Introducing a Framework for Automatically Differentiating Witness Accounts of Events from Social Media

Marie Truelove, Maria Vasardani, and Stephan Winter

Department of Infrastructure Engineering, The University of Melbourne, Australia; Emails: truelove@student.unimelb.edu.au (M.T.); maria.vasardani@unimelb.edu.au (M.V.); winter@unimelb.edu.au (S.W.)

SUMMARY

Identifying Witnesses of events from social media is an opportunity to crowdsource real-time information to enhance numerous applications including emergency response in a crisis, filtering sources for journalism, and enhancing marketing services. Using a sporting event broadcast live to a proportionally much larger audience, this research demonstrates a significant increase in the number of Witnesses identified posting from the event venue, in comparison to the number identified from geotags alone. This is achieved by considering the text and image content of micro-blogs as additional evidence. This paper also reports progress towards the automatic categorisation of the additional text and image evidence, and modelling and testing this evidence for corroboration or conflict, using Dempster-Shafter Theory of Evidence.

Keywords: Crowdsourcing, Social Media, Witness Accounts, Supervised Machine Learning, Dempster-Shafer Theory of Evidence

INTRODUCTION

Crowdsourcing information about events from social networks such as Twitter is recognised as an opportunity to harvest detailed real-time information, for example enhancing situational awareness for emergency response and management [18] and creating news summaries of large sporting spectacles [19]. However, these opportunities come with many problems to solve, including detecting the fraction of relevant micro-blogs, and assessing the credibility and location of the micro-bloggers who posted them. This research makes unique contributions by proposing a framework towards distinguishing those micro-blogs which are Witness Accounts (WA) of events. WA are defined as those micro-blogs which contain an observation of the event or its effects [17], for example a statement *I see the bushfire smoke!* or an image conveying the same information. The micro-blogger who posted the WA is considered a potential Witness to the event, and it can be inferred they are on-the-ground (OTG) [15], that is they in close proximity to the event [17]. Impact Accounts (IA) are defined for those micro-blogs which do no contain an observation of the event, but from which it can also be inferred that the micro-blogger who posted it is OTG. IA statements may be as explicit as *I'm being evacuated from my home due to the bushfire*. Formally modelling the witnessing fundamentals of observation and spatial relationship separately enables a generic model for a range of event types including unpredicted natural disasters to scheduled events broadcast live from dedicated venues, such as the case study presented in this paper. All micro-bloggers who post observations of the event whether viewed direct from the grandstands or via television are by definition Witnesses. The research in this paper questions whether it is possible to differentiate those Witnesses which are physically at the event from those watching a broadcast. Such differentiation is supported by micro-blogs with geotags, but typically they are present in only a fraction of micro-blogs, for example 1% [1]. This research demonstrates that including the text content and linked images as evidence, the sample of micro-blogs posted from the event location can be increased significantly from those identified by geotags alone. Additionally, this research questions whether text content and linked images can be automatically categorised, and used to test whether they corroborate

the inference they were posted from the event.

In order to automatically differentiate those micro-blogs which are WA or IA, and test the Witness categorisation of the micro-bloggers who posted them, a framework is proposed with the following parts:

1. Machine learning approaches to categorise micro-blogs with text and linked images that are likely WA or IA;
2. Combine the evidence extracted for each individual micro-blog to determine those which can be ranked as containing corroborating or conflicting evidence;
3. For each micro-blogger found to have posted micro-blogs containing evidence, combine these to rank their likely status as a Witness OTG; and
4. For likely Witnesses, seek further evidence, for example from micro-blogging history posted during the event.

This paper presents progress to date on parts 1) and 2). To demonstrate part 1) supervised machine learning approaches are used to categorise the text and image content. A model of the micro-blog text, linked images and geotags using Dempster-Shafer Theory of Evidence [3] is developed to demonstrate part 2). The results indicate a significant improvement on the recognition rate of micro-blogs posted from an event from geotags alone. And where multiple evidence is present for an individual micro-blog their combination does produce intuitive results, including identifying conflict due to GPS error. Enhancements and alternative approaches to those presented in this paper, as with parts 3) and 4) of the framework is the subject of future work.

