Integrating Machine Learning into a Crowdsourced Model for Earthquake-Induced Damage Assessment

Umaa Rebbapragada Umaa.D.Rebbapragada@jpl.nasa.gov
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91106

Thomas Oommen toommen@mtu.edu
Dept. of Geological Engineering, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931

Abstract

On January 12th, 2010, a catastrophic 7.0M earthquake devastated the country of Haiti. In the aftermath of an earthquake, it is important to rapidly assess damaged areas in order to mobilize the appropriate resources. The Haiti damage assessment effort introduced a promising model that uses crowdsourcing to map damaged areas in freely available remotely-sensed data. This paper proposes the application of machine learning methods to improve this model. Specifically, we apply work on learning from multiple, imperfect experts to the assessment of volunteer reliability, and propose the use of image segmentation to automate the detection of damaged areas. We wrap both tasks in an active learning framework in order to shift volunteer effort from mapping a full catalog of images to the generation of high-quality training data. We hypothesize that the integration of machine learning into this model improves its reliability, maintains the speed of damage assessment, and allows the model to scale to higher data volumes.

1. Introduction

Earthquakes are a major natural disaster associated with large-scale destruction and high mortality rates. The January 12th, 2010 Haiti Earthquake killed close to a quarter of a million people, left over a million individuals homeless, and leveled approximately 20 percent of the buildings in greater Port-au-prince in less than a minute (Eguchi et al., 2010). The latest events in New Zealand (September 4, 2010 and February 22, 2011) and Japan (March 11, 2011) confirm that both developing and developed countries are equally susceptible to large scale earthquake-induced damages, the impact of which can be greatly reduced with accurate and timely damage assessment.

There are four phases of disaster management: mitigation, preparedness, response, and recovery. The most critical and challenging phase is the response phase (Steinle et al., 2001), which consists of a series of both short-term and long-term measures taken to mitigate the effects of the disaster and reestablish normalcy. The overall cost of a disaster after the event, both in terms of fatalities and economic damage, is minimized when the response phase is quickly and efficiently managed (Kerle & Oppenheimer, 2002). Effective management of the response phase requires accurate and timely damage assessment.

The post-disaster needs assessment (PDNA) effort for Haiti led by the World Bank was the first to use crowdsourcing to map damage in remotely-sensed images. Crowdsourcing is an emerging business model largely enabled by Web 2.0 technologies that outsources tasks to an undefined collection of individuals via an open call (Howe, 2006). For the Haiti earthquake, an open call for volunteers from the geoengineering community was issued for the rapid assessment of damage in remotely sensed imagery. Each volunteer received a block of images, completed mapping it for earthquake-induced damage by the visual comparison of a pre- and post-event image, then received another block, and so on. In less than a week, volunteers mapped close to 30,000 severely-damaged buildings.

Despite its successes, the model used during the Haiti earthquake can be improved. With so many involved, and only one image assigned per volunteer, the reliabil-

Appearing in *ICML 2011 Workshop on Machine Learning for Global Challenges*, Bellevue, WA, USA, 2011.

ity of the mappings varied greatly with each volunteers expertise. Indeed, field analysis later revealed a large number of errors of omission (or false negatives) produced by the volunteer PDNA effort (EC et al., 2010). The volunteers competently mapped sites that were severely damaged or totally destroyed, but were less successful at finding moderate levels of damage.

This paper suggests a new model for crowdsourced PDNA of earthquake-induced damage that uses machine learning to semi-automate the effort and improve the reliability of the damage assessment. In contrast to the Haiti model, our model assigns a single image to multiple volunteers in order to robustly estimate both the damage in the image and the volunteer's reliability. However, the use of multiple volunteers per image can inflate the overall time of the effort. To maintain the speed of the assessment effort, we propose the semi-automation of the effort via an active learning framework which shifts the volunteer effort from mapping and classifying the entire database to generating reliable training data. This paper discusses relevant work in machine learning and research opportunities that arise from adapting existing machine learning methods for this problem setting.

2. Haiti PDNA Model

The use of crowdsourcing for earthquake-induced damage assessment was prototyped in the wake of the 2008 Sichuan, China earthquake. A "highly-motivated" network of scientists participated in a "community-based" damage assessment effort examining private aerial and commercial satellite data (Bevington et al., 2010). This effort spurred the development of the online Virtual Disaster Viewer (VDV) by ImageCat, Inc. for the analysis of pre- and post-event high-resolution imagery.

When the Haiti earthquake happened in 2010, high resolution remotely-sensed images provided comprehensive coverage of Haiti in real-time. The World Bank tasked ImageCat to assess damage in these now freely available images. ImageCat expanded the network of scientists used during the Sichuan earthquake, and formalized it as the Global Earth Observation Catastrophe Assessment Network (GEO-CAN). During the Haiti PDNA effort, GEO-CAN grew to include over 600 participants from academia and private industry worldwide. The dissemination of images via VDV to GEO-CAN reduced the overall time required for image review to less than a week, and proved to be promising model for rapid damage assessment.

