Ciclo de Coloquios 2016

El Departamento de Informática de la Universidad Técnica Federico Santa María invita a la comunidad Universitaria a un nuevo coloquio a realizarse el día **Martes 28 de Junio a las 12:00 Hrs.** en Laboratorio de Programación Avanzada (LPA, B-038) del Campus San Joaquín. Como es habitual, la charla se transmitirá por videoconferencia al Auditorio Claudio Matamoros (F-106) en Casa Central.

Título

Acquiring and Exploiting Lexical Knowledge for Twitter Sentiment Analysis

Expositor

Felipe José Bravo Márquez
University of Waikato, New Zealand

Mini Bio

Felipe Bravo-Marquez is currently doing his PhD at the Machine Learning Group in the University of Waikato, New Zealand. He received two engineering degrees in the fields of computer science and industrial engineering, and a master's degree in computer science, all from the University of Chile. He worked for three years as a research engineer at Yahoo! Labs Latin America. His main areas of interest are: data mining, machine learning, information retrieval, and sentiment analysis.

You can find his full list of publications at:
http://www.cs.waikato.ac.nz/~fjb11/

Resumen

The most popular sentiment analysis task in Twitter is the automatic classification of tweets into sentiment categories such as positive, negative, and neutral. State-of-the-art solutions to this problem are based on supervised machine learning models trained from manually annotated examples. These models are affected by a label sparsity problem, because the manual annotation of tweets is labour-intensive and time-consuming.

In this presentation, we discuss how to address the label sparsity problem by building two type of resources that can be exploited when labelled data is scarce: opinion lexicons and synthetically labelled tweets. We explain how to build Twitter-specific opinion lexicons by training words-level classifiers using representations that exploit different sources of information such as (a) the morphological information conveyed by part-of-speech (POS) tags, (b) associations between words and the sentiment expressed in the tweets that contain them, and (c) distributional representations calculated from unlabelled tweets. We also develop distant supervision methods for generating synthetic training data for twitter polarity classification by exploiting unlabelled tweets and prior lexical knowledge. Positive and negative training instances are generated by averaging unlabelled tweets annotated according to a given polarity lexicon. We study different mechanisms for selecting the candidate tweets to be averaged. Our experimental results show that the training data generated by the proposed models produce classifiers that perform significantly better than classifiers trained from tweets annotated with emoticons, a popular distant supervision approach for Twitter sentiment analysis.

Lugar y Fecha

Martes 28 de Junio de 2016, 12:00 Hrs.
Laboratorio de Programación Avanzada (B-038), DI Campus San Joaquín. UTFSM
Auditorio Claudio Matamoros (F-106), DI Casa Central, UTFSM (Video-conferencia)