BACKGROUND

Communication technologies have been described as space-adjusting techniques [14], as they enable events to be witnessed by proportionally much larger audiences than the capacity of the venues in which they are held. In these scenarios, unlike previous case studies such as those in [16], it is not possible to infer a Witness is OTG for the dominating category of observations, that is of the play on the field [19]. It has also been determined that the live broadcast delay of approximately 12 seconds cannot be detected in micro-blogs, ruling this feature out as a method to distinguish those witnessing via a broadcast [19] . In addition to sport, differentiating Witnesses of crisis events has gained much interest from researchers. A journalistic approach describes extracting observation features from text to identify Witnesses [2], whereas spatial presence in the city of the event is the criterion in other work [10].

Supervised Machine Learning for Categorisation

Natural language processing (NLP) using bag-of-words approaches from unigram, bigram and parts-of-speech (POS) models, can be utilised as baseline text categorisation features [10] [18]. These research report success, comparable in many scenarios to more sophisticated features [10] [18]. A visual bag-of-words approach to categorise images linked to micro-blogs has also been tested [9]. The disadvantage of bag-of-words approaches is that although the methodology can be applied generically, the resulting model is not generic, for example, a model developed from training data for a football game cannot be used for a bushfire. Approaches which extract semantic meaning, for example locative expressions from text [7] would enable a generic model, but their success to-date is limited in domains such as social media [7]. Detecting micro-blogs posted from OTG is also recognised as a unbalanced class problem [15]. Approaches taken to mitigate class imbalance typically involve balancing the data via sampling [10] [5] [18], or algorithmically introducing a miss-classification cost to the under-represented class [15].

Dempster-Shafer Theory of Evidence

Dempster-Shafters Theory of Evidence is one method that has found application in classifier fusion, and managing uncertainty and incomplete reasoning [3]. The theory models the power set for the frame of discernment of the hypothesis [3]. A mass function is assigned for each subset in the power set from which the belief interval can be derived [3]. The mass function can be assigned from various classifier results, including the overall accuracy, class statistics or individual instances [12]. The mass functions for independent evidence can then be combined [3]. Dempster's Rule of Combination has been shown to produce uninituitve results in scenarios with conflict [13], resulting in many enhancements being proposed including PCR6 [13] based on proportional conflict resolution.

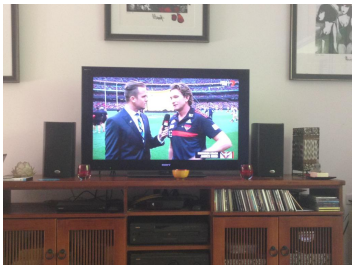
1 METHODOLOGY

Data Collection and Training Set Creation

The case study event is an Australian Football League (AFL) match played at the Melbourne Cricket Ground (MCG) on the annual ANZAC Day public holiday. In 2015, this match attracted a near capacity crowd of 88,398[1] and television ratings of 1.298 million[2]. The corpus was collected using the AFL's promoted hashtag #afldonspies, utilising the Twitter Data Analystics software packages [6]. Pre-processing samples the micro-blogs to those which can be identified as individual and original, that is not a retweet or posted by a non-individual such as the media [17]. To collect a sample of linked images, all micro-blogs in the corpus with a URL to Twitter or Instagram were inspected as these are more likely to contain WA [16]. To create the training set, two expert annotators coded the tweet text and linked images with one of three categories, examples for which are presented in Table 1. The three categories are:

1. No Evidence (NE) when no evidence of being posted from OTG or another place could be detected.
2. When evidence is detected, it is categorised as either evidence posted from OTG (E-OTG);
3. Or counter-evidence indicating that it is not posted from OTG (E-NOTG).