Figure 1 shows the workflow of the mapping effort performed in the aftermath of the Haiti earthquake. Each volunteer received a block of images, and was tasked with drawing polygons around earthquake-induced damage by visually inspecting the overlay of pre- and post-event images in VDV. In this model, each image is mapped by only one volunteer and results are accepted as-is.

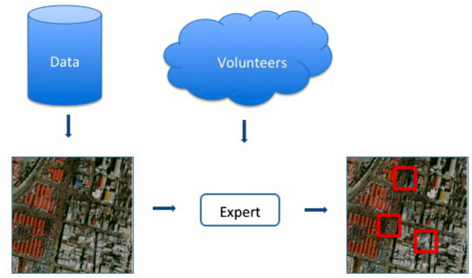

Figure 1. Workflow of Haiti Model

On March 10, 2010 a damage assessment report was released that also included results from field survey. In verifying the results of GEO-CAN's effort, the report cited that the PDNA effort on remotely-sensed data produced a large number of errors of omission (EC et al., 2010). The PDNA effort correctly mapped sites damaged at levels four (very heavy damage) and five (destruction), but were less successful mapping areas that were either moderately or less damaged (classes 1, 2 and 3).

3. Proposed Model

In our opinion, the primary drawback of the Haiti model is that the quality of the effort is subject to the variability of expertise in the crowd. There was no check on the credentials of the 600 volunteers who joined the Haiti effort, or whether they were even familiar with this type of mapping task on remote-sensing data. Even a qualified volunteer may not be familiar with all possible data sources (optical, radar, infrared and microwave). Scoring each volunteer according to their reliability on each data source may improve damage estimates and inform the image dissemination process.

3.1. Volunteer Reliability

Our model proposes the application of recent work on learning from multiple, imperfect experts to robustly estimate an example's label (Sheng et al., 2008), score each volunteer for reliability, or both (Raykar et al., 2009; Donmez et al., 2009; Dekel & Shamir, 2009; Yan et al., 2010). Much of this work is inspired by the advent of low-cost services (e.g., Mechanical Turk, Rent-A-Coder) (Sheng et al., 2008; Donmez et al., 2009) with a focus on tasks related to natural language processing (Donmez et al., 2009), or medical diagnosis (Raykar et al., 2009; Yan et al., 2010).

Each method differs in its approach. Sheng et al. (2008) assumes uniform labeler reliability, and examines the trade-off of getting another label versus another example. Raykar et al. (2009) uses an approach that jointly estimates expert reliability along with an example's label via the Expectation Maximization (EM) algorithm. Yan et al. (2010) extends this work by estimating labeler reliability per example. Donmez et al. (2009) modifies active learning to account for multiple, unreliable labelers and selects only those oracles that score above a certain threshold. Their approach trades off between exploration in the early rounds and exploitation in the later rounds, as the more reliable oracles are identified and targeted for labeling.

3.2. Semi-automation

Algorithm 1 Iterative Framework for Training Data Generation using Multiple Volunteers

1: **Given:** learning algorithm L, labeled images $\mathcal{L}$, unlabeled images $\mathcal{U}$, volunteers $\mathcal{V}$ with weights $\mathcal{W}$
2: $Train(L, \mathcal{L})$
3: **while** true **do**
4: $\quad X \leftarrow Select(\mathcal{U})$
5: $\quad$ **for** objects $x \in X$ **do**
6: $\qquad \hat{l} \leftarrow Label(X, \mathcal{V}, \mathrm{w})$
7: $\qquad \mathcal{L} \leftarrow \mathcal{L} \cup \{(x, \hat{l})\}$
8: $\qquad \mathcal{U} \leftarrow \mathcal{U} - \{x\}$
9: $\quad$ **end for**
10: $\quad$ Update($\mathcal{W}$)
11: $\quad Train(L, \mathcal{L})$
12: **end while**

The use of multiple experts to map a single image increases the magnitude of the effort and risks inflating the overall time of assessment. One can address this problem by recruiting more volunteers. However, it is likely that increasing data rates and availability will outpace the ability to recruit out of the geoengineering communities. In order to use multiple experts per image but maintain the speed of the current model, we propose shifting the volunteer effort to the generation of high-quality training data using an active learning-style framework (Donmez et al., 2009).

