EMNLP 2018

Proceedings of the 2018 EMNLP Workshop SMM4H:

3rd Social Media Mining for Health Applications Workshop & Shared Task

October 31, 2018
Brussels, Belgium

©2018 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-948087-77-3

Preface

Welcome to the 3rd Social Media Mining for Health Applications Workshop and Shared Task - SMM4H.

The total number of users of social media continues to grow worldwide, resulting in the generation of vast amounts of data. With nearly half of adults worldwide and two-thirds of all American adults using social networking, the latest Pew Research Report estimates that 26% of the total users have discussed health information and, of those, 42% have even discussed current medical conditions. Advances in automated NLP and machine learning present the possibility of utilizing this massive data source for biomedical and public health applications, if researchers address the methodological challenges unique to this media.

For its third iteration, the SMM4H workshop takes place in Brussels, Belgium, on November 1, 2018, and is co-located with the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). Following on the success of a session and accompanying Workshop on the topic that was hosted at the Pacific Symposium in Biocomputing (PSB) in 2016 and the AMIA Annual Conference in 2017, this workshop aims to provide a forum for the ACL community members to present and discuss NLP advances specific to social media use in the particularly challenging area of health applications, with a special focus given to automatic methods for the collection, extraction, representation, analysis, and validation of social media data for health informatics.

As for the previous years, the workshop includes shared tasks with a particular interest on social media mining for pharmacovigilance. This third execution of the SMM4H shared tasks comprises four subtasks. These subtasks involve annotated user posts from Twitter (tweets) and focus on the (i) automatic classification of tweets mentioning a drug name, (ii) automatic classification of tweets containing reports of first-person medication intake, (iii) automatic classification of tweets presenting self-reports of adverse drug reaction (ADR) detection, and (iv) automatic classification of vaccine behavior mentions in tweets. A total of 14 teams participated and 78 system runs were submitted. Deep learning-based classifiers were the primary approach, but feature-based classifiers and a few ensemble learning systems were also used.

We received very high quality submissions, and present 19 as long and short talks and posters. The organizing committee would like to thank the program committee, consisting of 13 researchers, for their thoughtful input on the submissions, as well as the organizers of EMNLP for their support and management. Finally, a huge thanks to all authors who submitted a paper for the workshop or participated in the shared tasks; this workshop would not have been possible without them and their hard work.

Graciela, Davy, Abeed, Michael

Organizing Committee:

Graciela Gonzalez-Hernandez, University of Pennsylvania
Davy Weissenbacher, University of Pennsylvania
Abeed Sarker, University of Pennsylvania
Michael Paul, University of Colorado-Boulder
Karen O'Connor, University of Pennsylvania

Program Committee:

Pierre Zweigenbaum, French National Center for Scientific Research (France)
Nigel Collier, University of Cambridge (UK)
Cecile Paris, Commonwealth Scientific and Industrial Research Organisation (Australia)
Azadeh Nikfarjam, Nuance Communications (USA)
Masoud Rouhizadeh, Johns Hopkins (USA)
Ari Z. Klein, University of Pennsylvania (USA)

Additional Reviewers:

Takeshi Onishi, Toyota Technological Institute at Chicago (USA)
Arjun Magee, Arizona State University (USA)
Haitao Cai, University of Pennsylvania (USA)

Table of Contents

Football and Beer - a Social Media Analysis on Twitter in Context of the FIFA Football World Cup 2018
Roland Roller, Philippe Thomas and Sven Schmeier 1

Stance-Taking in Topics Extracted from Vaccine-Related Tweets and Discussion Forum Posts
Maria Skeppstedt, Manfred Stede and Andreas Kerren 5

Identifying Depression on Reddit: The Effect of Training Data
Inna Pirina and Çağrı Çöltekin 9

Overview of the Third Social Media Mining for Health (SMM4H) Shared Tasks at EMNLP 2018
Davy Weissenbacher, Abeed Sarker, Michael J. Paul and Graciela Gonzalez-Hernandez 13

Changes in Psycholinguistic Attributes of Social Media Users Before, During, and After Self-Reported Influenza Symptoms
Lucie Flekova, Vasileios Lampos and Ingemar Cox 17

Thumbs Up and Down: Sentiment Analysis of Medical Online Forums
Victoria Bobicev and Marina Sokolova 22

Identification of Emergency Blood Donation Request on Twitter
Puneet Mathur, Meghna Ayyar, Sahil Chopra, Simra Shahid, Laiba Mehnaz and Rajiv Shah 27

Dealing with Medication Non-Adherence Expressions in Twitter
Takeshi Onishi, Davy Weissenbacher, Ari Klein, Karen O'Connor and Graciela Gonzalez-Hernandez 32

Detecting Tweets Mentioning Drug Name and Adverse Drug Reaction with Hierarchical Tweet Representation and Multi-Head Self-Attention
Chuhan Wu, Fangzhao Wu, Junxin Liu, Sixing Wu, Yongfeng Huang and Xing Xie 34

Classification of Medication-Related Tweets Using Stacked Bidirectional LSTMs with Context-Aware Attention
Orest Xherija 38

Shot Or Not: Comparison of NLP Approaches for Vaccination Behaviour Detection
Aditya Joshi, Xiang Dai, Sarvnaz Karimi, Ross Sparks, Cecile Paris and C Raina MacIntyre . . . 43

