Virtual Field Research with Social Media: A Pilot Case of Biometeorology

Christoph Kling, Sergej Sizov, Steffen Staab
University of Koblenz-Landau, Germany
Institute for Web Science and Technologies - WeST
{ckling,sizov,staab}@uni-koblenz.de

1. INTRODUCTION

In social sciences, field research is an important approach for analysing individual and social behaviour. In doing so, researchers are moving in natural environments of individuals to observe unbiased behavioural patterns. However, systematic field work is quite time consuming and expensive. Due to this, sample size often remains small, leading to a problem of external validity. In most of the cases, long term studies are not feasible. Also, there are hypotheses which are not verifiable by observation, for instance opinions or behaviour in rare situations. Our contribution aims to demonstrate the potential of Social Web for gaining comparably deep insights into human behaviour with significantly lower overhead. In fact, Social Media increasingly offers rich metadata, such as temporal and geospatial attributes. This information allows us to integrate additional environmental knowledge - such as weather data - and to interpret contents of Social Web as a replacement for traditional observations in field research

2. VIRTUAL FIELD RESEARCH IN BIOMETEOROLOGY

We consider the domain of *biometeorology* as a characteristic showcase of our investigation. Field studies demonstrate weather influence onto human temper, activity profile, opinion forming, and social interactions. Usually, studies in human biometeorology combine knowledge about persons and respective weather conditions and analyse correlations and patters of atmospheric influences, e.g. by means of statistical methods. Whilst data for the most important weather factors can be retrieved from weather stations and retrospective meteorological archives, knowledge about human behaviour is usually gained by means of comprehensive field observations or surveys. Consequently, real-life field studies are usually limited to particular locations (e.g. cities or countries) or even localized places of interest (e.g. city districts, tourism attractions, parks, etc.). Since biometeorological studies require systematic experiments under various weather conditions, it takes a long time (months of even years) to settle representative series of observations. At the same time, the flexibility of a real-life field study is quite restricted. We demonstrate how biometeorological virtual field research can be performed based on data from social media and consider two popular scenarios: affective analysis and activity analysis of observed persons.

3. AFFECT ANALYSIS IN MICROBLOGGING

Participant affectation in social media can be directly measured by means of text-based affective mining. To quantify weather impact on Social Web users in mood, we collected a large corpus of Twitter postings through the platform-specific search API. Using built-in Twitter API filter conditions on posting timestamps and user locations, we collected postings in 25 km radius around 230 meteorological stations in Germany for the period from April to June 2010. Messages containing URLs were filtered, presumably eliminating most of the spam messages.

We took two different approaches for classifying messages by affect classes. First, we used the "Affective Dictionary Ulm", a dictionary, developed to rate conversations in psychotherapy [1]. The dictionary is assigning twelve different affect subclasses of either positive or negative effect to a set of 24.861 words, which could be reduced to 1.974 stems. Messages were classified for each single affect when containing one of the words from that class in the dictionary. Those affect classes then are representing messages containing a word from an affective *vocabulary* used by a user, though not saying anything about the actual affective *state* of a user.
The second method to extract affectations of users by their messages was based on a typical classification, used in affective mining. First we conducted a survey with 130 students, rating 7.627 different twitter messages for positive, negative and tired affect of the user each on a five item Likert scale. 541 messages were rated two times or more so that a corrected agreement κ could be calculated using the interrater agreement by Fleiss[2]. Using five categories for each affect was not leading to an acceptable agreement, which was to be expected when taking the short length of twitter messages into account. Reducing each of the five degrees of the affect into two categories, for no affect and presence of the affect, leads to a reasonable κ between 0.41 and 0.76 for each rated affect. Knowing that we have a quite well rated set of twitter messages, we constructed a univariate model to predict the affective rating for the rest of the messages.

To extract "good" and "bad" weather, we run a principal component analysis on weather variables grouped by date which returned two main factors corresponding to warm and sunny weather with low precipitation for the first and cold, windy weather with precipitation for the second factor. We then found coherences between mood and weather by grouping messages by date and correlating the average mood of twitter users with corresponding weather factor values.