Table 1. Example text and image content for each category. (Source: twitter.com, access date: 25-26/04/15.)

No Evidence (NE)	Evidence OTG (E-OTG)	Evidence not OTG (E-NOTG)
Fletcher goes bang with a 60 metre monster! #AFLDonsPies 	*Not the best seats in the house but just glad to be here at @MCG #AFLDonsPies* 	*In front of TV with chips for next 3 hours! #AFLDonsPies* 

Supervised Machine Learning for Categorisation of Text and Images

Pre-processing of the text included word tokenisation and parts-of-speech tagging using Ark NLP [11]. WEKA's [4] default pre-processing filters were used to experiment with unigram and bigram models. WEKA default feature selection filters are utilised to reduce the number of redundant dimensions, and experiment with a range of classifiers indicated by previous research including Naive Bayes, Random Forest and Support Vector Machines (SVM). All experiments were completed with 10-fold cross validation. As expected, class imbalance was an issue in particular for the text corpus. Sampling to micro-blogs posted by micro-bloggers with at least one piece of evidence detected in the the training set was used to mitigate the imbalance. The classifier selected was that which maximises precision of E-OTG and E-NOTG classes, at the expense of recall if necessary, to minimise conflict due to miss-classification in the Dempster-Shafer modeling.

Categorisation of Geotags

Geotag evidence was cateogorised as E-OTG or E-NOTG based on whether it was contained within or in the immediate vicinity of the MCG, the place of the event. It was necessary to create a decision boundary for this categorisation, which was informed primarily by the boundaries of places bordering the MCG, for example train lines, roads and other venues.

[1] twitter.com/MCG/status/591859347891748865
[2] http://footyindustry.com/files/afl/media/tvratings/2015/2015AFLRatings.png

Dempster-Shafer Modelling of Evidence Extracted from Micro-blogs

The frame of discernment is modelled as {E-OTG, E-NOTG} with power set (null, E-OTG, E-NOTG, {E-OTG, E-NOTG}). The categorisation of NE is not modelled in the frame of discernment. For example, if a micro-blog has a geotag categorised as E-OTG, and text and image categorised as NE, the text and image do not corroborate or produce conflict with the E-OTG categorisation provided by the geotag. For demonstraton mass functions are set manually, with derivation from classifier results left to future work. The mass functions assigned to geotags represent greater certainty than that assigned for images, which are greater than that assigned for text. The combination rule PCR6 implemented in Matlab [8] is then used to compute the combinations for analysis, and again, decision algorithm testing left for future work.

2 RESULTS AND DISCUSSION

The corpus contained 3260 micro-blogs, 265 with linked images and 133 with geotags. Table 2 presents the categorisation results, both training and predicted by classifiers. The annotator agreement for the text and image content was high with Cohen's Kappa of 0.895 and 0.929 respectively. Combining the three content sources, or evidence, from the training data indicates the number of micro-blogs categorised as E-OTG and E-NOTG can be increased significantly from those with geotags alone. The increase for E-OTG is from 21 to 176 micro-blogs, and the increase for E-NOTG is from 112 to 241 micro-blogs. This corresponds to an additional 125 potential Witnesses OTG from 16. 54 tweets had more than one piece of evidence which could be checked for conflict. Conflict did exist for a fraction of tweets, found to be due to GPS error. The geotag indicated the micro-blog was posted from a nearby venue, when the image and text indicated it was posted from the MCG.

The combined results correctly predicted are fewer than the training data, but still a significant increase from those with geotags alone. The number of micro-blogs categorised as E-OTG increased from 21 to 125, and the number of micro-blogs categorised as E-NOTG increased from 112 to 182. This corresponded to an additional 77 potential Witnesses posting from OTG and an additional 50 potential Witnesses via the broadcast. From the predicted results of the classifiers, 26 micro-blogs had more than one piece of evidence, with five in conflict. In addition to GPS error, these conflicts are now also attributed to miss-classification.

