International Workshop on Social Media World Sensors (Sideways)

Workshop Programme

09:00 – 09:30 – Introduction by Workshop Chairs

09:30 – 10:00 – Dane Bell, Daniel Fried, Luwen Huangfu, Mihai Surdeanu, Stephen Kobourov, *Challenges for using social media for early detection of T2DM (invited talk)*

10:00 – 10:30 – Tim Kreutz and Malvina Nissim, *Catching Events in the Twitter Stream: A showcase of student projects*

10:30 – 11:00 Coffee break

11:00 – 11:30 – Udo Krushwitz, Ayman Al Helbawy, Massimo Poesio, *Exploiting Social Media to Address Fundamental Human Rights Issues (invited talk)*

11:30 – 12:00 – Emanuele Di Rosa, Alberto Durante, *App2Check: a Machine Learning-based system for Sentiment Analysis of App Reviews in Italian Language*

12:00 – 12:30 – Christian Colella, *Distrusting Science on Communication Platforms: Socio-anthropological Aspects of the Science-Society Dialectic within a Phytosanitary Emergency*

12:30 – 13:00 – Luca Vignaroli, Claudio Schifanella, K. Selcuk Candan, Ruggero Pensa, Maria Luisa Sapino, *Tracking and analyzing the "second life" of TV content: a media and social-driven framework (invited industrial demo)*

Editors

Mario Cataldi — Université Paris 8, France

Luigi Di Caro — Department of Computer Science – University of Turin, Italy

Claudio Schifanella — RAI – Centre for Research and Technological Innovation, Turin, Italy

Workshop Organizers/Organizing Committee

Mario Cataldi — Université Paris 8, France

Luigi Di Caro — Department of Computer Science – University of Turin, Italy

Claudio Schifanella — RAI – Centre for Research and Technological Innovation, Turin, Italy

Workshop Programme Committee

Luca Aiello — Yahoo! Labs, London
Andrea Ballatore — University of California, Santa Barbara, USA
Iván Cantador — Universidad Autónoma de Madrid, Spain
Federica Cena — University of Torino, Italy
Martin Chorley — Cardiff University, Wales
Emilio Ferrara — Indiana University Bloomington, USA
Simon Harper — University of Manchester, England
Dino Ienco — Irstea, UMR TETIS, Montpellier, France
Séamus Lawless — Trinity College Dublin, Ireland
Emmanuel Malherbe — Multiposting, France
Rosa Meo — University of Torino, Italy
Ruggero G. Pensa — University of Torino, Italy
Rossano Schifanella — University of Torino, Italy
Thomas Steiner — Google, USA
Luca Vignaroli — RAI – Centre for Research and Technological Innovation, Turin, Italy

Table of contents

Author Index

Introduction

It is our great pleasure to welcome you to the *2016 ACM Workshop on Social Media World Sensors - Sideways 2016,* which is held in conjunction with the *10th Edition of its Language Resources and Evaluation Conference.- LREC 2016* in Portoroz, Slovenia.

This second edition of the workshop aims at bringing together academics and practitioners from different areas to promote the vision of social media as *social sensors*. Nowadays, social platforms have become the most popular communication system all over the world. In fact, due to the short format of messages and the accessibility of these systems, users tend to shift from traditional communication tools (such as blogs, web sites and mailing lists) to social network for various purposes. Billions of messages are appearing daily in these services such as Twitter, Tumblr, Facebook, etc. The authors of these messages share content about their private life, exchanging opinions on a variety of topics and discussing a wide range of information news.

Even if this system cannot represent an alternative to the authoritative information media, considering the number of its users and the impressive response time of their contributions, they represent a sort of real-time news sensor that can also predate the best newspapers in informing the web community about the emerging topics and trends. In fact, the most important information media always need a certain amount of time to react to a news event; i.e. professional journalists require time, collaborators and/or technology support to provide a professional report. However, a user can easily report, in few characters, what is happening in front of the user's eyes, without any concern about the readers or the writing style. These aspects make social services the most powerful sensor for events detection and automatic news generation. The aim of this workshop was to ask researchers to enter into such view, by studying how social platforms can be used in real-time scenarios to detect emerging events and enrich them with contextual information.

First, we would like to thank the organizing committee of LREC 2016 for giving us the opportunity to organize the workshop. Second, we would like to thank our program committee members. And of course, we would like to thank all the authors of the workshop for submitting their research works and for their participation.

We hope you will enjoy the second edition of the Sideways workshop and the LREC Conference, and have a great time in Portoroz.

