Social Media as a Source of Meteorological Observations

OTTO HYVÄRINEN AND ELENA SALTIKOFF

Finnish Meteorological Institute, Helsinki, Finland

(Manuscript received 3 November 2009, in final form 19 March 2010)

ABSTRACT

An increasing number of people leave their mark on the Internet by publishing personal notes (e.g., text, photos, videos) on Web-based services such as Facebook and Flickr. This creates a vast source of information that could be utilized in meteorology, for example, as a complement to traditional weather observations. Photo-sharing services offer an increasing amount of useful data, as modern mobile devices can automatically include coordinates and time stamps on photos, and users can easily tag them for content. In this study, different weather-related photos and their metadata were accessed from the photo-sharing service Flickr, and their reliability was assessed. Case studies of hail detection were then performed. The position of hail detected in the atmosphere by radar was compared with positions of Flickr photos depicting hail on the ground. As a result of this preliminary study, the authors think that further exploration of the use of Flickr photographs is warranted, and the consideration of other social media as data sources can be recommended.

1. Introduction

The number of human observations of weather is decreasing worldwide (Fiebrich 2009; Eden 2009; Tuomenvirta 2004), as manned weather stations are replaced by automatic weather stations. Usually, the frequency and homogeneity of automated observations is better than that of the manual ones, but some parameters can still be better observed by human eyes, such as the precise type of hydrometeors. In this paper we argue that automatic observations could be augmented with data from users of Internet social network sites. Heymann et al. (2007) defined social network sites as those that capture and display content (i.e., ranging from photos to text) generated by untrusted users. Here we have used the term “social media” in a broader sense to describe services through which users can report details of their life and experiences and browse other users’ experiences, which are often sorted by tags and themes.

People of the generation born since 1982 have grown up using computer technology (Barnes 2009). Cell phones, text messaging, and the Internet are all part of their culture. Now they are acquiring more and more devices with good quality cameras and global positioning system (GPS) abilities. They share photos and reports with friends and strangers alike. Typical messages can be divided into two categories: “this is what I saw” and “this is what happened to me.” At first glance, these data are unreliable, unorganized, and uncontrolled. But the amount of data is huge and increasing and should not be ignored, and its reliability should be assessed. From these data we might be able to obtain first-hand, widespread, sometimes even near-real-time information of things that really matter to the people at their location. Results of the first application of social media to the sciences have been published. Ginsberg et al. (2009) showed that the influenza activity estimates made by Google Flu Trends based on the search queries of ordinary users correlate with the weekly statistics compiled by the U.S. Centers for Disease Control (CDC) from doctors’ reports, and that Google Flu Trends can estimate influenza activity with a shorter reporting lag than the weekly CDC reports.

Weather is related to many incidents described in social media. Some messages tell about weather indirectly: a person who tells how he fell and crashed his bike does not see it as a report of road conditions. People, who take photos of the hailstorm that interrupted their meal at an outdoor restaurant, do not think that they are making weather reports. People tend to take and share images, when something unusual happens. A September snowstorm in Paris probably gets more attention than a snowstorm in northern Scandinavia in February. Therefore, by observing the messages in social media, we may collect

Corresponding author address: Otto Hyvärinen, Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland. E-mail: otto.hyvarinen@fmi.fi

DOI: 10.1175/2010MWR3270.1

© 2010 American Meteorological Society
information of high-impact weather phenomena that would otherwise be imperfectly known, if at all, by the meteorological community.

Social media services have different characteristics, for example, timeliness varies between types of services. At the moment, text-oriented services, such as Twitter, are more up to date, while services for publishing photos and videos (e.g., Flickr and YouTube) react more slowly to events. The former might be used to real-time monitoring (Vieweg et al. 2010), while the latter are more suitable for posthoc research of phenomena and verification of forecasts.

This study concentrates on the use of free-form descriptive texts and metadata to extract meteorological information from photos on the Flickr service. Quantitative estimates of the reliability of the metadata are presented in section 3, and Flickr photos depicting hail on the ground are used in case studies for the validation of hail detection algorithms in section 4.

