An Approach for Identifying Crisis on Social Media

[1]Farooq Sunar Mahammad, [2] V. Uday Kumar,[3] Y. Tharun Kumar Reddy, [4] J. Sri Hari,[5]S. Viswanath Goud
1 Associate Professor, Department of CSE, Santhiram Engineering College, Nandyal
[2,3,4,5] IV.B.Tech CSE, Santhiram Engineering College, Nandyal

ABSTRACT: Today in all over the world using social media to varies purpose. Some people are post their situation in real time. Posts are different types in this paper mainly focused on crisis. Crisis are having some important in real-time. Hence, we proposed a novel that takes only the Crisis from the social media. AOMPC algorithm that takes the data from the social media and that divides data into irrelevant and relevant data based on the Crisis words. AOMPC algorithm performs on labeled and unlabeled data. AOMPC takes the two types of data (1) Synthetic data and (2) Social Media data. The experiments showed very good behavior of AOMPC for dealing with evolving, partly-labeled data streams.

Key Words: AOMPC, Crisis, Synthetic data, Social Media.

1. Introduction: The primary task is to identify the Crisis in real time that to prevent, identify and feedback for the Crisis. In order to execute these tasks efficiently, it is helpful to use data from various sources including the public as witnesses of emergency events. Such data would enable emergency operations centers to act and organize the rescue and response. In recent years, a number of research studies have investigated the use of social media as a source of information for efficient crisis management. A selection of such studies, among others, corona virus, encompasses Norway Attacks, Minneapolis Bridge Collapse, California Wild fire, Colorado Floods, and Australia Bush fires. In real-time several departments like (e.g. Police, cyber-crime) are using social media data for monitoring, gathering and unknown information from the public it uses. Hence we propose a learning algorithm, AOMPC that can illustrate about the crisis on the social media data. For these we are using a classification method. The classifier plays the role of filtering machinery. We propose a Learning Vector Quantization (LVQ) - like approach based on multiple prototype classification. The classifier operates online to deal with the evolving stream of data. The algorithm, named active online multiple prototype classifier (AOMPC), uses un- labeled and labeled data which are tagged through active learning. Data items which fall into ambiguous regions are selected for labeling by the user. The number of queries is controlled by a budget. The requested items help to direct the AOMPC classifier to a better discriminatory capability. While AOMPC can be applied to any streaming data, here we consider in particular SM data.

2. Literature Survey:

2.1 Data Mining and Machine Learning

Together with a suitable, discriminative distance or dissimilarity measure, prototypes can be used for the classification of complex, possibly high-dimensional data. Most frequently, standard Euclidean distance is employed as a distance measure. We discuss how LVQ can be equipped with more general dissimilarities. Moreover, we introduce relevance learning as a tool for the data-driven optimization of parameterized distances. Prototype-based models constitute a very successful family of methodological approaches in machine learning [1].They are appealing for a number of reasons: The extraction of information from previously observed data in terms of typical representatives, so called prototypes, is particularly transparent and intuitive, in contrast to many, more black-box like systems. The same is true for the working phase, in which novel data are compared with the prototypes by use of a suitable dis-similarity or distance measure. Prototype systems are frequently employed for the unsupervised analysis of complex data sets, aiming at the detection of underlying structures, such as clusters or hierarchical relations [2]. Competitive Vector Quantization or the well-known K-means algorithm is prominent examples for the use of prototypes in the context of unsupervised learning Potential goals of supervised machine learning are the assignment of data to categories in classification problems, or their characterization by a continuous target value in regression tasks. In both cases, the learning or training process relies on the availability of labeled example data. The aim is to extract relevant information and represent it in terms of a hypothesis for the unknown target function. The obtained hypothesis can then be applied to novel data in a working phase.