The call for papers attracted submissions from India, Europe, Africa, and the United States.

The program committee reviewed and accepted the following:

Venue or Track	Reviewed	Accepted	
Full Papers	5	3	60%
Short Papers	2	0	0%
Total	7	3	42,8%

We also encourage attendees to attend the invited talk presentation.

• *Challenges for using social media for early detection of T2DM* - Dane Bell - University of Arizona

• *Exploiting Social Media to Address Fundamental Human Rights Issues* - Udo Krushwitz, Ayman Al Helbawy, Massimo Poesio - University of Essex

- *Tracking and analyzing the "second life" of TV content: a media and social-driven framework -* Luca Vignaroli, Claudio Schifanella, K. Selcuk Candan, R. Pensa, Maria Luisa Sapino – RAI Research Center, Arizona State University, University of Turin

Putting together *Sideways 2016* was a team effort. We thank the authors for providing the content of the program and we are grateful to the program committee who worked very hard in reviewing papers and providing feedback for authors. Finally, we thank the hosting organization.

We hope that you will find this program interesting and that the workshop will provide you with a valuable opportunity to share ideas with other researchers and practitioners from institutions around the world.

Co-Chairs
Luigi Di Caro
Mario Cataldi
Claudio Schifanella

Challenges for using social media for early detection of T2DM

Dane Bell, Daniel Fried, Luwen Huangfu, Mihai Surdeanu, Stephen Kobourov

University of Arizona
Tucson, AZ 85721, USA
{dane, dfried, huangfuluwen, msurdeanu, kobourov}@email.arizona.edu

Abstract

Twitter and other social media data are utilized for a wide variety of applications such as marketing and stock market prediction. Each application and appropriate domain of social media text presents its own challenges and benefits. We discuss methods for detecting obesity, a risk factor for Type II Diabetes Mellitus (T2DM), from the language of food on Twitter on community data, the peculiarities of this data, and the development of individual-level data for this task.

1. Introduction

This project is in aid of implementing a system that can detect individuals who are likely to be at high risk for preventable Type II Diabetes Mellitus (T2DM), a life-shortening disease that generates fatal complications that is common in the developed world. The system is part of an effort to nudge (Thaler and Sunstein, 2008) individuals at risk for T2DM to make changes to their diet and exercise level to prevent or delay the disease's onset. The central hypothesis of this work is that (features of) individuals' tweets about food correlate with their real-world food consumption, which is in turn correlated with their likelihood of developing T2DM. We began by learning to detect obesity, a factor often implicated in the rising rate of T2DM diagnosis in the United States. Through work on community-level data, we found that this hypothesis was supported, but our machine learning model for detecting obesity rates at a state level proved not to transfer well to individuals. For this reason, we sought to engage with tweeters and other individuals on social media sites to help collect individual data through the use of a novel, 20-questions-style quiz generated semiautomatically from a classifier trained on community-level Twitter data. We discuss our approach to these challenges as well as future directions.

2. Community-level detection

In order to begin detecting obesity, we began with models over community-level data, namely cities and US states. Using the Twitter API, we gathered ca. 3.5 million tweets containing relevant hashtags such as #dinner and #breakfast, of which 16% (562,547 tweets) could be assigned a location within a US state (Fried et al., 2014).

As is typical in Twitter data, our tweets required significant preprocessing, most importantly in removing Uniform Resource Locators (URLs) and @mentions of user handles. We experimented with different feature sets, including limiting our features to hashtags, food words, or both. We also used Latent Dirichlet Analysis (Blei et al., 2003) to mitigate sparsity, with 200 topics added to our feature set. In all cases, we used Support Vector Machines (SVM) with a linear kernel (Vapnik and Vapnik, 1998).

A model was trained to predict whether a state was above or below the national median for overweight rate according to a Kaiser Commission on Medicaid and the Uninsured (KCMU) analysis[1]. In addition to predicting community-level obesity at an accuracy of 80%, this dataset was able to predict whether a state had greater or lower than median diabetes rate (69% accuracy). Similar models were able to predict the less obviously related variables of location and political party affiliation.

Different sets of features were optimal for each of these factors, sometimes favoring all words and sometimes only food words, for example, although the addition of LDA topics was beneficial in all cases. Table 1 shows the top 20 features of the SVM model for classification, displaying intuitively appropriate correspondence between diets (*fried* vs. *vegan*) and rate of obesity and diabetes.