2. Data source: Tagged photos in the Flickr service

Flickr is an image-sharing service established in 2004, and is at present owned by the Internet services provider Yahoo!. At the time of writing, more than one billion photos were available for viewing, 6% of these having geographical information included. Even though most of the photos were taken in North America and western Europe, the distribution of the photos is truly global (Fig. 1). Of all the existing social media services, Flickr was chosen for this study because the considerable amount of photos meant the probability of finding useful data was high; additionally the application programmer’s interface (API) allowed the data to be processed automatically. Images can have extensive metadata attached, typically at least a short description of the photo, when it was uploaded (“posted”) and information about the camera settings (see the example in the appendix). Automatically saved technical details in the exchangeable image file format (EXIF) can include the date when the photo was taken and geographical information from the GPS. In addition to these parameters, the user can manually add freely chosen “tags” that describe the photo. On one hand, this makes it easier for users to add tags and encourages their use; on the other hand, homonyms and synonyms can make their utilization later harder. For our purposes, the date, the geographical information, and the tags were the most important. The tags are usually less ambiguous than the short description.

3. Quantifying the reliability of Flickr metadata

Before photos and tags can be used, the quality of the metadata must be assessed. For example, the layman who labeled the photos and the scientist who uses them might not always have the same conception of different weather phenomena. Another important question is the reliability of temporal and spatial information: are the photos taken when and where the metadata indicate? a. Unambiguity of tags

To compare the conception of different weather phenomena by Flickr users with that of people with a background in meteorology, a random sample was taken of geotagged photos using eight different weather-related
tags from 2007 to 2009 (Fig. 2). To better ensure that the photos were independent from each other, photos were only included in the sample if they were taken more than 8 h apart or if the spatial distance between them was more than 0.05. This was done because people often upload more than one image of the same event. The aim was to assemble sample of 200 photos for each group, but for some tags not enough of photos were available. The authors of this paper (both with a degree in meteorology and experience with professional weather observations) then classified the photos in each tag group into three categories: “agree,” “disagree,” and “ambiguous.” Photos labeled agree clearly showed the weather phenomenon of the tag, while cases labeled disagree showed something else. The third class, ambiguous, was reserved for cases where we thought the photographer had selected the tag thinking about the right phenomenon, even though the photo was about something else. These included images of tornado damage under the tag “tornado,” snow seen on distant mountains tagged as “snow,” and stratus clouds on hillsides tagged as “fog.” Some cloud images were labeled with clearly the wrong cloud type (e.g., cumulus as stratus, cirrocumulus as cirrus) and these were classified as ambiguous. On the other hand, all cases of reduced visibility (e.g., mist, fog, haze, etc.) tagged as fog were deemed agreements, and both hail and graupel were deemed to be agreements with “hail.”

The best hit rate was found for the word snow, as it appears to be a generally well-known concept. The second best was “cumulonimbus”—a word probably used only by people with some meteorological knowledge, and referring to a relatively distinctly shaped cloud. The worst correspondence was found for the words “cirrus,” “stratus,” and “tornado”—words that are so well known that people outside meteorology use them to name brands, buildings, and sports teams, and yet find less easy to recognize visually. This is seen by the large number of “misses” in cases tagged as tornado, stratus, and cirrus. Our conclusion is that Flickr should be used to search phenomena easily recognizable by laymen. Correctly identified photos of cloud genera were sparse. Additionally, the results for tornado, for example, could be improved by focusing on a specific geographical area or time period where or when tornadoes are thought to have occurred. In such cases, even photos of tornado damage could be used (in this study they were classified as ambiguous—they can give accurate information of the location, but not of the time when the tornado passed that place).

b. Estimation of location accuracy

If a photo file contains information of the location where the photo was taken, Flickr can automatically extract it. The geospatial information in the photos can be added either when the photo was taken by a GPS device or later by comparing the time stamps of the photos and locations recorded with an auxiliary GPS tracking device. The user can also enter the location manually using the graphical user interface. Based on this, it can be speculated that the main possible reasons for errors in the location are a malfunction of the GPS device, a badly synchronized clock in the camera and the GPS device, and human error in entering the location. Unfortunately, the different ways of entering the location are not explicitly shown in the metadata, but must be indirectly deduced, for example, from the existence of EXIF tags with location-related information (if no such information is present, the location was probably entered manually).