Subsequently, we compared key observations with recent real-life field research [5, 3]. In line with most experimental studies, our results quantify major effects of good and bad weather: good weather raising the positive affected vocabulary use, as well as bad weather lowering the positive affective state and raising the negative affected vocabulary use.

4. VIRTUAL ACTIVITY ANALYSIS

As a second example for a virtual field study we have digressed from text-oriented towards image-focused social media. With regard to biometeorology, researchers in field studies have e.g. observed comfort zones for people outside based on weather, activities and clothing. In many cases, human activities are directly reflected by rates of content generation in social media. As an example, we considered Flickr photos for one particular location (*Slottsskogen* park in Göteborg, Sweden). By querying the Flickr API for photos located within the urban area of Göteborg, we obtained a set of 23.977 photos. Subsequently, we compared the number of photo sets taken within the area of the Slottsskogen park with the number of photos taken in the whole urban area. The result clearly shows that people tend to visit the park at warm weather conditions. For depicted people, we calculated the so-called insulation index of human clothing (clo) by using the empirical ASHRAE standard 55-92. The average observed clothing insulation showed how people wear less insulated clothes when temperature is raising. The results support findings of the recent real-life study by Thorsson et al. [4] for same location (Slottsskogen park).

5. APPLICATIONS BASED ON VIRTUAL FIELD STUDIES

Findings of virtual field studies not only are of scientific value but also can directly trigger new practical applications due to their realtime nature and scalability.

We found that people are visiting some locations more often depending on weather. This finding enables the calculation of a weather dependent ranking in popularity of places.

Creating such a ranking can be done straight forward by extracting clusters from geospatial information of photographies and ranking these clusters by a simple algorithm: We assign normalised vectors of weather variables to each picture set, then we calculate the distance of this vector to a given weather condition vector and iteratively return the cluster number of the closest photo set regarding weather. If a certain cluster was returned c times, it will be recommended next. This way, weather as well as visitor frequency determine the rank of a location.

Applying this algorithm on 13.761 photos from Berlin crawled in Flickr returned plausible results: We calculated the rank of a swimming pool area, called the "Badeschiff Treptow" which is known to be a popular place at warm weather only. Looking at the ranking given different temperatures and full sunshine (Table 1), the location is the least crowded place at cold days, resulting on the worst or near worst rank. Raising the temperature is raising the rank of the location significantly, peaking in rank 15 at 30°C.
This example shows how findings of virtual field studies can be translated into real applications.

Temp. °C	-10°	-5°	0°	5°	10°	15°	20°	25°	30°
Rank	40	38	36	33	26	21	24	35	15

Table 1: Visitor frequency rank for a swimming pool area ("Bathing ship Treptow") on sunny days, depending on temperature. 40 locations were compared.

6. BENEFITS

We demonstrated the new method of virtual field studies by reproducing basic findings of biometeorology. This is why the results are mostly trivial, whilst the method is new. It is up to other researchers to apply our method on new, interesting hypotheses.

Key advantages of the proposed methodology are low organizational overhead, flexibility of experimental design, as well as comprehensive coverage of locations and timeframes. The use of Social Web as a playground for virtual field studies offers new opportunities to practitioners of field research as well as to designers of social information systems. On one hand, design and organization of field studies can be substantially improved by virtualizing large-scale problem settings. On the other hand, gained insights can be used for optimising Social Web services and improving user support.

7. REFERENCES

[1] L. Devillers, L. Lamel, and I. Vasilescu. *Textanalyse. Anwendungen der computerunterstützten Inhaltsanalyse. Beiträge zur 1. Textpack-Anwenderkonferenz*. Opladen, 1992.

[2] J. Fleiss et al. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971.

[3] M. C. Keller, B. L. Fredrickson, O. Y. S. Cote, K. Johnson, J. Mikels, A. Conway, and T. Wager. A warm heart and a clear head: The contingent effects of weather on mood and cognition. 2005.

[4] S. Thorsson, M. Lindqvist, and S. Lindqvist. Thermal bioclimatic conditions and patterns of behaviour in an urban park in goeteborg, sweden. *International Journal of Biometeorology*, 48:149–156, 2004. 10.1007/s00484-003-0189-8.

[5] D. Watson. *Mood and temperament*. New York: Guilford Press, 2000.