3. Transfer to individuals

Although the previous experiments showed that tweets about food contained information about our variables of interest, the performance of the community-trained models on manually annotated individual Twitter accounts was at chance. This made it clear that a corpus of individually annotated Twitter accounts was necessary for accurate prediction, and we devised a 20-questions-style quiz site based on our community-level data to serve two purposes: evaluation on individuals, and data collection for new models.

SVMs do not produce models that are easily converted into natural-language questions, but tree-based classifiers such as random forest classifiers do. Through further experimentation, we discovered that a small number of relatively shallow decision trees with discrete features could perform comparably to our prior SVM model when predicting state-level overweight rates (Bell et al., 2016). The high performance of these models (78% accuracy, compared to 80% of our previous work and baseline accuracy of 51%) came in spite of their simplicity and interpretability: the best performing model used a 7-tree decision forest with maximum depth 3 and three-bin discrete features.

These trees were converted semiautomatically into natural languages questions, so that a feature based on the word *brunch* became "How often do you eat brunch?" with three multiple-choice Likert scale (Likert, 1932) answers such as *Practically never*. Figure 1 illustrates one tree of the decision forest. The questions that were asked depended cru-

---

[1] http://kff.org/other/state-indicator/
adult-overweightobesity-rate/

1

Class	Highest-weighted features
overweight: +	i, day, my, great, one, *American Diet* (chicken, baked, beans, fried), #snack, *First-Person Casual* (my, i, lol), cafe, *Delicious* (foodporn, yummy, yum), *After Work* (time, home, after, work), house, chicken, fried, *Breakfast* (day, start, off, right), #drinks, bacon, call, eggs, broccoli
overweight: -	*You, We* (you, we, your, us), #rvadine, #vegan, make, photo, dinner, #meal, #pizza, *Giveaway* (win, competition, enter), new, *Restaurant Advertising* (open, today, come, join), #date, happy, #dinner, 10, jerk, check, #food, #bento, #beer
diabetes: +	*Mexican* (mexican, tacos, burrito), *American Diet* (chicken, baked, beans, fried), #food, *After Work* (time, home, after, work), #pdx, my, lol, #fresh, *Delicious* (foodporn, yummy, yum), #fun, morning, special, good, cafe, #nola, fried, bacon, #cooking, all, beans
diabetes: -	#dessert, *Turkish* (turkish, kebab, istanbul), #foodporn, #paleo, #meal, *Paleo Diet* (paleo, chicken, healthy), i, *Giveaway* (win, competition, enter), *I, You* (i, my, you, your), your, new, today, #restaurant, *Japanese* (ramen, japanese, noodles), some, jerk, #tapas, more, *Healthy DIY* (salad, chicken, recipe), *You, We* (you, we, your, us)

Table 1: Top 20 highest-weighted features in descending order of importance for each dataset from Fried et al. (2014), for both the positive and negative classes. For example, "overweight: +" indicates the most representative features for being overweight, whereas "overweight: -" shows the most indicative features for **not** being overweight. The features include LDA topics, with manually assigned names (*italicized*) for clarity, and a few of their most common words within parentheses.
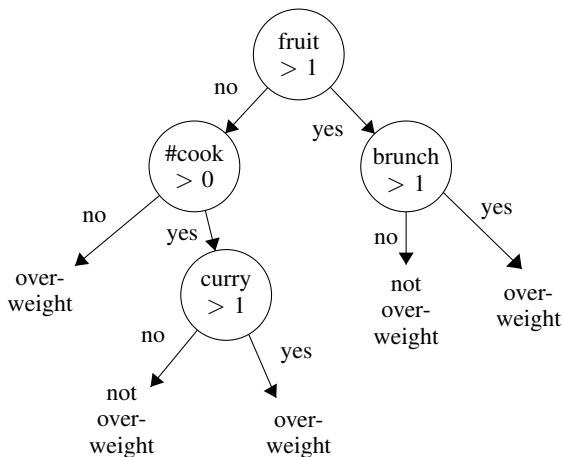

Figure 1: A decision tree from the random forest classifier trained using state-level Twitter data.

cially on which answers the quiz takers provided, as in the tree's binary choices based on three quantized values, 0, 1, and 2. Notice that this conversion relies on our underlying assumption that mentions of a food word are correlated with the consumption of that food. The 20-questions-style quiz based on these questions allowed us to evaluate how

well the community-trained classifier applied to individuals, resulting in high accuracy (79%) which was nevertheless lower than the baseline (82%) in our highly biased sample. We interpret this to indicate that, as in our initial experiments, individuals are not highly representative of their regions (or vice versa), meaning that individual-level training information is necessary for good individual classification. Fortunately, by engaging with quiz takers, the site also afforded an opportunity to collect individual-level training data. Quiz takers optionally provided their (public) social media accounts and their height and weight, from which we can calculate body mass index (BMI), as well as other demographic information: location, age, and gender. This will allow data collection from these accounts with permission to train more direct, individual-based classifiers.