Table 2. Number of micro-blogs categorised for each content individually and in combination. The number of miss-classified micro-blogs in the class are presented in parenthesis.

Content	E-OTG		E-NOTG		NE	
	Training	Predicted	Training	Predicted	Training	Predicted
Geotag	21	-	112	-	-	-
Image	95	95 (11)	26	17 (10)	146	173 (27)
Text	99	34 (10)	129	66 (6)	3032	1088 (143)
Combination	176	125 (15)	241	182 (17)	2328	876 (132)

From the classifier experimentation it was found the WEKA default SVM, feature selection filter, and a unigram model maximised precision of the E-OTG and E-NOTG classes for text content. Using the methodology described by [9] the SVM classifier was additionally selected for the image content. The precision and recall results for each class are presented in Table 3. These results indicate the image categorisation failed for the E-NOTG class, which is attributed to the insufficient number of samples in the training data. For future experiments this category could be removed, or the sample increased from other events, both options are to be tested in future work. In comparison, the E-OTG category proved acceptable for both precision and recall. The better precision for text E-NOTG compared to E-OTG could in part be explained by the topics contained in these micro-blogs were dominated by explicit statements critiquing the television coverage or the medium via which the broadcast was being viewed, enabling a more representative unigram model. In comparison, there was not a dominate topic for E-OTG. More robust feature development based on previously identified witnessing characteristics [16] is being developed in future work. Additionally, the results indicate improvements could be made if the class imbalance were further addressed.

Table 3. Class precision and recall results for adopted classifier.

Corpus	Statistic	E-OTG	E-NOTG	NE
Text	Precision	0.706	0.909	0.869
	Recall	0.242	0.465	0.984
Image	Precision	0.844	0.412	0.844
	Recall	0.896	0.280	0.896

Table 4 presents example mass functions manually assigned for text and image evidence and combination mass functions corresponding to predicted combination results. Manual analysis concludes that these results appear intuitive, for example, when there are multiple evidence present supporting a categorisation, the increased values indicate corroboration. Additionally, when conflict exists the values reflect this, suggesting the conflict redistribution of the PCR6 algorithm is appropriate for the modelled scenario. Hybrid methods for deriving the mass functions which can model the uncertainty of the evidence in addition to the automatic classification results are in progress.

Table 4. Example mass functions assigned or PCR6 combination results for evidence combinations detected in micro-blogs. (- indicates no data.)

Evidence Categorisat.			Mass Function			Comment
Text	Image	Geotag	E-OTG	E-NOTG	E-OTG,E-NOTG	
E-OTG	-	-	**0.70**	0.15	0.15	Assigned Mass Funct.
E-OTG	-	E-OTG	**0.95**	0.04	0.01	
E-OTG	E-OTG	E-OTG	**0.97**	0.02	0.01	
E-OTG	E-OTG	E-NOTG	**0.55**	**0.43**	0.02	GPS error example
E-OTG	-	E-NOTG	**0.35**	**0.64**	0.01	Miss-classified Text
-	E-NOTG	-	0.10	**0.80**	0.10	Assigned Mass Funct.
E-NOTG	E-NOTG	-	0.07	**0.91**	0.02	

3 CONCLUSION

This paper presented progress on a framework to automatically extract WA and IA of events from social media. Baseline supervised machine learning techniques to categorise text and images were demonstrated, enabling micro-blogs posted from OTG or via the broadcast to be identified in signficantly greater numbers than with geotags alone. Additionally, a method based on Dempster-Shafer Theory of Evidence was demonstrated to combine the extracted evidence to test corroboration or conflict in the categorisation of the micro-blogs. Many areas for enhancements are identified, including machine learning approaches that further mitigate class imbalance and enable generic model development. In addition to seeking these enhancements, future work will include modeling the combination of evidence from multiple micro-blogs to identify the status of potential Witnesses.

