Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Spatio-Temporal Trend Analysis of the Brazilian Elections based on Twitter Data

Bruno Justino Garcia Praciano

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Orientador
Prof. Dr. -Ing João Paulo Lustosa da Costa

Brasília
2014

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Spatio-Temporal Trend Analysis of the Brazilian Elections based on Twitter Data

Bruno Justino Garcia Praciano

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Prof. Dr. -Ing João Paulo Lustosa da Costa (Orientador)
ENE/UnB

Prof. Dr. Rafael Timóteo de Sousa Júnior
ENE/UnB

Dr. -Ing Ricardo Kerhle Miranda
ENM/UnB

Prof. Dr. José Edil Guimarães de Medeiros
Coordenador do Curso de Engenharia da Computação

Brasília, 24 de dezembro de 2014

Dedicatória

Na *dedicatória* o autor presta homenagem a alguma pessoa (ou grupo de pessoas) que têm significado especial na vida pessoal ou profissional. Por exemplo (e citando o poeta): *Eu dedico essa música a primeira garota que tá sentada ali na fila. Brigado!*

Agradecimentos

Nos *agradecimentos*, o autor se dirige a pessoas ou instituições que contribuíram para elaboração do trabalho apresentado. Por exemplo: *Agradeço aos gigantes cujos ombros me permitiram enxergar mais longe. E a Google e Wikipédia.*

Resumo

O *resumo* é um texto inaugural para quem quer conhecer o trabalho, deve conter uma breve descrição de todo o trabalho (apenas um parágrafo). Portanto, só deve ser escrito após o texto estar pronto. Não é uma coletânea de frases recortadas do trabalho, mas uma apresentação concisa dos pontos relevantes, de modo que o leitor tenha uma ideia completa do que lhe espera. Uma sugestão é que seja composto por quatro pontos: 1) o que está sendo proposto, 2) qual o mérito da proposta, 3) como a proposta foi avaliada/validada, 4) quais as possibilidades para trabalhos futuros. É seguido de (geralmente) três palavras-chave que devem indicar claramente a que se refere o seu trabalho. Por exemplo: *Este trabalho apresenta informações úteis a produção de trabalhos científicos para descrever e exemplificar como utililzar a classe LaTeX do Departamento de Ciência da Computação da Universidade de Brasília para gerar documentos. A classe `UnB-CIC` define um padrão de formato para textos do CIC, facilitando a geração de textos e permitindo que os autores foquem apenas no conteúdo. O formato foi aprovado pelos professores do Departamento e utilizado para gerar este documento. Melhorias futuras incluem manutenção contínua da classe e aprimoramento do texto explicativo.*

Palavras-chave: Big Data, Aprendizado de Máquina Supervisionado, Análise de Sentimentos, Máquina de Vetor de Suporte

Abstract

O *abstract* é o resumo feito na língua Inglesa. Embora o conteúdo apresentado deva ser o mesmo, este texto não deve ser a tradução literal de cada palavra ou frase do resumo, muito menos feito em um tradutor automático. É uma língua diferente e o texto deveria ser escrito de acordo com suas nuances (aproveite para ler `http://dx.doi.org/10.6061%2Fclinics%2F2014(03)01`). Por exemplo: *This work presents useful information on how to create a scientific text to describe and provide examples of how to use the Computer Science Department's LaTeX class. The `UnB-CIC` class defines a standard format for texts, simplifying the process of generating CIC documents and enabling authors to focus only on content. The standard was approved by the Department's professors and used to create this document. Future work includes continued support for the class and improvements on the explanatory text.*

Keywords: Big Data, Supervised Machine Learning, Sentiment Analysis, Support Vector Machine

Contents

Chapter 1

Introduction

1.1 Motivation

Com a invenção e popularização da internet tem revolucionado as sociedades com o passar do tempos, pois agora é possível conectar várias pessoas, e realizar trocar de informações em tempo realizar e com o custo muito baixo em relação aos veículos tradicionais de mídia [**?**].

As notícias tem sido compartilhadas de maneira muito rápida e eficente e com a utilização massiva das redes de relacionamento, as pessoas podem trocar ideias e opiniões acerca de determinado assunto e com isso facilitar o acesso de todos. Com o passar dos anos as redes sociais já fazem parte da vida de várias pessoas, e com isso as relações interpessoais modificaram-se e esse mundo tem gerado muitos dados de fácil e livre acesso [1].

1.2 Problems

1.3 Objectives

1.4 Related work

1.5 Chapters description

Chapter 2

Concepts on Machine Learning and Text Mining

2.1 Basic Concepts

2.1.1 Training and testing phases

2.1.2 Learning of paradigms

2.1.3 Performance measures

2.2 Machine Learning

2.2.1 Support Vector Machine

2.2.2 Naive Bayes

2.2.3 Decision Trees

2.2.4 Logistic Regression

2.3 Natural Language Processing

2.3.1 Pre-Processing

2.3.2 Tokenization

2.3.3 Stemming

2.3.4 Stop Words

2.3.5 Bag of Words

2.3.6 Term Frequency

2.3.7 Term Frequency - Inverse Document Frequency (TF-IDF)

Chapter 3

Proposed Framework

3.1 Crawling and Tweet Extraction

3.2 Data Pre-Processing

3.3 Lexical Dictionary

3.4 Sentiment Classification

3.5 Data Visualization

Chapter 4

Results

4.1 Perfomance evaluation of trend analysis

4.2 N-Fold Cross Validation

4.3 Error evaluation of the sentiment analysis via SVM

4.4 Spatio Trend Analysis

4.5 Election Results

Chapter 5

Conclusion

Este documento serve de exemplo da utilização da classe `UnB-CIC` para escrever um texto cujo objetivo é apresentar os resultados de um trabalho científico. A sequência de ideias apresentada deve fluir claramente, de modo que o leitor consiga compreender os principais conceitos e resultados apresentados, bem como encontrar informações sobre conceitos secundários.

Referências

[1] Araniti, G., I. Bisio e M. De Sanctis: *Towards the reliable and efficient interplanetary internet: A survey of possible advanced networking and communications solutions*. Em *2009 First International Conference on Advances in Satellite and Space Communications*, páginas 30–34, July 2009. 1

Appendix A

A.1 Appendix