4. Discussion

The greatest challenges for obesity detection are familiar from other NLP work. There is the persistent problem of non-human accounts (e.g., businesses, organizations, and bots) which add noise to the training data. The signal fighting against that noise is also imperfect, notably in its sparsity, since the average user has on the order of hundreds of tweets, of which only a very small percentage regard food. However, tweets that do not mention food may still be useful for obesity detection, much in the way that food tweets can significantly predict political affiliation (Fried et al., 2014) through indirect cultural connections.

Future work will include taking more information into account in the models. With individual-level data, we can capitalize on users' locations, photo, user handle, bio, and age, all of which are informative, though optional, parts of a Twitter profile. With these as well as features generated from the tweets themselves, classifiers can be constructed for intermediate factors such as gender which will in turn add valuable features for obesity classification. This in turn will improve our ability to develop a valuable public health tool for detecting and preventing T2DM efficiently through social media.

5. References

Bell, D., Fried, D., Huangfu, L., Surdeanu, M., and Kobourov, S. (2016). Towards using social media to identify individuals at risk for preventable chronic illness. In *LREC 2016*.

Blei, D., Ng, A., and Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., and Bell, D. (2014). Analyzing the language of food on social media. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 778–783. IEEE.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, 140:1–55.

Thaler, R. and Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Vapnik, V. N. and Vapnik, V. (1998). *Statistical learning theory*, volume 1. Wiley New York.

Catching Events in the Twitter Stream:
A showcase of student projects

Tim Kreutz and Malvina Nissim

Center for Language and Cognition Groningen – Language Technology
Rijksuniversiteit Groningen, The Netherlands
`t.j.kreutz@student.rug.nl,m.nissim@rug.nl`

Abstract

A group of bachelor students in information science at the University of Groningen applied off-the-shelf tools to the detection of events on Twitter, focusing on Dutch. Systems were built in four socially relevant areas: sports, emergencies, local life, and news. We show that (i) real time event detection is a feasible and suitable way for students to learn and employ data mining and analysis techniques, while building end-to-end potentially useful applications; and (ii) even just using off-the-shelf resources for such applications can yield very promising results.

1. Introduction

The availability of a constant flow of information in the form of short texts makes it in theory possible to collect real time information about virtually all sorts of events that are being written about. The attractiveness of this is evident, and so is its potential social utility. ReDites, an event detection and visualisation system (Osborne et al. 2014), is a prime example of this, as it was developed in order to help information analysts identify security-related events.

However, how the twitter stream can be successfully exploited to this end isn't straightforward, neither conceptually nor practically. First, specific events must be detected, and related tweets clustered. This step must rely on a definition of what an event is, which is often application dependent. Second, the retrieved tweets must be filtered and processed to minimise noise, both in terms of pertinence as well as in terms of the noise typical to the nature of tweets. Third, in order to produce a meaningful tool, the output must be evaluated at a development stage, and made as usable as possible for the end user in its final form, for example providing customisation and visualisation features.

In this short paper, we report a series of efforts within a bachelor programme in information science, where a group of nine students developed different systems with different specific aims, all exploiting real time tweet-derived information, and all socially relevant. The systems work with Dutch tweets, but their architecture is virtually language-independent, as long as tweets and basic language processing tools are available. Suggesting novel methodologies or applications for event detection wasn't our primary concern when developing the systems and when writing this contribution. Instead, by describing this collection of different student projects in the area of real time event detection towards social utility, we have a twofold aim. First, we show that even leveraging off-the-shelf tools and basic pre-processing can yield interesting, promising, and even unexpected results, with a variety of applications. Second, as we have observed that event detection on Twitter has been a useful and suitable task for students, helping them gain familiarity with data mining and data analysis while putting together end-to-end systems, we hope to inspire others to embark on similar exercises.

2. Applications

The University of Groningen extracts Dutch tweets from the Twitter Firehose and provides access to the stream for students and employees (Tjong Kim Sang 2011). Students used the backlog of available tweets to select a specific snapshot or demonstrated their systems using the most recent tweets.