Algorithm 1 outlines a generic algorithm which is initialized with a learning algorithm L, a small set of labeled images $\mathcal{L}$, the unlabeled images $\mathcal{U}$, and a set of expert volunteers $\mathcal{V}$ with associated scores $\mathcal{W}$. After training the classifier (line 2), learning proceeds with the selection of images for analysis by $\mathcal{V}$. Typically, active learning selects examples on which the classifier is most uncertain (Lewis & Catlett, 1994). Once an image is selected, all volunteers in $\mathcal{V}$ analyze the pre- and post-event image, identify damaged objects and label them. The predictions of the volunteers in $\mathcal{V}$ are unified according to their reliability scores $\mathcal{W}$ (e.g., weighted majority vote). The unified label $\hat{l}$ is adopted, the training set $\mathcal{L}$ and unlabeled pool $\mathcal{U}$ are updated, L is re-trained, reliability scores $\mathcal{W}$ are updated, and the process repeats until terminated. The approach of (Donmez et al., 2009) best fits our proposed model since it is an active learning framework that uses multiple experts.

3.3. Image Segmentation

Our second time-saving solution is to use image segmentation and change detection (Shapiro, 2001) to find damaged objects (e.g., buildings, roads) in the post-event image, rather than have a volunteer identify them and draw polygons manually. This frees the volunteer to focus on the classification of damage states, and may reduce errors of omission. The challenge of using image segmentation for earthquake-induced damage assessment is that damaged areas tend to expose bare rock and debris, which are easily confused with "rocky outcrops, roads, water bodies, and river beds" (Martha et al., 2010). Image segmentation may not produce objects that represent true damage unless methods are improved to recognize post-earthquake landscape features. However, this challenge enables an interesting research direction: use the volunteer pool to provide feedback on *both* the newly-segmented objects and their classification. This requires integrating the image segmentation process into the iterative learning framework and simultaneously unifying object boundaries and their predictions.

4. Conclusion

This paper presents an alternate model for crowdsourced PDNA that improves reliability by amassing multiple opinions on the damage present in each image, while maintaining time efficiency through the use of machine learning methods for image segmentation and classification. In proposing this model, we have identified two interesting areas of research. The first is the development of image segmentation methods that are robust to earthquake-induced damage, and the second is the integration of expert feedback on both image segmentation and classification into an active learning framework for multiple experts.

We are currently investigating these issues and plan to evaluate them on remotely-sensed data used for mapping damage in the aftermath of the Haiti and New Zealand earthquakes. For both earthquakes we have field survey results that serve as ground truth. We have access to images mapped via GEO-CAN for the Haiti PDNA effort, and plan to simulate crowdsourcing results for the New Zealand earthquake.

Acknowledgments

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2011. Government sponsorship acknowledged.

References

Bevington, J., Adams, B., and Eguchi, R. GEO-CAN debuts to map haiti damage. *Imaging Notes*, 25(2), 2010.

Dekel, O. and Shamir, O. Good learners for evil teachers. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 233–240, 2009.

Donmez, P., Carbonell, J. G., and Schneider, J. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pp. 259–268, New York, NY, USA, 2009.

EC, JRC, UNITAR, UNOSAT, GFDRR, and CNIGS. Building damage assessment report: Haiti earthquake 12 january 2010, 2010. URL http://www.eqclearinghouse.org/20100112-haiti/wp-content/uploads/2010/02/PDNA_damage_assessment_report_v03-1.pdf.

Eguchi, R. T., Gill, S. P., Ghosh, S., Svekla, W., Adams, B. J., Evans, G., Toro, J., Saito, K., and Spence, R. 2010. The january 12, 2010 haiti earthquake: A comprehensive damage assessment using very high resolution areal imagery. In *8th International Workshop on Remote Sensing for Disaster Management*, 2010.

Howe, J. The rise of crowdsourcing. *Wired Magazine*, 14(6), June 2006.

Kerle, N. and Oppenheimer. Satellite remote sensing as a tool in lahar disaster management. *Journal of Disaster Studies Policy and Management*, 26(2): 140–160, 2002.

Lewis, D.D. and Catlett, J. Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of the International Conference on Machine Learning*, pp. 148–156, 1994.

Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., and Kumar, K. V. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology*, 116(1-2):24–36, March 2010.

Raykar, V.C., Yu, S., Zhao, L.H., Jerebko, A., Florin, C., Valadez, G.H., Bogoni, L., and Moy, L. Supervised learning from multiple experts: Whom to trust when everyone lies a bit. In *Proceedings of the 26th International Conference on Machine Learning*, pp. 889–896, 2009.

Shapiro, L. *Computer vision*. Prentice Hall, Upper Saddle River NJ, 2001.

Sheng, V. S., Provost, F., and Ipeirotis, P. G. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 614–622, 2008.

Steinle, E., Kiema, J., and Leebemann, J. 2001. Laserscanning for analysis of damages caused by earthquake hazards. In *Proceedings of the OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models*, Stockholm, 2001.

Yan, Y., Rosales, R., Fung, G., Schmidt, M., Hermosillo, G., Bogoni, L., Moy, L., and Dy, J.G. Modeling annotator expertise: Learning when everybody knows a bit of something. In *International Conference on Artificial Intelligence and Statistics*, pp. 932–939, 2010.