Time-Critical Analysis of Evolving Social Media Streams During Sudden-Onset Events

Muhammad Imran
Qatar Computing Research Institute (HBKU), Doha, Qatar.
mimran@hbku.edu.qa

Abstract. Sudden-onset emergencies such as natural or man-made disasters bring uncertainties in which time-critical information needs emerge from formal response organizations, affected communities and other concerned population [2]. The growing adaption of Information and Communication Technologies (ICT) and Social Networks such as Twitter, Facebook has created numerous opportunities to disseminate and consume critical information during an on-going situation. However, time-critical analysis of high-velocity social media streams containing high-volume data involves solving multiple challenges including real-time parsing of brief and informal messages, handling information overload issues, and classifying [3], summarizing [10, 9], and prioritizing different types of information.

Time-Critical Analysis of Live Streams: Live data analysis (online processing) is usually performed over a stream of data relevant to the event, often provided in real-time or with a short delay. The trade-off between retrospective and real-time data analysis is a matter of accuracy versus latency. However, in time-critical situations, the urgency with which the output of an analysis is required remains high, thus online algorithms using stream processing architectures can play a vital role in supporting real-time decision-making [3, 8].

Crowdsourced Stream Processing (CSP): A significant drawback of traditional stream processing systems is that they rely entirely on automated algorithms of data processing, and as such they are limited by the processing capabilities of these algorithms. Indeed, it is often the case that data stream to be processed is imprecise, highly variable and previously unseen, yet the decision based on this data needs to be made in real-time. In this case, automated processing on its own may produce undesirable results, harming the decision-making process it needs to support. CSP systems that combine high-speed processing with human intelligence can produce better and accurate results as compared to traditional stream processing techniques [5].

Handling Evolving Social Media Streams: An emerging challenge in the online classification of social media data streams is to keep the categories used for classification up-to-date [1]. Identification of novel concepts in an online supervised classification setting demands iteration over past observed data items, while due to the infinite length property of data streams, it is impractical to store past data. Approaches employing unsupervised techniques and human-experts could potentially be useful [4].

Domain Adaptation and Transfer Learning: Scarcity of labeled data at the onset of a crisis delays machine learning processes. Past events labeled data (e.g., [6]) can be employed to train more robust machine learning models using

domain adaptation (for same/similar tasks) or transfer learning (for novel tasks) techniques [7].

Brief Biography. Muhammad Imran is a Research Scientist at the Qatar Computing Research Institute (QCRI) where he leads the Crisis Computing group from both science and engineering directions. His interdisciplinary research focuses on Human-Computer Interaction (HCI), real-time Information Retrieval (IR), and Stream Processing areas. He studies the user-generated content on Social Networks (e.g. Twitter) during mass convergence events to develop novel computational models, techniques, and technologies helpful for stakeholders to gain situational awareness and actionable information.

Dr. Imran has published over 40 research papers in top-tier international conferences and journals. Two of his papers have received the Best Paper Award. He is a Co-Chair of the Social Media Studies track of the ISCRAM international conference (2014-2017) and has served as Program Committee (PC) of many major conferences and workshops (ICWSM, ACM DH, ICWE, SWDM, etc.). Dr. Imran has worked as a Post-Doctoral researcher at QCRI (2013-2015). He received his PhD in Computer Science from the University of Trento (2013).

References

1. M. Imran and C. Castillo. Towards a data-driven approach to identify crisis-related topics in social media streams. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1205–1210. ACM, 2015.
2. M. Imran, C. Castillo, F. Diaz, and S. Vieweg. Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)*, 47(4):67, 2015.
3. M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg. AIDR: Artificial intelligence for disaster response. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 159–162. ACM, 2014.
4. M. Imran, S. Chawla, and C. Castillo. A robust framework for classifying evolving document streams in an expert-machine-crowd setting. In *Proceedings of the 18th International Conference on Data Mining (ICDM)*, 2016.
5. M. Imran, I. Lykourentzou, Y. Naudet, and C. Castillo. Engineering crowdsourced stream processing systems. *arXiv preprint arXiv:1310.5463*, 2013.
6. M. Imran, P. Mitra, and C. Castillo. Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, may 2016.
7. M. Imran, P. Mitra, and J. Srivastava. Cross-language domain adaptation for classifying crisis-related short messages. In *Proceedings of ISCRAM*, 2016.
8. L. Neumeyer, B. Robbins, A. Nair, and A. Kesari. S4: Distributed stream computing platform. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 170–177. IEEE, 2010.
9. K. Rudra, S. Banerjee, N. Ganguly, P. Goyal, M. Imran, and P. Mitra. Summarizing situational tweets in crisis scenario. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, pages 137–147. ACM, 2016.
10. K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, and S. Ghosh. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 583–592. ACM, 2015.