The definition of an event was very application dependent, but always informed by some secondary information, be that peaks in Twitter usage, shared time and location of tweets or overlap with news headlines. Practically and generally speaking, tweets about a single event overlap in some way, and finding sophisticated ways to detect this overlap was a core challenge for all projects.

For none of the developed applications, typical noise in tweets was of any particular concern, although hashtags and URLs were often removed or replaced by placeholders in preprocessing.

Visualization for end users was provided by a few of the projects (Kreutz 2015; de Kleer 2015; Pool 2015), where live demonstrations of their results consisted for example in a website listing relevant tweets per news headline, or maps where categorised events are clustered by location.

The students' systems that we describe showcase four different application areas: sports, emergencies, local events, and news.

2.1. Sports

Sports fans like to stay up to date with real time scores by accessing websites such as `livescore.com`, which provides live overviews of football matches for major Leagues.[1] The overview consists of tables including major events in the game, like goals and yellow/red cards. It takes a lot of time to manually input in-game events, which is why obtaining reliable real-time updates can be expensive and automatising this process leveraging real time Twitter data becomes attractive. Three of the projects were concerned with automatically reproducing such tables by using the stream of Twitter data to automatically detect and classify in-game events. We describe here one of the developed systems, where matches of the Dutch national team in the

---

[1]Companies such as Livescore.com buy real-time information for prices that vary according to the prestige of the League.

2014 World Cup of soccer are used as a case to demonstrate the approach (Kuiper 2015).

The selected five matches featured a total of 19 goals, 15 yellow cards and no red cards, so only the first two types of events were predicted. For each match, two hours of Dutch Twitter data from the first minute of the match was collected, resulting in a total of 4.376 relevant tweets.

Kuiper (2015) makes use of past sports event and automatic annotation to save on the effort that would go into annotating such a large set of tweets. Firstly, only tweets that contain hashtags referring to a certain match are considered. The convention of such a hashtag is using the first three characters of the involved teams (#SpaNed for Spain versus the Netherlands). Secondly, the timestamp of each tweet is compared to the timeline of the actual match. As such, all tweets posted up to three minutes after an actual event took place will be annotated as being about the event. In the training data, distribution of the classes no event, goal and yellow card were respectively 55%, 37% and 8%.

The data was then modified to allow for a more general application of the system. This involved replacing specific scores and players with placeholders. Occurrence of specific keywords that denote an event were used as features, along with the length of the tweet and the tf-idf term vector.

Beyond detecting events in single tweets, detecting events has to do with grouping relevant tweets in the right way. Detection of peaks in Twitter activity has been used to detect events (Corney, Martin, and Goker 2014; Van Oorschot, Van Erp, and Dijkshoorn 2012). Specifically, Chakrabarti and Punera (2011) demonstrate how tweet volume signifies important events in sport matches. However, in soccer it is more likely that two important events occur in close proximity which is problematic for peak detection, since the events will be grouped as one event. To more accurately distinguish between events, Kuiper (2015) implements a rule-based system that looks at tweet content during peaks. If at least fifteen% of tweets are classified as goal-tweets, the rule based system determines whether the mentioned score is logically probable (a match with score 1-1 logically progresses to either 1-2 or 2-1) and updates the score. This way, the score is updated before a potential second goal, allowing consecutive score updates to be detected.

Using the Multinomial Naive Bayes implementation in Scikit learn (Pedregosa et al. 2011), with the set of features in Table 1, classification of individual tweets yields an f-score of .843. For the matches of the Dutch national team, sixteen out of seventeen goals were detected in the right minute (f-score .940), but detection of yellow cards was harder (f-score .500). Overall the results are encouraging for future extensions.

Kuiper (2015) moreover shows the potential of automatic detection of subevents for sports. Specifically with regards to grouping and distinction of isolated events, it gives an idea of the considerations to be made and the challenges to overcome.

2.2. Emergencies

Twitter allows for detection of real-time sub-events in sports because relevant tweets follow these events almost

Table 1: Features for Kuiper (2015)'s sports event detection system.

