Extracting Spatial Information from Social Media in Support of Agricultural Management Decisions

Nevena Golubovic
UCSB Computer Science
nevena@cs.ucsb.edu

Chandra Krintz
UCSB Computer Science
ckrintz@cs.ucsb.edu

Rich Wolski
UCSB Computer Science
rich@cs.ucsb.edu

Sara Lafia
UCSB Geography
slafia@geog.ucsb.edu

Thomas Hervey
UCSB Geography
thomas.hervey@ucsb.edu

Werner Kuhn
UCSB Geography
werner@geog.ucsb.edu

ABSTRACT

Farmers face pressure to respond to unpredictable weather, the spread of pests, and other variable events on their farms. This paper proposes a framework for data aggregation from diverse sources that extracts named places impacted by events relevant to agricultural practices. Our vision is to couple natural language processing, geocoding, and existing geographic information retrieval techniques to increase the value of already-available data through aggregation, filtering, validation, and notifications, helping farmers make timely and informed decisions with greater ease.

CCS Concepts

•**Information systems** → *Decision support systems;*

Keywords

information retrieval, natural language processing, gazetteer, social media, agricultural analytics

1. INTRODUCTION

Farmers are confronted with many environmental challenges such as difficult weather conditions (i.e. frost, chill hours) and the spread of pests (i.e. insects, invasive weeds) on a daily basis. These challenges require immediate decisions. Farmers increasingly rely on heterogeneous online services, including weather stations and news articles, for environmental information. Techniques are needed to chain these services together and produce a decision-support product with the potential to save farmers time and increase their confidence in decision-making.

We propose a service that aggregates a multitude of data inputs and notifies farmers when events of interest are likely to occur. The proposed system increases farmers' awareness of events, aggregating local spatial information and allowing farmers to make spatially-informed decisions.

GIR 16, October 31-November 03 2016, Burlingame, CA, USA

© 2016 ACM. ISBN 978-1-4503-4588-0/16/10...$15.00

DOI: http://dx.doi.org/10.1145/3003464.3003468

2. ARCHITECTURE

Users provide geographic **areas of interest** (i.e. farm parcel as a shape file) to look up **events of interest** (i.e. frost, pests) by keyword from sources that farmers currently reference, such as news articles and social media. The spatial extents and keywords provided by farmers are used to infer local relevant input sources to include, such as news feeds from neighboring counties that a farmer has not referenced before [3]. When user keywords and source events are matched and their extents intersect or overlap, the system sends an event notification together with all contributing sources, summarized in Figure 1. The following sections provide further details for each processing step.

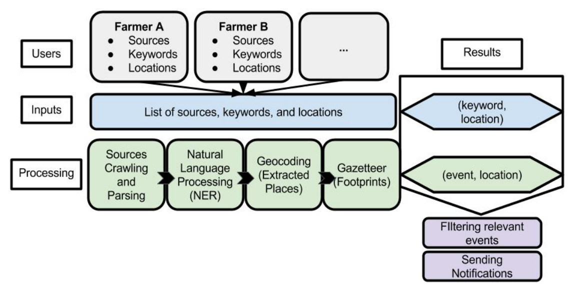

Figure 1: Chaining services to generate spatially informed event notifications from unstructured text.

Natural Language Processing (NLP) – APIs (Twitter[1]), and local news services are periodically called to extract relevant place names and events based on the keywords. Named Entity Recognition (NER) is used to extract toponyms from the given input sources. Services such as GeoTxt[2] couple several NER tools [5].

Geocoding and Footprint Generation – The extracted toponym is referenced against OpenStreetMap[3] to retrieve a spatial identifier, which is then passed to the Overpass API[4] that retrieves the footprint geometry in the form of a polygon.

Semantic Matching – The service uses farmer-specified events of interest to extract only thematically relevant in-

[1]https://dev.twitter.com/rest/public
[2]http://www.geotxt.org/
[3]https://www.openstreetmap.org/
[4]https://overpass-turbo.eu/

formation. The National Agricultural Library Thesaurus[5] provides a controlled vocabulary for categories of thematic terms that map to events of interest. The keywords are enhanced with a service that identifies synonyms, such as Word2Vec[6], discovering synonym strings like *rain* in conjunction with *rainstorm*.