To assess the quality of the location information, the location of the photo must be compared with known landmarks. As landmarks, famous sights at popular tourist destinations are an attractive solution, as the amount of photos is adequate and the exact coordinates are readily available from various sources. Therefore, geotagged photos of popular sights around the world from the years 2007 to 2009 were fetched for the assessment (Fig. 3). Sights were chosen to be popular and as unambiguous as possible. Naturally, the tags used in the search affect the results very much. Here a rather simple combination of tags was used, partly because this way the amount of photos was large, partly to get an overall subjective
estimate of the quality of the photos. For each landmark, only one photo from each photographer was included.

The results were promising, as for all the landmarks the median error in location was less than 1 km, with most of the errors being less than 2 km. Photos of the Fontana di Trevi had the smallest error, as the place itself is small, well defined, and cannot be photographed from a great distance, compared with the Eiffel Tower, which is visible from the other side of Paris. It is therefore reasonable to assume that most of the errors arose not from the erroneous location information but because the tag does not describe place where photo was taken but rather the subject of the photo. Very small errors, less than 10 m, were spurious. Comparison between manually geotagged and automatically geotagged photos (classified by the existence of EXIF tags with location-related information) showed no significant difference between groups, so users are very careful at geotagging their photos, at least for famous landmarks.

Encouragingly, no cases were found where the latitude and longitude coordinates were incorrectly switched (real longitude being reported as latitude, and vice versa). To make this estimation relevant for this study, it must be assumed that famous landmarks and weather events are geotagged with the same accuracy. However, this might not be the case for the manually geotagged photos, especially if the events take place in the unfamiliar surroundings that are hard to locate afterward. For automatically geotagged photos no difference should be accepted.

### c. Estimation of temporal accuracy

The time in Flickr tags is always the local time when the photo was taken, not, for example, the UTC time. As in the previous section, some reference is needed to which the time from the tags can be compared. The clock faces of the Big Ben clock tower in London, United Kingdom, are a popular sight, of which a large number of photos can be retrieved from Flickr. From these photos the position of the clock hands can be extracted. This would make an interesting project to perform automatically; however, because of the limited resources available, this was done by hand, which reduced the amount of data extracted. Because most of photos, if not all, were taken by tourists who, for most part, are possibly not very diligent in setting the time zones of their cameras or other devices, only the difference in minutes between the time stamp of the photo taken and the minute hand of the clock was compared. The reasonable assumption is that a small time difference in minutes between the photo and the clock image is not spurious, but rather indicates that the device is on time. The time difference of photos without geolocation and with geolocation, both manually and automatically given, was investigated. Photos with geolocation but without geolocation information in EXIF were classified as manually geolocated. Only one photo from each user was taken in the sample. A sample of about 100 photos was sought for all three categories, but in the case of automatically geolocated photos, only 54 cases were found, after blurred and otherwise unusable photos were discarded.

At first a rough estimate of errors in time information was done by comparing the “taken” time to the “posted” time. About 1% of images were supposed to be taken after they had been posted on Flickr, which cannot be true. About 2% of images had the same taken and posted time, which is also suspect, as some time would be expected to take place between the photo being taken and it being posted to Flickr. The temporal uncertainty was further inspected in the three categories: automatically geotagged photos, manually geotagged photos, and photos without any geolocation. In all three groups (Fig. 4), the median of the time difference was zero, but the spread, or scale, varied. The scale of the time difference was the smallest for the automatically geotagged
photos, while there was no significant difference between the scales of manually geotagged photos and those without any geolocation. Perhaps no difference between photos lacking geotagging and manually geotagged photos would be expected, because they were essentially from the same source (i.e., photos taken with devices for which the exact time was not very important). On the other hand, there could be a real difference between automatically geotagged photos and other photos, as it is easier for some devices with geolocation abilities (e.g., cell phones) to update their time and it is also more important for users to manually synchronize the time of their cameras to GPS time to get better geolocation for their photos. This indeed seems to be the case, as the time difference of the scales of automatically geotagged and other photos was statistically significant. The distributions of the differences were very non-Gaussian, and therefore Mood’s nonparametric test for a difference in scale parameters (Conover 1999) was used. The null hypothesis “no difference in scales” could be rejected at $p < 0.01$. All in all, for about 90% of the photos the error scale is less than 15 min; for automatically geotagged photos, the error is even less. Furthermore, as it is not obvious why photos of famous landmarks and weather events should have the different characteristics of temporal accuracy, these results can be used also for photos depicting weather events.

