VITOR ARINS PINTO

REDES NEURAIS CONVOLUCIONAIS DE PROFUNDIDADE PARA RECONHECIMENTO DE TEXTOS EM IMAGENS DE CAPTCHA.

Florianópolis

7 de agosto de 2016

VITOR ARINS PINTO

REDES NEURAIS CONVOLUCIONAIS DE PROFUNDIDADE PARA RECONHECIMENTO DE TEXTOS EM IMAGENS DE CAPTCHA.

Trabalho de Conclusão de Curso submetido ao Programa de graduação da Universidade Federal de Santa Catarina para a obtenção do Grau de Bacharel em Sistemas de Informação.

Orientadora: Luciana de Oliveira Rech

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA

Florianópolis

7 de agosto de 2016

AGRADECIMENTOS

RESUMO

Atualmente muitas aplicações na Internet seguem a política de manter alguns dados acessíveis ao público. Para isso é necessário desenvolver um portal que seja robusto o suficiente para garantir que todas as pessoas possam acessá-lo. Porém as requisições feitas para recuperar dados públicos nem sempre vêm de um ser humano. Empresas especializadas em Big data possuem um grande interesse em fontes de dados públicos para poder fazer análises e previsões a partir de dados atuais. Com esse interesse, Web Crawlers são implementados. Eles são responsáveis por consultar fontes de dados milhares de vezes ao dia, fazendo diversas requisições a um site. Tal site pode não estar preparado para um volume de consultas em um período tão curto de tempo. Com o intuito de impedir que sejam feitas consultas por programas de computador, as instituições que mantêm dados públicos investem em ferramentas chamadas CAPTCHA (teste de Turing público completamente automatizado, para diferenciação entre computadores e humanos). Essas ferramentas geralmente se tratam de imagens contendo um texto qualquer e o usuário deve digitar o que vê na imagem. O objetivo do trabalho proposto é realizar o reconhecimento de texto em imagens de CAPTCHA através da aplicação de redes neurais convolucionais.

ABSTRACT

Currently many applications on the Internet follow the policy of keeping some data accessible to the public. For this it is necessary to develop a portal that is robust enough to ensure that all people can access this data. But the requests made to recover Public Data not always come from a human. companies specializing in Big data have a great interest in data from public sources in order to make analyzes and forecasts from current data. With this interest, Web Crawlers are implemented. They are responsible for querying data sources thousands of times a day, making several requests to a site. This site may not be prepared for a volume of inquiries in a short period of time. In order to prevent queries to be made by computer programs, institutions that keep public data invest in tools called CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*). These tools usually deal with images containing text and the user must enter what he or she sees in the image. The objective of the proposed work is to perform the text recognition in CAPTCHA images through the application of convolutional neural networks.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

CAPTCHA ..... *Completely Automated Public Turing test to tell Computers and Humans Apart*
HDF ..... *Hierarchical Data Format*
GPU ..... *Graphics Processing Unit*
AWS ..... *Amazon Web Services*
IA ..... *Inteligência Artificial*
DCNN ..... *Deep Convolutional Neural Networks*
ReLU ..... *Rectified Linear Unit*

SUMÁRIO

1 INTRODUÇÃO

Redes neurais artificiais clássicas existem desde os anos 60, como fórmulas matemáticas e algorítimos. Atualmente os programas de aprendizado de máquina contam com diferentes tipos de redes neurais. Um tipo de rede neural muito utilizado para processamento de imagens é a rede neural convolucional de profundidade. O trabalho em questão tratará da utilização e configuração de uma rede neural convolucional de profundidade para reconhecimento de textos em imagens.

1.1 PROBLEMA

Com o aumento constante na quantidade de informações geradas e computadas atualmente, percebe-se o surgimento de uma necessidade de tornar alguns tipos de dados acessíveis a um público maior. A fim de gerar conhecimento, muitas instituições desenvolvem portais de acesso para consulta de dados relevantes a cada pessoa. Esses portais, em forma de aplicações na Internet, precisam estar preparados para receber diversas requisições e em diferentes volumes ao longo do tempo.