Feature	Type	Reason
TF-IDF	Float	Indicate the importance of a word in relation to the document in which the word has been found.
Length	Integer	Tweets posted after important events tend to be shorter.
Goal	Boolean (1 or 0)	Tweets containing the word "goal" or "scoort" or "doelpunt" could very well be tweets describing a goal.
Score	Boolean (1 or 0)	Tweets containing scores such as "2-1" or "3-2" could very well be tweets describing a goal if they are not predictions.
Predict	Boolean (1 or 0)	Tweets containing words such as "gok" or "prono" could very well be predictions instead of goals.
Yellow	Boolean (1 or 0)	Tweets containing words such as "geel" or "gele" could very well be indicators of a given yellow card.

instantly. The delay between a real-time occurrence and its social resonance are thus minimal. This adds to the social relevance of detection of events that are particularly time-sensitive, such as emergency situations. This section will look at two different emergency scenarios: earthquakes in the Dutch province of Groningen, and detection of context for events reported by Dutch emergency services.

2.2.1. Earthquakes

The detection of earthquakes on Twitter has been extensively documented (Sakaki, Okazaki, and Matsuo 2010) for Japan, where the tweet density is high and earthquakes occur relatively frequently. The research focuses on detection of earthquakes and extraction of the time that it occurred, along with the location. Earthquakes with a magnitude of 3.0 or higher on the Richter scale were successfully detected in 96% of the cases, and real-time detection led to notifying civilians faster than the Japan Meteorological Agency could, in most cases.

Detection of earthquakes in Groningen has only recently become relevant since gas extraction in the province led to a 200% increase in earthquakes over the past ten years (Kuipers 2015). This has lately sparked debate in politics and media and increased public involvement. Detection of earthquakes using Twitter can thus contribute to timely updates, but it may also map public sentiments.[2]

To develop his system for detecting earthquake events via Twitter, Kuipers (2015) used data from the Dutch Meteorological Institute (KNMI) from January 2014 until April 2015. The data contained 60 earthquakes with a magnitude of 1.2 or higher on the Richter scale, their timestamp and location of the epicenter. Weaker earthquakes are generally considered intangible for humans, and hence not useful for the research as there would be no tweets about them. A preselection of Twitter data was made by selecting tweets containing the words 'beving' or 'aardschok' that were tweeted up to four hours after the occurrence of an earthquake.

Using Weka (Hall et al. 2009), a Naive Bayes classifier

---

[2]In the context of earthquakes in the Groningen area, this is interesting also in the context of NAM's compensation duties for earthquake-caused damage to local properties.

4

was trained on the annotated tweets and tested via cross-validation. Results show that tweets are correctly classified as relevant or irrelevant to a given earthquake in over 91% of the cases. Among the most distinguishing features are the mention of a location in the Groningen or Drenthe province (boolean) which usually signals an actual earthquake, and the mention of political terms (boolean) which usually signals no actual earthquake. Further features used as potential indicators of relevant tweets are mentions of numbers, which can signal a specific time or magnitude, and certain signal words that are used to signal the sensation of experiencing an earthquake ('voel', 'tril', 'knal').

2.2.2. Emergency services

P2000[3] is a live repository of all emergency services active in a given area. All reports are publicly available and real-time updated communications of and between Dutch police, ambulance and fire department services are available.

The work described in (Louwaars 2015) is concerned with matching user tweets to reports from emergency services in the same area. The rationale behind this is that such matches could be used to diminish delay in notifying stakeholders, or adding context to official, quantitative reports. The real-time nature of Twitter makes it particularly suitable for detecting time-sensitive events like emergencies.

One month of emergency reports and tweets were downloaded from the P2000 website in April 2015 for the larger Groningen area. This resulted in 700 'matches' of reports to one or more tweets with a similar location and time. A Naive Bayes model was trained on 80% of the annotated data, using simply word occurrences as features, to classify tweets as 'relevant' or 'irrelevant'. Testing on the remainder 20% resulted in a significant improvement over the baseline (75% of tweets were annotated as irrelevant) with a final accuracy of 91%.

A critical reflection on these results can be that Louwaars (2015) observes that for some events the amount of tweets is too low to draw any solid conclusion. He further indicates that few of the relevant tweets comes from 'real' twitter users, with substantial data coming from automated emergency service accounts. As a solution, the system could be trained on emergencies that can draw from a larger pool of tweets, indicating more severe cases or emergencies that occur in more densely populated locations.

2.3. Local events

The meta-data attached to tweets can be useful for certain instances of event detection. Pool (2015) and de Kleer (2015) show that using the relatively low frequency of geo-tagged tweets, it is possible to cluster various sorts of events on the local scene, classify them and map where they occur in real-time.

Detecting events using geo-locations from Twitter has previously been done by Walther and Kaisser (2013) and applying a similar approach to Dutch tweets is plausible because the Netherlands has one of the highest twitter accounts to population ratio (Pool 2015).