Spatial Filtering – Geometric operations are performed to check for the overlap of the farmer specified areas of interest (i.e. farm footprint parcels) and the area of the news event. Python packages[7], [8] are used to check if the shape returned from the Overpass API intersects with the parcel shape, delineated by the farmer. In the case of an overlap, an event is considered relevant. Additionally, user can specify a search radii for their area of interest, creating *zones of interest*. This enables the generation of a notification about severe rainfall two counties away.

Notification – Users are notified of events in the form of a text message or e-mail. Along with the notification, the user is provided with a list of snippets from tweets, weather station APIs, and other contributing input sources. The end-user can then take action after assessing the severity of the event notification.

3. MOTIVATING USE CASE

After conducting interviews with citrus farmers we discovered that telephone alerts from the Weather Services are among the most commonly used means for monitoring frost. While mostly reliable, there have been incidents where these alerts were received too late and frost settled, reducing crop yield, and other times when they were unnecessary, en-queuing wind machine operation costs. In order to collect additional information to help with operation decisions, farmers must go online and manually check multiple forecast sources and assess the likelihood of being impacted. Since many weather reports are in a text and/or tabular form, they often report conditions at a regional granularity by county or town at best.

By contrast, the farmer can subscribe to the decision-support service by providing his parcel boundary, selecting ”frost” and ”chill” as event keywords, and selecting a text message as the notification medium. Figure 2 shows how a notification is generated from an extracted place name footprint and combined with an event keyword, resulting in a list of push notifications sent to farms intersecting the affected region.

Depicted in figure 2, the service automates much of the search process by using the events of interest keywords and selected data sources to extract the relevant data. Next, spatial filtering, geocoding and footprint generation are used to determine which station data are within a user defined radius to the parcel. Finally, a text notification is sent stating which sources suggest likely frost.

4. DISCUSSION AND FUTURE WORK

We have proposed a framework for data aggregation that can help farmers take precautionary measures to address

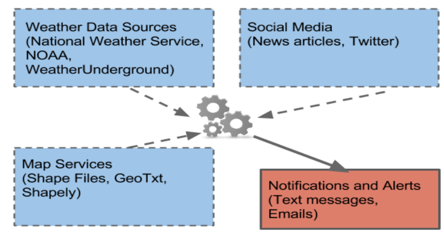

Figure 2: Extracting relevant notifications from selected sources and areas of interest.

challenges on their farms. We plan to implement this framework by combining NLP, geocoding, and Geographical Information Systems (GIS). Anticipated challenges in this proposal include quality assurance for volunteered geographic information from OpenStreetMap, and reliability of information published in social media sources like Twitter [1]. We anticipate a high variability in the granularity when using the geometry corresponding with individual farms to match features of a regional scale. Named places extracted from tweets will correspond to many granularities as well [4].

This framework can be further extended by providing users with ways to contribute feedback and suggest responses to given notifications. This information could then be used to build a recommendation system informed by location, environmental conditions, and actions taken by other farmers. Collaborative feedback will help farmers decide what techniques to use in order to mitigate the anticipated risk. Another unique challenge is to investigate the various performance tradeoffs between using lightweight keyword matching operations, and expensive GIS operations (shapefile processing) [2].

5. REFERENCES

[1] Bishr, M. and Kuhn, W. (2013). Trust and reputation models for quality assessment of human sensor observations. International Conference on Spatial Information Theory. Springer International Publishing.

[2] Kessler, C., Janowicz, K., and Bishr, M. (2009). An agenda for the next generation gazetteer: Geographic information contribution and retrieval. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.

[3] Lieberman, M. D., Samet, H., and Sankaranarayanan, J. (2010). Geotagging with local Lexicons to build indexes for textually-specified spatial data. Proceedings - International Conference on Data Engineering.

[4] Wallgrun, J. O., Hardisty, F., MacEachren, A. M., Karimzadeh, M., Ju, Y., and Pezanowski, S. (2014). Construction and First Analysis of a Corpus for the Evaluation and Training of Microblog/Twitter Geoparsers. Proceedings of the 8th Workshop on Geographic Information Retrieval.

[5] Finkel, J.R., Grenager, T. and Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics.

[5]https://catalog.data.gov/dataset/nal-agricultural-thesaurus
[6]http://deeplearning4j.org/word2vec
[7]https://pypi.python.org/pypi/Shapely
[8]https://pypi.python.org/pypi/pyshp