Devido a popularização de ferramentas e aplicações especializadas em Big data, empresas de tecnologia demonstram interesse em recuperar grandes volumes de dados de diferentes fontes públicas. Para a captura de tais dados, Web crawlers são geralmente implementados para a realização de várias consultas em aplicações que disponibilizam dados públicos.

Para tentar manter a integridade da aplicação, as organizações que possuem estas informações requisitadas investem em ferramentas chamadas CAPTCHA (teste de Turing público completamente automatizado para diferenciação entre computadores e humanos). Essas ferramentas frequentemente se tratam de imagens contendo um texto qualquer e o usuário precisa digitar o que vê na imagem.

O trabalho de conclusão de curso proposto tem a intenção de retratar a ineficiência de algumas ferramentas de CAPTCHA, mostrando como redes neurais convolucionais podem ser

aplicadas em imagens a fim de reconhecer o texto contido nestas imagens.

1.2 OBJETIVOS

1.2.1 *OBJETIVO GERAL*

Analisar o treinamento e aplicação de redes neurais convolucionais de profundidade para o reconhecimento de texto em imagens de CAPTCHA.

1.2.2 *OBJETIVOS ESPECÍFICOS*

- Estudar trabalhos correlatos e analisar o estado da arte;
- Entender como funciona cada aspecto na configuração de uma rede neural convolucional;
- Realizar o treinamento e aplicação de uma rede neural artificial para reconhecimento de CAPTCHAs.

1.3 ESCOPO DO TRABALHO

O escopo deste trabalho inclui o estudo e análise de redes neurais convolucionais de profundidade para reconhecimento de texto em imagens de CAPTCHA. Não está no escopo do trabalho analisar outras formas de inteligência no reconhecimento de texto. Também não está no escopo do trabalho o estudo, análise ou implementação da aplicação de redes neurais convolucionais para outros tipos de problemas.

1.4 METODOLOGIA

Para realizar o proposto, foram feitas pesquisas em base de dados tais como IEE Xplorer e ACM Portal. Adquirindo assim maior conhecimento sobre o tema, estudando trabalhos relacionados.

Com base no estudo do estado da arte, foram feitas pesquisas e estudos para indicar caminhos possíveis para desenvolvimento da proposta de trabalho.

1.5 ESTRUTURA DO TRABALHO

Para uma melhor compreensão e separação dos conteúdos, este trabalho está organizado em 6 capítulos.

O capítulo 2 apresenta a fundamentação teórica, com as definições das abordagens de desenvolvimento de aprendizado de máquina e redes neurais. Os conceitos de tipos de redes neurais.

No capítulo 3 está a proposta de experimento a ser realizado. Assim como uma breve ideia dos resultados esperados e a forma de avaliação dos mesmos.

O capítulo 4 contém as informações do desenvolvimento do sistema de reconhecimento de imagens de CAPTCHA. Também a apresentação dos dados obtidos através das metodologias escolhidas na seção anterior.

No capítulo 5 são apresentados os resultados da aplicação do sistema de reconhecimento de imagens de CAPTCHA e as ameaças que podem comprometer o acesso à dados públicos disponbilizados.

Por fim, no capítulo 6 estão as conclusões obtidas através dos resultados deste trabalho e as sugestões para trabalhos futuros relacionados.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 APRENDIZADO DE MÁQUINA

Aprendizado de máquina, ou *Machine Learning*, é uma área da computação que emergiu de estudos relacionados ao reconhecimento de padrões e inteligência artificial. Nesta área é contemplado o estudo e implementação de algoritmos que conseguem aprender e fazer previsões baseadas em dados. Esses algoritmos funcionam através da construção de um modelo preditivo que tem como entrada um conjunto de treinamento com dados de observações quaisquer. Desse modo as previsões são feitas orientadas aos dados e não a partir de instruções estáticas de um programa.

2.2 REDES NEURAIS

Diante das ferramentas disponíveis que tratam de aprendizado de máquina, uma delas é a rede neural artificial.

