

Identification of Microblogs Prominent Users during Events by Learning Temporal Sequences of Features

Imen Bizid
L3i Lab
La Rochelle, France
imen.bizid@univ-lr.fr

Nibal Nayef
L3i Lab
La Rochelle, France
nibal.nayef@univ-lr.fr

Patrice Boursier
IUMW
Kuala Lumpur, Malaysia
patrice@iumw.edu.my

Sami Faiz
LTSIRS Lab
Tunis, Tunisia
sami.faiz@insat.rnu.tn

Antoine Doucet
L3i Lab
La Rochelle, France
antoine.doucet@univ-lr.fr

ABSTRACT

During specific real-world events, some users of microblogging platforms could provide exclusive information about those events. The identification of such prominent users depends on several factors such as the freshness and the relevance of their shared information. This work proposes a probabilistic model for the identification of prominent users in microblogs during specific events. The model is based on learning and classifying user behavior over time using Mixture of Gaussians Hidden Markov Models. A user is characterized by a temporal sequence of feature vectors describing his activities. The features computed at each time-stamp are designed to reflect both the on- and off-topic activities of users. To validate the efficacy of our proposed model, we have conducted experiments on data collected from Twitter during the Hérault floods that have occurred in France. The achieved results show that learning the time-series of users' actions is better than learning just those actions without temporal information.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*selection process, retrieval models*;
H.3.1 [Information Systems]: User/Machine Systems—*human factors, human information processing*

Keywords

Prominent microblogs users identification, Learning temporal user behavior, MoG-HMM classification

1. INTRODUCTION

Microblogging platforms represent a rich source of information indispensable to manage specific events. These plat-

forms are seen as the perfect ground to mine relevant and exclusive information during such events. Although these microblogs such as Twitter provide many characterizing features about their content – such as the number of retweets or favorites etc. –, such features generally refer to popular users content rather than reflecting the real importance of that content. Therefore, in the context of a specific event – such as a disaster –, it is more logical to associate the relevance and quality of messages with the user's prominence in that specific event [11, 8, 15].

We define *prominent user identification* in microblogging platforms as identifying key users who provide fresh and relevant information during specific events regardless of their popularity in the platform. To the best of our knowledge, this task has never been studied in this context. However, the related problems of finding influencers have been widely explored in the state-of-the-art [10, 4]. Other works have dealt with the identification of topical authorities and domain experts [9, 13, 6]. In these related contexts, key users were identified by analyzing either the network structure using time consuming centrality algorithms such as PageRank [7, 12, 14] and HITS [1], or using the historic topical users activities independently of their temporal characteristics and off-topic activity [10, 13].

These general purpose existing approaches would give a misleading image of users behavior in real scenarios. Users are mainly evaluated according to their on-topic activity while neglecting their off-topic one. Such practice would promote official media channels toggling between several topics and which are not necessarily sharing fresh information. Moreover, users are typically represented in terms of the quantity of their produced and forwarded information independently of the temporal distribution of this information. This would give the same description for users interacting at an early stage of the event by sharing fresh information and other users posting the same information at its end.

This work is thus designed to alleviate these shortcomings. More specifically, we present the following contributions: (1) a novel representation of microblogs user behavior as a temporal sequence of features that characterize both the on- and off-topic user activities, (2) a probabilistic model for the identification of prominent users in microblogs during specific events. This model is based on learning and classifying the mentioned representation using a Mixture of Gaussians Hidden Markov Model (MoG-HMM).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CIKM'15 October 19–23, 2015, Melbourne, VIC, Australia
Copyright 2015 ACM 978-1-4503-3794-6/15/10 ...\$15.00
DOI: <http://dx.doi.org/10.1145/2806416.2806612>.