2.2 Multiple Prototype Classification and LVQ Classification

A prototype-based classification approach operates on data items mapped to a vector representation (e.g., vector space model for text data). Data points are classified via prototypes considering similarity measures. Prototypes are adapted based on items related/similar to them. A Rocchio classifier [3] is an example of a single prototype-based classifier. It distinguishes between two classes, e.g., "relevant" and "irrelevant". In real world-scenarios, due to the nature of the data, it is often not possible to describe the data with a single prototype-based classifier. Multiple prototype classifiers (i.e., several prototypes) are needed.Self-organizing maps (SOM) introduced by Kohonen [4] are an unsupervised version of prototype based classification, also known as LVQ. In this case, prototypes are initialized (e.g., randomized) and adapted. SOM was also used for SM analysis in the context of crisis management to identify important hotspots [5].LVQ has been applied to several areas, e.g., robotics, pattern recognition, image processing, text classification etc. [6], [4], [7]. LVQ - in the context of similarity representation, rather than vector-based representation - is analyzed by Hammer et al. [8]. Mokbel et al. [9] describe an approach to learn metrics for different LVQ classification tasks. They suggest a metric adaptation strategy to automatically adapt metric parameters. Bezdek et al. [10] review several offline multiple prototype classifiers, e.g., LVQ, fuzzy LVQ, and the deterministic Dog-Rabbit(DR) model. The latter limits the movement of prototypes and is similar to our approach. However, in contrast to our approach, DR uses offline adaptation of the learning rate. The time based learning rate of our algorithm considers concept drift (i.e., changes of the incoming data) directly during the update of the prototypes. In contrast to the previous approaches, Bouchachia [11] proposes an incremental supervised LVQ-like competitive algorithm that operates online. It consists of two stages. In the first stage (learning stage), the notions of winner reinforcement and rival repulsion are applied to update the weights of the prototypes. In the second stage (control stage), two mechanisms, staleness and dispersion are used to get rid of dead and redundant prototypes. A summary of different prototype based learning approaches can be found in Biehl et al. [12]. In this study, we deal with online real-time classification and we propose a multi-prototype quantization algorithm, where the winning prototype is adapted based on the input. In particular, the algorithm relies on online learning and active learning.

2.3 Social Media Analysis for Crisis Management

Recent research studies SM from several technical perspectives. Due to space limitations, we describe existing SM analysis frame works mostly in the context of crisis management, although there are several frameworks in other contexts, e.g., Twitterbeat [13] andHarVis [14]. Backfried et al. [15] describe an analysis approach based on visual analytics for combining information from different sources with a specific focus on multilingual issues. Vieweg and Hodges [16], [17] describe the Artificial Intelligence for Disaster Response (AIDR) platform, where persons annotate incoming tweets (similar to Amazon Mechanical Turk). The tweets are then used to train classifiers to identify more relevant tweets. AIDR allows classifying incoming tweets based on different information categories, e.g., damage report, casualties, advising, etc. Chen et al. [18] analysis tweets related to Flu to identify topics for predicting the Flu-peak. Neppalli et al. [19] perform sentiment analysis based on social media related to Hurricane Sandy. The work shows that sentiment of users is related to the distance of the Hurricane to the users. Twitcident described by Abel et al. [20] is a framework to search and filter Twitter messages through specific profiles (e.g., keywords). Terpstra et al. [21] show the usage of Twitcident in crisis management. Tweak-the-Tweet introduced by Starbirdetal. [22] Defines a grammar which can be easily integrated in tweets and therefore automatically parsed. Also, TEDAS described by Li et al. [23] is a system to detect high-level events (e.g., all car accidents in a certain time period) using spatial and temporal information. Yin et al. [24], [25] design a situational awareness platform for SM. Tweets are analyzed based on bursty keywords to identify emergent incidents. Ragini et al. [26] combine several techniques to identify people in danger. They examined rule based classification and several machine learning approaches, like SVM, for hybrid classification. Additional information on social media analysis in different crises can be found in Reuter and Kaufhold [27]. Due to the importance of SM, it is our aim to support emergency management when using the content of SM platforms. Currently, there are systems with crowd-sourcing platform characteristics, but no procedure (like active learning) is available to directly involve emergency management personnel in filtering relevant information.

3. Proposed System:

It is interesting to note that people from emergency departments (e.g., police forces) already use SM to gather, monitor, and to disseminate information to inform the public. Hence, we propose a learning algorithm, AOMPC that relies on active learning to accommodate the user's feedback upon querying the item being processed. Since AOMPC is a classifier, the query is related to labeling that item. The purpose of the project is the purpose of CBCF with the IPU model is to improve recommendation performance such as precision, recall, and F1 score by carefully exploiting different preferences among users. Specially, we formulate a constrained optimization problem in which we aim to maximize the recall (or equivalently F1 score) for a given precision. To this end, users are divided into several clusters based on the actual rating data and Pearson correlation coefficient. Afterward, we give each item an incentive/penalty according to the preference tendency by users within the same cluster. Our experimental results show a significant performance improvement over the baseline CF scheme without clustering in terms of recall or F1 score for a given precision.