Redes neurais artificiais são conjuntos de modelos inspirados por redes neurais biológicas, usados para aproximar funções que dependem de um número muito grande de entradas. De acordo com Mackay[1], Redes neurais geralmente são especificadas utilizando 3 coisas:

- **Arquitetura:** Especifica quais variáveis estão envolvidas na rede e quais as relações topológicas. Por exemplo, as variáveis envolvidas em uma rede neural podem ser os pesos das conexões entre os neurônios.
- **Regra de atividade:** A maioria dos modelos de rede neural tem uma dinâmica de atividade com escala de tempo curta. São regras locais que definem como as *atividades* de neurônios mudam em resposta aos outros. Geralmente a regra de atividade depende dos parâmetros da rede.
- **Regra de aprendizado:** Especifica o modo com que os pesos da rede neural muda conforme o tempo. O aprendizado normalmente toma uma escala de tempo maior do que

a escala referente a dinâmica de atividade. Normalmente a regra de aprendizado dependerá das *atividades* dos neurônios. Também pode depender dos valores que são objetivos definidos pelo usuário e valores iniciais dos pesos.

Tomando imagens como exemplo, uma rede neural para reconhecimento de texto pode ter como entrada o conjunto de pixels da imagem. Depois de serem atribuídos os pesos para cada item da entrada, os próximos neurônios serão ativados mediante a função de atividade pré-definida. Os pesos são recalculados através da regra de aprendizado e todo processo é repetido até uma condição determinada pelo usuário.

2.3 CONVOLUÇÕES

Para entender redes neurais convolucionais, é necessário primeiro entender o que são convoluções. Segundo Olah[2], uma convolução pode ser vista como um somatório das probabilidades de resposta de duas funções algébricas. Tendo como definição padrão de convolução a seguinte expressão:

$$(f * g)(c) = \sum_{a} f(a) \cdot g(c - a) \tag{2.1}$$

Onde $\boldsymbol{f}$ e $\boldsymbol{g}$ são duas funções, $\boldsymbol{c}$ é o parâmetro de entrada para a função final e $\boldsymbol{a}$ é um parâmetro de entrada escolhido para uma das funções, geralmente uma diferença temporal.

Como podemos considerar que imagens são funções bidimensionais, é comum realizar transformações por meio de convoluções. Estas convoluções são executadas com uma função local pequena chamada de “kernel”.

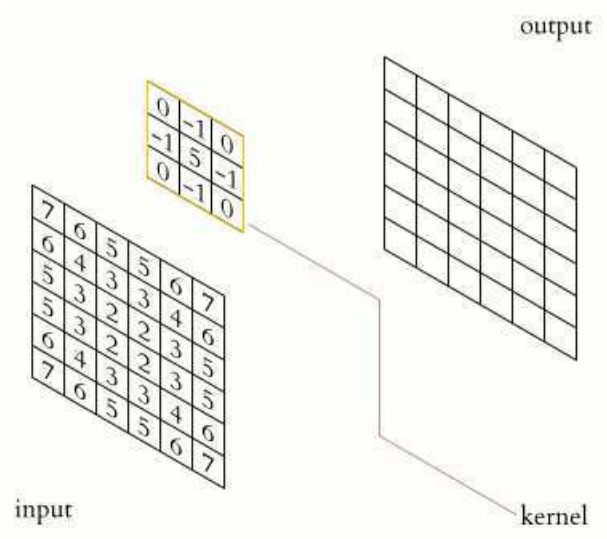

Figura 1: Aplicação de uma função “kernel” sobre a função de uma imagem que é o “input”.

2.3.1 *APLICAÇÃO EM REDES NEURAIS*

Redes neurais convolucionais são muito similares a redes neurais comuns. De acordo com Karpathy[3]:

> “Arquiteturas de redes convolucionais assumem explicitamente que as entradas são imagens, o que nos permite cifrar algumas propriedades dentro da arquitetura. Essas então fazem a função de ativação mais eficiente de implementar e reduz drasticamente a quantidade de parâmetros na rede.” (KARPATHY[3], 2015, tradução nossa).

Portanto para o caso de reconhecimento de texto em imagens, as redes neurais convolucionais fazem muito sentido.

2.4 APRENDIZADO EM PROFUNDIDADE

O aprendizado em profundidade permite que modelos computacionais compostos por múltiplas camadas de processamento possam aprender representações de dados com múltiplos níveis de abstração[4].