REFERENCES

[1] CHENG, Z., CAVERLEE, J., AND LEE, K. You are where you tweet: A content-based approach to geo-locating Twitter users. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (2010), ACM, pp. 759–768.

[2] DIAKOPOULOS, N., DE CHOUDHURY, M., AND NAAMAN, M. Finding and assessing social media information sources in the context of journalism. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (2012), pp. 2451–2460.

[3] GARGIULO, F., MAZZARIELLO, C., AND SANSONE, C. *Multiple Classifier Systems: Theory, Application and Tools*. Spinger-Verlag, 2013, ch. 10, pp. 335–378.

[4] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. H. The WEKA data mining software: An update. *SIGKDD Explorations 11*, 1 (2009), 10–18.

[5] KUMAR, S., HU, X., AND LIU, H. A behaviour analytics approach to identifying tweets from crisis regions. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (2014), pp. 255–260.

[6] KUMAR, S., MORSTATTER, F., AND LIU, H. *Twitter Data Analytics*. Springer, 2014.

[7] LIU, F., VASARDANI, M., AND BALDWIN, T. Automatic identification of locative expressions from social media text: A comparative analysis. In *Proceedings of the 4th International Workshop on Location and the Web (LocWeb)* (2014), pp. 9–16.

[8] MARTIN, A. Implementing general belief function framework with a practical codification for low complexity. In *Advances and Application os DSmT for Information Fusion*, F. Smarandache and J. Dezert, Eds., vol. 3. American Press Rehoboth, 2009, pp. 217–273.

[9] MCLEAN, S. Identifying witness account in social media using imagery. Master's thesis, 2015. The University of Melbourne.

[10] MORSTATTER, F., LUBOLD, N., PON-BARRY, H., PFEFFER, J., AND LIU, H. Finding eyewitness tweets during crises. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science* (2014).

[11] OWOPUTI, O., OCONNOR, B., DYER, C., GIMPEL, K., SCHNEIDER, N., AND SMITH, N. A. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of NAACL 2013* (2013).

[12] PARIKH, C. R., PONT, M. J., AND JONES, N. B. Application of Dempster-Shafer theory in condition monitoring applications: a case study. *Pattern Recognition Letters 22* (2001), 777–785.

[13] SMARANDACHE, F., AND DEZERT, J. On the consistency of PCR6 with the averaging rule and its application to probability estimation. In *Proceedings of the 16th International Conference on Information Fusion* (2013), pp. 1119–1126.

[14] SPENCER, J. E., AND THOMAS, W. L. J. *Cultural Geography*. John Wiley & Sons, Inc., 1969.

[15] STARBIRD, K., GRACE, M., AND LEYSIA, P. Learning from the crowd: Collaborative filtering techniques for identifying on-the-ground Twitterers during mass disruptions. In *Proceedings of the 9th International ISCRAM Conference* (2012), pp. 1–10.

[16] TRUELOVE, M., VASARDANI, M., AND WINTER, S. Testing a model of witness accounts in social media. In *Proceedings of the 8th Workshop on Geographic Information Retrieval* (2014), no. 10.

[17] TRUELOVE, M., VASARDANI, M., AND WINTER, S. Towards credibility of micro-blogs: characterising witness accounts. *GeoJournal 80*, 3 (2015), 339–359.

[18] VERMA, S., VIEWEG, S., CORVEY, W. J., PALEN, L., MARTIN, J. H., PALMER, M., SCHRAM, A., AND ANDERSON, K. M. Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (2011), pp. 385–392.

[19] ZHAO, S., ZHONG, L., WICKRAMASURIYA, J., AND VASUDEVAN, V. Human as real-time sensors of social and physical events: A case study of Twitter and sports games. Tech. rep., Rice University and Motorola Labs, 2011.