Neural DrugNet
Nishant Nikhil and Shivansh Mundra 48

IRISA at SMM4H 2018: Neural Network and Bagging for Tweet Classification
Anne-Lyse Minard, Christian Raymond and Vincent Claveau 50

Drug-Use Identification from Tweets with Word and Character N-Grams
Çağrı Çöltekin and Taraka Rama 52

Automatic Identification of Drugs and Adverse Drug Reaction Related Tweets
Segun Taofeek Aroyehun and Alexander Gelbukh 54

UZH@SMM4H: System Descriptions
Tilia Ellendorff, Joseph Cornelius, Heath Gordon, Nicola Colic and Fabio Rinaldi 56

Deep Learning for Social Media Health Text Classification
Santosh Tokala, Vaibhav Gambhir and Animesh Mukherjee 61

Using PPM for Health Related Text Detection
Victoria Bobicev, Victoria Lazu and Daniela Istrati 65

Leveraging Web Based Evidence Gathering for Drug Information Identification from Tweets
Rupsa Saha, Abir Naskar, Tirthankar Dasgupta and Lipika Dey 67

CLaC at SMM4H Task 1, 2, and 4
Parsa Bagherzadeh, Nadia Sheikh and Sabine Bergler 70

Conference Program

Wednesday, October 31, 2018

9:00–9:10 *Introduction*
Graciela Gonzalez

9:10–9:30 *Football and Beer - a Social Media Analysis on Twitter in Context of the FIFA Football World Cup 2018*
Roland Roller, Philippe Thomas and Sven Schmeier

9:30–9:50 *Stance-Taking in Topics Extracted from Vaccine-Related Tweets and Discussion Forum Posts*
Maria Skeppstedt, Manfred Stede and Andreas Kerren

9:50–10:10 *Identifying Depression on Reddit: The Effect of Training Data*
Inna Pirina and Çağrı Çöltekin

10:30–11:00 ***Tea Break and Poster Session***

11:00–11:20 *Overview of the Third Social Media Mining for Health (SMM4H) Shared Tasks at EMNLP 2018*
Davy Weissenbacher, Abeed Sarker, Michael J. Paul and Graciela Gonzalez-Hernandez

11:20–11:40 *Changes in Psycholinguistic Attributes of Social Media Users Before, During, and After Self-Reported Influenza Symptoms*
Lucie Flekova, Vasileios Lampos and Ingemar Cox

11:40–12:00 *Thumbs Up and Down: Sentiment Analysis of Medical Online Forums*
Victoria Bobicev and Marina Sokolova

12:00–12:20 *Identification of Emergency Blood Donation Request on Twitter*
Puneet Mathur, Meghna Ayyar, Sahil Chopra, Simra Shahid, Laiba Mehnaz and Rajiv Shah

12:20–12:30 *Dealing with Medication Non-Adherence Expressions in Twitter*
Takeshi Onishi, Davy Weissenbacher, Ari Klein, Karen O'Connor and Graciela Gonzalez-Hernandez

12:30–14:00 ***Lunch***

14:00–14:20 *Detecting Tweets Mentioning Drug Name and Adverse Drug Reaction with Hierarchical Tweet Representation and Multi-Head Self-Attention*
Chuhan Wu, Fangzhao Wu, Junxin Liu, Sixing Wu, Yongfeng Huang and Xing Xie

Wednesday, October 31, 2018 (continued)

14:20–14:40 *Classification of Medication-Related Tweets Using Stacked Bidirectional LSTMs with Context-Aware Attention*
Orest Xherija

14:40–15:00 *Shot Or Not: Comparison of NLP Approaches for Vaccination Behaviour Detection*
Aditya Joshi, Xiang Dai, Sarvnaz Karimi, Ross Sparks, Cecile Paris and C Raina MacIntyre

15:00–15:10 *Neural DrugNet*
Nishant Nikhil and Shivansh Mundra

15:10–15:20 *IRISA at SMM4H 2018: Neural Network and Bagging for Tweet Classification*
Anne-Lyse Minard, Christian Raymond and Vincent Claveau

15:20–15:30 *Drug-Use Identification from Tweets with Word and Character N-Grams*
Çağrı Çöltekin and Taraka Rama

15:30–16:00 ***Tea Break and Poster Session***

16:00–16:10 *Automatic Identification of Drugs and Adverse Drug Reaction Related Tweets*
Segun Taofeek Aroyehun and Alexander Gelbukh

16:10–16:20 *UZH@SMM4H: System Descriptions*
Tilia Ellendorff, Joseph Cornelius, Heath Gordon, Nicola Colic and Fabio Rinaldi

16:20–16:30 *Deep Learning for Social Media Health Text Classification*
Santosh Tokala, Vaibhav Gambhir and Animesh Mukherjee

16:30–16:40 *Using PPM for Health Related Text Detection*
Victoria Bobicev, Victoria Lazu and Daniela Istrati

16:40–16:50 *Leveraging Web Based Evidence Gathering for Drug Information Identification from Tweets*
Rupsa Saha, Abir Naskar, Tirthankar Dasgupta and Lipika Dey

16:50–17:00 *CLaC at SMM4H Task 1, 2, and 4*
Parsa Bagherzadeh, Nadia Sheikh and Sabine Bergler