4. Case studies with hail detection

Hailstorms are short-lived, local phenomena, and a lack of reference surface observations is a major problem for comparisons of different hail detection algorithms (Tuovinen et al. 2009). The Finnish Meteorological Institute (FMI) uses a hail detection algorithm that identifies convective cores in the upper parts of clouds with weather radars (Saltikoff et al. 2010). The method is based on work by Waldvogel et al. (1979) and compares the altitude of echoes stronger than 45 dBZ with the altitude of the freezing level. The higher the storm echoes rise, the larger is the probability of occurrence of hail. In addition, a new method of hydrometeor classification based on dual-polarization radars is being tested in a joint project of FMI and Helsinki University. The method uses the polarimetric moments along a radial for classifying hail, snow, and other hydrometeor types (Keränen et al. 2007). Both these methods suffer from a fundamental property of radar beams: because of the curvature of the earth, even the lowest possible measurement gets higher and higher above the ground with increasing distance from the radar. Hence, the further a location is from the radar, the harder it is to be sure that the hailstones detected at higher altitude will actually reach the ground and thus provide a positive verification case for the algorithm.

For hail algorithm testing, photos from Flickr gave much needed ground level evidence. For case studies, photos without precise GPS coordinates can sometimes be located by recognizable targets (e.g., a golf course, an open-air restaurant). But for the processing of larger amounts of images, coordinates are needed. To select the two example cases used in this study, a search was made with the relevant keywords in Finnish (e.g., “raekuuro” for “hailstorm”), and the resulting 16 cases were sorted by the availability of radar data, sufficient metadata in the photos, copyright license to allow further use of the images, and, finally, the aesthetic values of the images. The use of the Finnish language by and large restricted the search results to the area covered by the FMI radar network. Two cases having different characteristics were selected: one was within the coverage of the dual-polarization radar, while the other was far from the radar.

For the study of the first case, in southern Finland, two approaches were explored: the hydrometeor classification based on dual-polarization radar (Fig. 5a) and a cross section of radar reflectivity from the volume scan of a conventional radar (Fig. 5b). Both indicate the formation of hail aloft, and indeed, the only available surface observation, a photo posted in Flickr (Fig. 5c), showed hail on the ground. The photo was taken with a cell phone with a built-in GPS, so the location is very precise. According to the time stamp, it was taken 12 min after the radar images, which explains the slight mispositioning of the convective core, moving eastbound.
FIG. 5. Using a single Flickr photo as ground truth for hail observations: a case in Helsinki, Finland, 11 May 2009. (a) Hydrometeor classification based on Helsinki University dual-polarization radar, elevation angle 1.5°. (b) Cross section of radar reflectivity from the volume scan of the FMI Vantaa radar, directed toward the convective core at Vuosaari, 15 km from the radar, marked with the line in (a). (c) A photo of melting hail on cafe tables at Vuosaari at the location marked with a circle in (a) and (b). (The photo is by Ilari Sani, made available under the Creative Commons Attribution License.)
available. But often, because of the observation geometry of radar, this is not the case.

For the study of the second case, in western Finland, a hail photograph from Flickr was compared with a cross section westward through the measurement volume of radar at Vimpeli, western Finland (Fig. 6a). The radar echoes above the golf course at Vaasa (105 km from Vimpeli at 1459:39 UTC at the location marked with a circle in (a). In FMI’s hail warning system, the echo in blue denotes hail with a probability of 30% (45 dBZ reaches 3.8 km). (The photo is by Timo Kytinä, made available under the Creative Commons Attribution-Noncommercial-No Derivative Work License.)