4. Results and Discussion

The automation has been done of the proposed system and the results are like discussed below

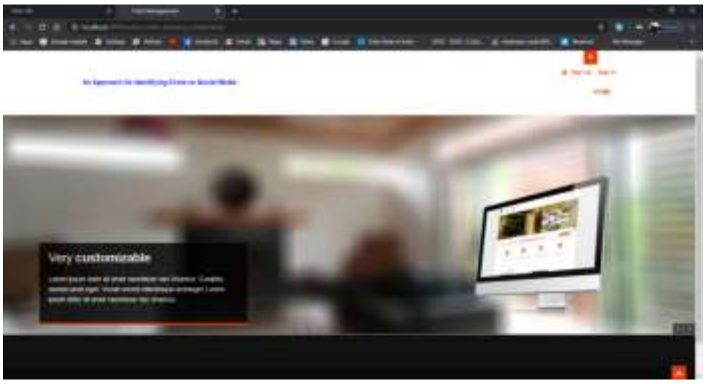

Fig1: Home Page

Above fig is homepage. It has the sign in and signup options. If you want sign in enter the credentials like email and password and phone number and click on the login button. It will creates the account if you want to login first the administrator can accept the user then only can login it can be shown in below fig.

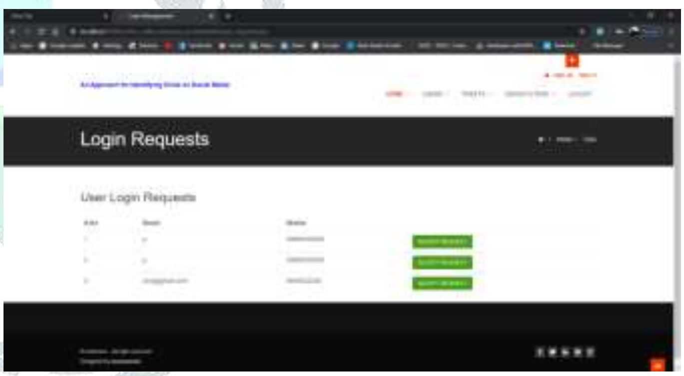

Fig 2. Accepting the users

Fig 3. AOPMC Graph & Tweets

Above fig is main important because it decides the crisis in the application. AOMPC graph can shows the increment of the crisis in real environment. It shows who the user that tweets on that topic is. Users are tweets on the trending crisis. It is also there in social media but the data is noisy. The advantage of AOMPC compared to the other algorithms is the continuous processing of data streams and incremental update of knowledge, where the existing prototypes act as memory for the future.

5. Conclusion

Hence we proposed this novel for managing the Crisis in real environment from the social media and also our framework. In these AOMPC divides the data input to relevant and irrelevant data. AOMPC is Budget less because it uses the fewer quires for data input systems.

6. Future Enhancement

In future may be tweets can give the lifespan of the tweets and delete that automatically. And also try to reduce the budget and work on automatically identify the real tweets and fake tweets. With all this work we can improve recognizing the crisis tweets in a short

References

[1] Schneider, Petra, Michael Biehl, and Barbara Hammer. "Adaptive relevance matrices in learning vector quantization." Neural computation 21.12 (2009): 3532-3561.

[2] Cruz-Vega, Israel, and Hugo Jair Escalante. "An online and incremental GRLVQ algorithm for prototype generation based on granular computing." Soft Computing 21.14 (2017): 3931-3944.

[3] Manning, Christopher D., PrabhakarRaghavan, and HinrichSchütze. Introduction to information retrieval.Cambridge university press, 2008.

[4] Kohonen, Teuvo. "The self-organizing map." Proceedings of the IEEE 78.9 (1990): 1464-1480.

[5] Pohl, Daniela, AbdelhamidBouchachia, and Hermann Hellwagner. "Social media for crisis management: clustering approaches for sub-event detection." Multimedia tools and applications 74.11 (2015): 3901-3932.

[6] Denecke, Alexander, et al. "Online figure–ground segmentation with adaptive metrics in generalized LVQ." Neurocomputing 72.7-9 (2009): 1470-1482.

[7] Azara, Mohammed, Tamer Fatayer, and Alaa El-Halees. "Arabie text classification using Learning Vector Quantization."2012 8th International Conference on Informatics and Systems (INFOS).IEEE, 2012.

[8]Hammer, Barbara, et al. "Learning vector quantization for (dis-) similarities."Neurocomputing 131 (2014): 43-51.

