnett, V. and Lewis T. (1994). Outliers in Statistical Data, 3rd Edition, Wiley Editorial Offices.

Bojanowski, P.; Grave, E.; Joulin, A. and Mikolov, T. (2016). Enriching Word Vectors with Subword Information, Journal CoRR, Vol. 1607.04606.

BusinessInsider (2013). "These Are The Six Most Discussed Topics On Social Media", Disponível em: http://www.businessinsider.com/most-discussed-topics-on-social-media-2013-5. Acesso em: 21 de Março de 2018.

Chandola, V.; Banerjee, A. and Kumar, V. (2009). Anomaly Detection: A Survey. Journal ACM Computing Surveys, Volume 41, Issue 3.

Ding, X.; Liu, T. and Duan, J. (2015). Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network, In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence Mining, pp 2389-2395.

Gao , Y.; Zhao, S.; Yang, Y. and Chua, T-S. (2015). Multimedia Social Event Detection in Microblog. 21st International Conference MultiMedia Modeling, Volume 8935, Issue 4, pp 269-281. doi:10.1007/978-3-319-14445-0_24.

Gonzalez, R. C. and Woods, R. E. (2007), Digital Image Processing, 3rd Edition, Pearson Press.

Kingma, D. P. and Ba J. L. (2015). Adam: A Method for Stochastic Optimization, ICLR 3rd International Conference for Learning Representations, San Diego.

Koehn, P. (2017). Statistical Machine Translation, Disponível em: http://mt-class.org/jhu/assets/papers/neural-network-models.pdf. Acesso em: 22 de Março de 2018.

Krizhevsky, A.; Sutskever, I. and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems.

LATimes (2016). "LATimes–NewsDaily", Disponível em:http://latimesblogs.latimes.com/technology/2011/08. Acesso em: 20 de Março de 2018.

Levy, O.; Goldberg, Y. and Dagan, I. (2015). Improving Distributional Similarity with Lessons Learned from Word Embeddings, Transactions of the Association for Computational Linguistics. 3. 211-225.

Merriam-Webster (2018), An encyclopedia britannica company. Disponível em: https://www.merriam-webster.com/. Acesso em: 01 de Março de 2018.

Pennington, J.; Socher, R. and Manning D. M. (2014). GloVe: Global Vectors for Word Representation.

Pimentel, M. A. F.; Clifton, D. A.; Clifton L. and Tarassenko, L. (2014). A Review of Novelty Detection, Elsevier, Signal Processing.

Preisach, C.; Burkhardt, H.; Schimidt-Thieme, L. and Decker, R. (2008). Data Analysis, Machine Learning and Applications, Spring-Verlag Press.

Ribeiro, M.; Lazzaretti, A. E.and Lopes, H. S. (2017). A study of deep convolutional auto-encoders for anomaly detection in videos. Pattern Recognition Letters, Volume 105, pp 13-22.

Ruocco, M. and Ramampiaro, H. (2014). A scalable algorithm for extraction and clustering of event-related pictures. Journal Multimedia Tools and Applications, Volume 70, Issue 1, pp 55-88. doi:10.1007/s11042-012-1087-z.

Ruta, D. and Gabrys, B. (2000). An Overview of Classifier Fusion Methods, Computing and Information Systems, pp 1–10.

Zhang, M.; Wu, J.; Huifeng, L.; Peng, Y. and Yanan, S. and Gabrys, B. (2017). The Application of One-Class Classifier Based on CNN in Image Defect Detection, Procedia Computer Science, pp 341–348.

Zhou, S.; Shen, W.; Zeng, D.; Fang, M.; Yuanwang, W. and Zhijiang, Z. (2016). Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes, Signal Processing: Image Communication, Volume 47, pp 358–368.