Figure 1: Visualisation of automatically classified events in The Netherlands, in April 2015 (de Kleer 2015).

All geo-tagged tweets from a month of Dutch Twitter data were used for training. This resulted in a total of 566.549 geo-tagged tweets. The geo-information was translated into a geoHash that denotes a specific area, and tweets with a similar geoHash and comparable timestamp were grouped and added to a list of event candidates. To handle the hard borders of the geohash area, candidates with matching timestamps in adjacent areas were then merged

Two judges annotated the event candidates in the training data and the test data with the following labels: *No event*, *Meeting*, *Entertainment*, *Incident*, *Sport* and *Other*, with the most frequent category being *No event* (triggering a 46% baseline). Inter-annotator agreement was measured via Cohen's Kappa (Cohen 1960). Features that were found to be most useful for this task were the most frequent words, the location using only the first five characters of the geo-Hash, the average word overlap between tweets in an event candidate and the average word overlap between different users in an event candidate.

Several models were built and tested on development data, with the final system being a Naive Bayes model which yielded an accuracy of 84% on test data. Especially considering the inter annotator agreement was measured at $K = 0.87$, this is a very good result. This research also shows that meta-information from tweets can successfully be used to detect events. The end-user output of the system is a map with classified events (Figure 1).

2.4. News

Twitter data has also been used to detect real time commentary on news events. News media websites often feature their own social plugins which allow readers to discuss news items. These discussions are relatively structured and easy to relate to the news article. However, when people post their commentary to Twitter, it becomes problematic to link the tweet back to the article and to group all relevant

discussion together.

In (Kreutz 2015), RSS feeds of the three most popular Dutch news websites are used to detect similar content on Twitter. The RSS feeds give access to 41 headlines and related abstracts at the same time. Each of the news items is then compared to the last hour of Dutch Twitter data to extract reaction, opinions and other meta-commentary that users posted.

To deal with the computational effort involved in making this many comparisons (an hour of Dutch twitter data often contains more than 30,000 tweets), a first module of the system makes a pre-selection of candidates to be considered. The candidates are made up of the 25 tweets with the highest cosine similarity compared to the title of the news items. Generally, these 25 tweets contain tweets that are too similar to the title (a retweet for example), too dissimilar (are not about the article) and actual relevant tweets that add meta-commentary. It is the latter type that one would want to detect and use, while discarding the former two as non-relevant.

To distinguish between relevant and non-relevant tweets, four non linear machine learning algorithms were trained and tested. Four news articles from the 21st of May 2015 were selected to be compared to 24 hours of Dutch Twitter data from the same day. For training and testing, 250 candidate tweets were selected using the approach mentioned above. This cut off point was chosen because after annotation it became clear that for each of the articles, the number of relevant tweets that could be found after rank 250 was negligible.

In the 1,000 candidates, 593 were annotated as relevant to the articles which results in a baseline of 59.3%. The system was trained on six features: (1) the difference in timestamp between the publication of the article and the publication of the tweet, (2) the cosine similarity between the title of the article and the tweet, (3) the difference in length between the title of the article and the tweet, (4) the cosine similarity between the abstract of the article and the tweet, (5) the amount over overlapping Named Entities, (6) the cosine similarity between bigrams in the abstract of the article and the tweet.

Named entities were extracted by means of a Named Entity Recognizer trained on the CoNLL2002 Dutch corpus using NLTK (Bird, Klein, and Loper 2009). A Random Forest classifier performed the best on the test data with an F-score of 0.874. It also helped to determine the most important features in the task. The list of features mentioned above adheres to this order, timestamp difference being the most predictive feature.

The selection of viable candidates before automatic classification proves successful in reducing computational effort, while still keeping the detection of relevant commentary possible. This approach is therefore suitable for a real-time application of the system. demonstrates this by applying the system to a specifically dedicated website that updates its news articles and tweets hourly (`http://www.nieuwstwiets.nl`, Figure 2).
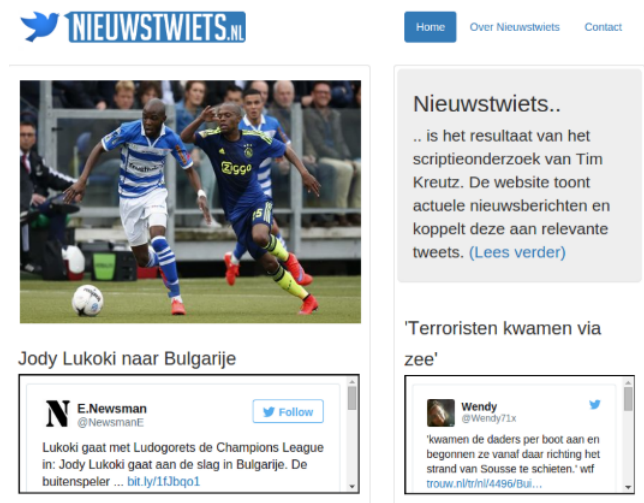