A solução de *Deep learning* permite que computadores aprendam a partir de experiencias e entendam o mundo em termos de uma hierarquia de conceitos, com cada conceito definido em termos da sua relação com conceitos mais simples. Juntando conhecimento de experiência, essa abordagem evita a necessidade de ter operadores humanos especificando formalmente todo o

conhecimento que o computador precisa. A hierarquia de conceitos permite que o computador aprenda conceitos complexos construindo-os à partir de conceitos mais simples. Desenhando um gráfico que mostra como esses conceitos são construídos em cima de outros, o gráfico fica profundo, com muitas camadas. Por esta razão, essa abordagem para IA é chamada de Aprendizado em profundidade[5].

2.5 REDES NEURAIS CONVOLUCIONAIS DE PROFUNDIDADE

A grande vantagem na abordagem de redes neurais convolucionais de profundidade (DCNN) para reconhecimento é que não é necessário um extrator de características desenvolvido por um ser humano. Nas soluções de [6] e [7] é possível perceber que foram usadas diversas camadas para o aprendizado das características.

Em arquiteturas de profundidade, as funções de ativação dos neurônios são unidades lineares retificadas (ReLU). Isso simplifica o uso de *backpropagation* e evita problemas de saturação, fazendo o aprendizado ficar muito mais rápido.

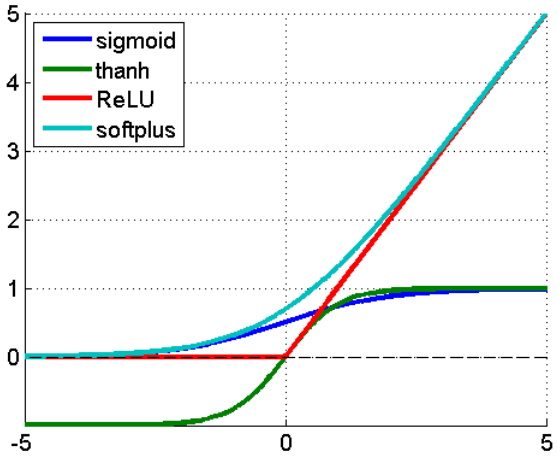

Figura 2: Comparação de funções de ativação.

Ao combinar o aprendizado em profundidade com redes convolucionais, conseguimos tratar problemas muito mais complexos de classificação em imagens. Assim problemas mais simples, como o reconhecimento de textos, podem ser resolvidos cada vez mais rápido e facilmente.

Citar Hinton[8] e Jaderberg[9]

3 PROPOSTA DE EXPERIMENTO

Para realizar o experimento será necessário treinar um modelo de rede neural que seja capaz, ou esteja próximo, de decifrar um CAPTCHA. Para isso serão efetuadas três etapas básicas e comuns quando se trabalha com redes neurais. Primeiro será coletado o maior número possível de imagens de CAPTCHA. Em seguida será gerado um dataset com as características dessas imagens junto com a classe em que pertence. A partir daí podemos realizar a configuração e treinamento da rede neural. E por fim será testada a acurácia do modelo mediante imagens de teste.

3.1 COLETA DE IMAGENS

A coleta de imagens será feita através de um script em Python que irá acessar a página que possui uma imagem de CAPTCHA e assim fazer o download da mesma. O script ficará em loop até que seja feita a coleta de um número suficiente de imagens.

3.2 GERAÇÃO DO CONJUNTO DE DADOS

Para criar os arquivos de conjunto de dados (ou "dataset"), será necessário utilizar o formato de arquivo **HDF**, mais especificamente a versão 5 (HDF5). O formato de arquivo de dados hierárquico possibilita a manipulação de conjuntos de dados extremamente grandes e complexos.

3.2.1 *PRÉ-PROCESSAMENTO*

A fase de pré-processamento das imagens é mínima e é feita em tempo de execução da geração do conjunto de dados.

- **Escala de cinza**

Ao gerar um array representativo da imagem, apenas é considerado um valor de escala de cinza da imagem, assim padronizando os valores de intensidade de pixels entre 0 e 1.

