Deep Sentiment Analysis on Tumblr

Anthony Hu

Department of Statistics
University of Oxford
Oxford, United Kingdom

September 2017

Declaration

The work in this thesis is based on research carried out at the Department of Statistics, University of Oxford. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

Many thanks to my supervisor Seth Flaxman and to Kellogg College.

Deep Sentiment Analysis on Tumblr

Anthony Hu

Submitted for the degree of Master of Science in Applied Statistics
September 2017

Abstract

This thesis proposes a novel approach to sentiment analysis using deep neural networks on both images and text.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Chapter 2

Tumblr data

2.1 Overview of the data

Tumblr's posts were extracted using the official API thanks to their tags that were taken as the ground truth. The tags represent the user's emotion: happy, sad, angry, surprised, scared or disgusted. The data extraction took several weeks due to the API's limitations: 1,000 requests per hour and 5,000 requests per day, with each request containing 20 posts. The final dataset has about one million posts and six different emotions.

Here is an example of a post with its associated image:

Chapter 3

Visual recognition

3.1 Section 1

Chapter 4

Natural Language Processing

Text analysis

4.1 Section 1

Chapter 5

Recurrent Neural Networks for text generation

5.1 Section 1

Chapter 6

Useful maths in LaTeX

6.1 Equations related

$$\frac{\partial u_1}{\partial t} = \Delta w_1 \quad \text{in } \Omega, t > 0, \tag{6.1}$$

$$\frac{\partial u_2}{\partial t} = \Delta w_2 \quad \text{in } \Omega, t > 0, \tag{6.2}$$

where

$$w_1 = \frac{\delta F(u_1, u_2)}{\delta u_1}, \tag{6.3}$$

$$w_2 = \frac{\delta F(u_1, u_2)}{\delta u_2}, \tag{6.4}$$

$$
\begin{aligned}
F(u_1, u_2) = {}& b_1 u_1^4 - a_1 u_1^2 + c_1 |\nabla u_1|^2 \\
& + b_2 u_2^4 - a_2 u_2^2 + c_2 |\nabla u_2|^2 \\
& + D \left(u_1 + \sqrt{\frac{a_1}{2b_1}} \right)^2 \left(u_2 + \sqrt{\frac{a_2}{2b_2}} \right)^2.
\end{aligned}
\tag{6.5}
$$

$$U_1^n = \sum_{i=1}^{J} U_{1,i}^n \eta_i, \quad W_1^n = \sum_{i=1}^{J} W_{1,i}^n \eta_i, \tag{6.6}$$

$$U_2^n = \sum_{i=1}^{J} U_{2,i}^n \eta_i, \quad W_2^n = \sum_{i=1}^{J} W_{2,i}^n \eta_i, \tag{6.7}$$

We also use the following notation, for $1 \leq q < \infty$,

$$L^q(0,T;W^{m,p}(\Omega)) := \left\{ \eta(x,t) : \ \eta(\cdot,t) \in W^{m,p}(\Omega), \int_0^T \|\eta(\cdot,t)\|_{m,p}^q \, dt < \infty \right\},$$

$$L^\infty(0,T;W^{m,p}(\Omega)) := \left\{ \eta(x,t) : \eta(\cdot,t) \in W^{m,p}(\Omega), \ \operatorname*{ess\,sup}_{t \in (0,T)} \|\eta(\cdot,t)\|_{m,p} < \infty \right\},$$

Cases

$$|v|_{0,r} \leq C|v|_{0,p}^{1-\mu} \|v\|_{m,p}^{\mu}, \quad \text{holds for } r \in \begin{cases} [p,\infty] & \text{if } m - \frac{d}{p} > 0, \\ [p,\infty) & \text{if } m - \frac{d}{p} = 0, \\ [p,-\frac{d}{m-d/p}] & \text{if } m - \frac{d}{p} < 0. \end{cases} \quad (6.8)$$

6.2   Writing

Lemma 6.2.1 Let $u,v,\eta \in H^1(\Omega)$, $f = u - v$, $g = u^m v^{n-m}$, $m,n = 0,1,2$, and $n - m \geq 0$. Then for $d = 1,2,3$,

$$\left| \int_\Omega fg\eta dx \right| \leq C|u-v|_0 \, \|u\|_1^m \, \|v\|_1^{n-m} \, \|\eta\|_1. \quad (6.9)$$

Proof: Note that using the Cauchy-Schwarz inequality we have

$$|(u)^m v^{n-m}|_{0,p} \leq \begin{cases} |u|_{0,2mp}^m \, |v|_{0,2(n-m)p}^{(n-m)} & \text{for } n - m \neq 0, \ \text{and } m \neq 0, \\ |u|_{0,mp}^m \text{ or } |v|_{0,(n-m)p}^{(n-m)} & \text{for } m = 0, \ \text{or } n - m = 0 \text{ respectively.} \end{cases}$$

Noting the generalise Hölder inequality and the result above we have

$$\left| \int_\Omega fg\eta dx \right| \leq |u-v|_0 \, |u^m v^{n-m}|_{0,3} \, |\eta|_{0,6},$$

$$\leq |u-v|_0 \ |\eta|_{0,6} \begin{cases} |u|_{0,6}^2 & \text{for } m = 2, \\ |u|_{0,6} \, |v|_{0,6} & \text{for } m = 1, \\ |v|_{0,6}^2 & \text{for } m = 0, \end{cases}$$

$$\leq C|u-v|_0 \, \|u\|_1^m \, \|v\|_1^{n-m} \, \|\eta\|_1,$$

where we have noted (6.8) to obtain the last inequality. This ends the proof. □

We consider the problem:

(**P**)   Find $\{u_i, w_i\}$ such that $u_i \in H^1(0, T; (H^1(\Omega))') \cap L^\infty(0, T; H^1(\Omega))$ for *a.e.* $t \in (0, T)$, $w_i \in L^2(0, T; H^1(\Omega))$

$$\left\langle \frac{\partial u_1}{\partial t}, \eta \right\rangle$$

Chapter 7

Conclusions

Bibliography

[1] J. W. Barrett and J. F. Blowey (1995), *An error bound for the finite element approximation of the Cahn-Hilliard equation with logarithmic free energy*, Numerische Mathematics, **72**, pp 1–20.

[2] J. W. Barrett and J. F. Blowey (1997), *Finite element approximation of a model for phase separation of a multi-component alloy with non-smooth free energy*, Numerische Mathematics, **77**, pp 1–34.

[3] J. W. Barrett and J. F. Blowey (1999a), *An improve error bound for finite element approximation of a model for phase separation of a multi-component alloy*, IMA J. Numer. Anal. **19**, pp 147-168.

[4] P. G. Ciarlet (1978), **The Finite Element Method for Elliptic Problems**, North-Holland.

[5] J. L. Lions (1969), **Quelques Méthodes de Résolution des Problémes aux Limites**, Dunod.

Appendix A

Basic and Auxiliary Results

A.1   Basic Results