In these examples, only one photo was used for verification of each case, as we were confident of their spatial and temporal accuracy. Used this way, photos gave us more evidence of the existence of hail, but the quality of the radar method cannot be determined from photos alone, as it is rarely possible to use the photos to depict nonevents (i.e., to show that there was no hail).

5. Discussion and conclusions

In this article we used data that were originally not created to be used as weather observations. Of course, such data can be wildly inaccurate or misleading. Date information especially can be unreliable, as users may not bother or know how to set the time of their devices. Even the geographical information can be misleading, as the number of photos wrongly located at 0° latitude, 0° longitude show (Fig. 1). The tags themselves depend on the integrity and competence of the users. In social media, there are even “spam tags” where users try to maximize their publicity by tagging their data with popular but misleading tags (Heymann et al. 2007), but in the course of this study, we found no traces of them in Flickr. The tags used in this study were not of such a kind that they could be used to lure users to inappropriate content. Additionally, as we have shown above, the metadata for Flickr photos are, for the most part, reasonably reliable, especially those that have been created with automatic GPS devices.

Especially in projects where remote sensing data are compared with surface observations, some tolerance must be accepted. Therefore, the temporal and spatial uncertainty of Flickr data (Figs. 3 and 4) should be compared with the resolution of other mesoscale observations. When testing the hail classification algorithm of a polarimetric Weather Surveillance Radar-1988 Doppler (WSR-88D), Schuur et al. (2003) used two hail-chasing cars. They accepted cases where data were recorded within the storm’s radius of influence, which varied from 3.2 to 5 km depending on the speed of storm movement. In time, they allowed a 12-min time window. Special arrangements, such as chase cars or very dense hail pad networks, can only be used for scientific experiments with a limited time and area. When creating a systematic hail pad calibration procedure for operational hail suppression in Greece, Dalezios et al. (1991) used a “normal grid” with 30 × 30 cm pads at intervals of 4.5 km. They noted that one of the main drawbacks of a hail pad system is the lack of time resolution—they changed the pads every 10 days, and after each day when hail was known to have occurred. Compared with these tolerances, the uncertainties of Flickr images are quite tolerable. In general, the needs of a particular research project drive the needs for the accuracy of the particular methods.

In some sense, we are at the mercy of service providers and their policies. It would have been very interesting to
do this experiment with the free-form short messages or “micro blogs” of Twitter, but, at the time of writing, an archive of only one week of old messages was available for developers and researchers. Use of Twitter in research would need more of a proactive approach, where researchers themselves would have to collect all messages for a certain time period and from a certain geographical area. In general, data can also disappear without warning as users lose interest in sharing their data, fill up their quotas and delete older documents, or even when network sites discontinue their services.

This new type of data exposes the researcher to issues previously unfamiliar to many people in the field of meteorology (e.g., copyright and privacy issues have to be taken into account). By default the authors own the copyright to the shared documents, and researchers cannot, for example, publish individual images without permission. Some data are shared with a more permissive license (e.g., Creative Commons; Fitzgerald et al. 2007), which can ease these issues. The images used in this paper were already available under such a permissive license. Even so, we contacted the authors, who in fact had no objections to the use of their photos and released their real names for publication. Generally, privacy issues are taken into account by Flickr letting the author decide if the photo is viewable by all or only by selected users. The same decision can be made separately for the geolocation information of the photos. However, this does not solve the problem of the privacy of the people in the photos, as it is not known whether they have given permission for their images to appear in public. The researcher thus needs to exercise some discretion when using photos showing recognizable people. This kind of research might even require approval from Human Subjects Committees or other similar regulatory bodies. However, from our experience the copyright issues are the more pressing. Creative Commons offers a variety of nuanced licenses. The attribution rules determine how the author must be acknowledged, and the other options may limit the further processing of the image.