[9]Mokbel, Bassam, et al. "Metric learning for sequences in relational LVQ." Neurocomputing 169 (2015): 306-322.

[10]Bezdek, James C., et al. "Multiple-prototype classifier design." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 28.1 (1998): 67-79.

[11]A.Bouchachia,"LearningwithIncrementality,"inProc.oftheInt'l Conf. on Neural Information Processing, 2006, pp. 137–146.

[12]Biehl, Michael, Barbara Hammer, and Thomas Villmann. "Prototype-based models in machine learning." Wiley Interdisciplinary Reviews: Cognitive Science 7.2 (2016): 92-111.

[13]E. Shook, K. Leetaru, G. Cao, A. Padmanabhan, and S. Wang, "Happy or Not: Generating Topic-based Emotional Heatmaps for Culturomics using CyberGIS," in 2012 IEEE 8th Int'l Conf. on E-Science, oct. 2012, pp. 1 –6.

[14]U. Ahmad, A. Zahid, M. Shoaib, and A. AlAmri, "Harvis: An integrated social media content analysis framework for youtube platform," Information Systems, vol. 69, pp. 25 – 39, 2017.

[15]G. Backfried, J. Gollner, G. Qirchmayr, K. Rainer, G. Kienast, G. Thallinger, C. Schmidt, and A. Peer, "Integration of Media SourcesforSituationAnalysisintheDifferentPhasesofDisaster Management: The QuOIMA Project," in Eur. Intel. and Security Informatics Conf., Aug 2013, pp. 143–146.

[16]M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, "AIDR: Artificial Intelligence for Disaster Response," in Proc. of the Companion Publication of the 23rd Int'l Conf. on World Wide Web, ser. WWW Companion '14, April 2014, pp. 159–162.

[17]S. Vieweg and A. Hodges, "Rethinking Context: Leveraging Human and Machine Computation in Disaster Response," Computer, vol. 47, no. 4, pp. 22–27, Apr 2014.

[18]L.Chen,K.S.M.TozammelHossain,P.Butler,N.Ramakrishnan, and B. A. Prakash, "Syndromic Surveillance of Flu on Twitter Using Weakly Supervised Temporal Topic Models," Data Mining and Knowledge Discovery, vol. 0, no. 3, pp. 681–710, May 2016.

[19]V. K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, and S. Stehle, "Sentiment Analysis during Hurricane Sandy in Emergency Response," International Journal of Disaster Risk Reduction, vol. 21, pp. 213 – 222, 2017.

[20]F. Abel, C. Hauff, G.-J.Houben, R. Stronkman, andK.Tao,"Semantics+Filtering+Search=Twitcident.ExploringInformationinSocialWebStreams,"inProc.ofthe23rdACMConf.on Hypertext and Social Media. ACM, 2012, pp. 285–294.

[21]T. Terpstra, A. de Vries, R. Stronkman, and G. L. Paradies,"TowardsaRealtimeTwitterAnalysisduringCrisesforOperational Crisis Management," in Proc. of the 9th Int'l ISCRAM Conf., Vancouver, April 2012.

[22]K.StarbirdandJ.Stamberger,"TweaktheTweet:LeveragingMicroblogging Proliferation with a Prescriptive Syntax to Support Citizen Reporting," in Proc. of the 7th Int'l ISCRAM Conf., Seattle, USA, May 2010.

[23]R. Li, K. H. Lei, R. Khadiwala, and K.-C. Chang, "TEDAS: A Twitter-based Event Detection and Analysis System," in IEEE 28th Int'l Conf. on Data Engineering (ICDE), 2012, pp. 1273–1276.

[24]"UsingSocialMediatoEnhanceEmergencySituationAwareness," IEEE Intelligent Sys., vol. 27, no. 6, pp. 52–59, 2012.

[25]J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Emergency Situation Awareness from Twitter for Crisis Management," Int'l Workshop on Social Web for Disaster Management (SWDM), In conjunction with WWW'12, no. 99, p. 1, 2012.

[26]J. R. Ragini, P. R. Anand, and V. Bhaskar, "Mining Crisis Information: A Strategic Approach for Detection of People at Risk through Social Media Analysis," International Journal of Disaster Risk Reduction, vol. 27, pp. 556 – 566, 2018.

[27]C. Reuter and M. Kaufhold, "Fifteen Years of Social Media in Emergencies: A Retrospective Review and Future Directions for Crisis Informatics," Journal of Contingencies and Crisis Management, vol. 26, no. 1, pp. 41–57, 2018.