- **Redimensionamento**

  Ao gerar o array que representa a imagem, é feito um cálculo para diminuir a imagem com base em uma escala. Essa escala será configurada à partir de um valor padrão para a largura e altura das imagens.

3.2.2 *CONJUNTO DE DADOS DE TESTE*

Para o treinamento será necessário um conjunto separado para teste que não possui nenhuma imagem presente no conjunto de treinamento. Para isso será coletada uma amostra aleatória com cerca de 2% do número de imagens do conjunto de treinamento para geração do conjunto de dados de teste.

3.3 TREINAMENTO

Após gerado um arquivo contendo o conjunto de dados, é possível trabalhar no treinamento do modelo da rede neural. Para isso será usado o Framework **TensorFlow**[10] destinado à *Deep Learning* e um script em Python que fará uso das funções disponibilizadas pela biblioteca do TensorFlow. Assim realizando o treinamento até atingir um valor aceitável de acerto no conjunto de teste. O resultado do treinamento será um arquivo binário representando o modelo que será utilizado para avaliação posteriormente.

3.3.1 *INFRAESTRUTURA*

Com o intuito de acelerar o processo, foi utilizada uma máquina com **GPU** para o treinamento. A máquina foi adquirida em uma *Cloud* privada da AWS. A GPU utilizada se trata de uma *NVIDIA GRID K520* com 1.536 cores e 4GB de memória de vídeo. Como processador a máquina possui um *Intel Xeon E5-2670 (Sandy Bridge)* com 8 cores, e ainda possui 15GB de memória RAM [11][12].

3.4 AVALIAÇÃO DE ACURÁCIA

Para a avaliação, uma nova amostra de imagens será coletada do mesmo modo que foram coletadas as imagens para treinamento. Essa amostra terá uma quantidade maior de imagens do

que o conjunto de teste.

Com essas amostra de imagens, será feita a execução do teste do modelo contra cada uma das imagens, assim armazenando uma informação de erro ou acerto do modelo. Ao final da execução será contabilizado o número de acertos e comparado com o número total da amostra de imagens para avaliação. Resultando assim em uma porcentagem que representa a acurácia do modelo gerado.

REFERÊNCIAS

[1] MACKAY, D. J. C. *Information Theory , Inference And Learning Algorithms*. Cambridge University Press, 2005. ISBN 9780521670517. Disponível em: <http://www.inference.phy.cam.ac.uk/mackay/itila/>.

[2] OLAH, C. *Understanding Convolutions*. 2014. Acessado: 18/07/2016. Disponível em: <http://colah.github.io/posts/2014-07-Understanding-Convolutions/>.

[3] KARPATHY, A. *CS231n Convolutional Neural Networks for Visual Recognition*. Acessado: 06/08/2016. Disponível em: <http://cs231n.github.io/convolutional-networks/>.

[4] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. Disponível em: <https://www.cs.toronto.edu/ hinton/absps/NatureDeepReview.pdf>.

[5] BENGIO, I. G. Y.; COURVILLE, A. Deep learning. Book in preparation for MIT Press. 2016. Disponível em: <http://www.deeplearningbook.org>.

[6] KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks.

[7] GOODFELLOW, I. J. et al. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. Disponível em: <http://arxiv.org/pdf/1312.6082v4.pdf>.

[8] HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A fast learning algorithm for deep belief nets. Disponível em: <https://www.cs.toronto.edu/ hinton/absps/fastnc.pdf>.

[9] JADERBERG, M. et al. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition. Disponível em: <http://arxiv.org/pdf/1406.2227v4.pdf>.

[10] TENSORFLOW. *TensorFlow — an Open Source Software Library for Machine Intelligence*. Acessado: 03/08/2016. Disponível em: <https://www.tensorflow.org/>.

[11] AWS. *Linux GPU Instances - Amazon Elastic Compute Cloud*. Acessado: 02/08/2016. Disponível em: <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using_cluster_computing.html#install-nvidia-driver>.

[12] AWS. *EC2 Instance Types – Amazon Web Services (AWS)*. Acessado: 02/08/2016. Disponível em: <https://aws.amazon.com/ec2/instance-types/#gpu>.