Here we concentrated on extracting information from the metadata of Flickr photos. Extensive research has been done in extracting information from images themselves or from free-flowing text in the spheres of image processing, pattern recognition, and information retrieval (e.g., Gonzalez and Woods 2008; Manning et al. 2008; Bishop 2006); utilizing such methods would be an interesting line to pursue in future research. Similar work has been carried out in content analysis where the reliability and validity of, for example, coding human communication into discrete categories has been discussed more thoroughly than in this study (e.g., Krippendorff 2004). Flickr photos do not arrive from its users as a batch but in a steady stream. Of the photos used in this study, around 50% were available 1 day after the photo was taken and around 75% were available after 1 week. Other popular social media communities, such as Twitter and Facebook, may be more real-time oriented. But the use of Twitter in research is hampered by the short archive, while Facebook is by its nature very concerned with privacy, and thus much less data is available to researchers than in Flickr. Here our passive, unobtrusive approach used in this study could be complemented with a more active approach in which users are encouraged to submit relevant up-to-date information to researchers or even to forecasters (e.g., Ferree et al. 2009). For submitting information and discussing results, all social media services provide tools for like-minded users to gather together, usually called “groups” or “channels.” Technically this is not difficult and weather enthusiasts like storm spotters would be surely motivated to join. However, widening the participation to the general public may prove to be difficult if the service does not reward the users somehow. For storm spotters excitement of participating in research is a reward in itself, but if this is enough for the general public remains to be seen. Our unobtrusive approach has the advantage that we reach users not directly interested in weather observations. The third, even more direct, approach is the active collection of observations by contacting the businesses and residences in the area under study using telephone interviews as in Ortega et al. (2009). In the future, this might be performed utilizing social media and the geolocation of users, but the services must mature considerably before this is technically feasible and takes into account the privacy concerns. All these three approaches have their uses, and do not exclude each other. Again, which approach is the preferred one depends on the needs of the researchers.

In our experience, Flickr data are useful for case studies when photos are available but meteorological data are not. For more quantitative studies, we would be more cautious in recommending Flickr. Sometimes the data retrieval method proved to be unreliable, sometimes metadata were lacking. But we encourage others to investigate this data source. The importance of this data will grow in coming years as users, services, and devices grow more sophisticated. If a large enough sample size can be obtained with a small enough amount of erroneous data, then Flickr or any similar data source will prove to be very useful. Especially for the study of mesoscale phenomena, such as hail and tornadoes, any regular network of weather stations is too sparse. This new data could, in part, augment regular observations. But there are also parameters, which nobody has ever
even tried to observe systematically, such as how slippery pavements are, and the real challenge for future scientists is to try to imagine truly novel uses for such data.

Acknowledgments. We thank Profs. S. Joffre, D. M. Schultz, and T. Vesala and Drs. E. O’Connor, S. Niemelä, and A. Huuskonen and Mr. Janne Kotro for advice and comments. The research was partly supported by the Finnish Funding Agency for Technology and Innovation, Tekes, and by the Academy of Finland. The insightful suggestions of the reviewers were very helpful.

APPENDIX

An Example of Metadata in Flickr

An example of a result [in Extensible Markup Language (XML)] for a Flickr API query for a photo (uploaded by one of the authors) is shown in Table A1. Only relevant information for this study is shown. Each photo has an identification number (line 1) and an owner (line 3). The owner writes the title (line 4), the description (line 5), and tags (lines 10–11), and decides on the license (line 2, the code 4 means Creative Commons Attribution 2.0 Generic), who may see the photo (line 6), and if its geolocation is public (line 14). Flickr extracts the time the photo was taken (line 8) from the EXIF information in the photo (complete EXIF information available with a separate query) and adds the time the photo was received by Flickr (in the form of a UNIX timestamp, line 7). In this case the photo and its location are visible to everyone, and the license allows others to copy, distribute, and to adapt the photo, provided the owner is acknowledged. Note how tags with white space or punctuation are converted to versions without them. These stripped versions are used in queries and in constructing uniform resource locators (URLs).

REFERENCES

Keränen, R., E. Saltikoff, V. Chandrasekar, S. Lim, J. Holmes, and J. Selzler, 2007: Real-time hydrometeor classification for the operational forecasting environment. Preprints, 33rd Int. Conf. on Radar Meteorology, Cairns, Australia, Amer. Meteor.