Figure 2: Visualisation of relevant tweet commentary on news events on `nieuwstwiets.nl` (Kreutz 2015).

3. Discussion and Conclusions

In this overview we have reported efforts of bachelor students in the field of automatic event detection exploiting the (Dutch) Twitter stream.

Besides the differences in fields of application, this overview gives insight in the considerations that were made in dealing with the inherent challenges in event detection. For the emergency detection and the detection of local events, geo-information of tweets was used. Since this information is not always available, this sometimes resulted in very little data to work with. For the detection of subevents in soccer, peaks of tweets with certain key patterns were used. This worked well for important subevents (goals) and worse for minor subevents (yellow cards).

The students used similar ways to remove noise from tweets, by removing or replacing URLs and hashtags. Even when hashtags were crucially used to detect events, such as in (Kuiper 2015), they were then normalised at a second stage in order to make the approach general and portable.

Finally, evaluation showed good results in all the theses. Some students chose to apply their findings in a real time setting by visualizing them. This lead to demonstration of the systems by Pool (2015) and de Kleer (2015) on Event-Detective and Kreutz (2015) on `nieuwstwiets.nl`.

With this exercise we observed that real time event detection on social media is a field that students can successfully experiment with. Although the aim was not to build the next generation event detection applications, the choices that the students made in the course of such a research reflect some of the core challenges and considerations central to this task and we believe are useful lessons for future endeavors, both from a research and a teaching perspective.

References

Bird, Steven, Ewan Klein, and Edward Loper (2009). *Natural Language Processing with Python*. O'Reilly Media.

Chakrabarti, Deepayan and Kunal Punera (2011). "Event Summarization Using Tweets." In: *ICWSM* 11, pp. 66–73.

Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales". In: *Educational and Psychological Measurement* 20.1, p. 37.

Corney, David, Carlos Martin, and Ayse Goker (2014). "Spot the Ball: Detecting Sports Events on Twitter". In: *Proceedings of Advances in Information Retrieval: 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands*. Ed. by Maarten et al. Rijke. Springer, pp. 449–454.

de Kleer, David (2015). "EventDetective: detectie, verrijking en visualisatie van Twitter events". Bachelor Thesis in Information Science. University of Groningen.

Hall, Mark et al. (2009). "The WEKA Data Mining Software: An Update". In: *SIGKDD Explor. Newsl.* 11.1, pp. 10–18. ISSN: 1931-0145.

Kreutz, Tim (2015). "Detecting news event commentary on Twitter". Bachelor Thesis in Information Science. University of Groningen.

Kuiper, Jacco (2015). "Real-time automatic detection of soccer match events using Twitter". Bachelor Thesis in Information Science. University of Groningen.

Kuipers, Rolf (2015). "'En we schudden weer'". Bachelor Thesis in Information Science. University of Groningen.

Louwaars, Olivier (2015). "P2000 locatiedata als classifier voor tweets". Bachelor Thesis in Information Science. University of Groningen.

Osborne, Miles et al. (2014). "Real-time detection, tracking, and monitoring of automatically discovered events in social media". In: *Proceedings of ACL 2014: System Demostrations*. Association for Computational Linguistics, pp. 37–42.

Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12, pp. 2825–2830. ISSN: 1532-4435.

Pool, Chris (2015). "Detecting local events in the Twitter stream". Bachelor Thesis in Information Science. University of Groningen.

Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo (2010). "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors". In: *Proceedings of WWW '10*. New York, NY, USA: ACM, pp. 851–860.

Tjong Kim Sang, Erik (2011). "Het gebruik van twitter voor taalkundig onderzoek". In: *TABU: Bulletin voor Taalwetenschap* 39.1/2, pp. 62–72.

Van Oorschot, Guido, Marieke Van Erp, and Chris Dijkshoorn (2012). "Automatic extraction of soccer game events from Twitter". In: *Proceedings of Detection, Representation and Exploitation of Events in the Semantic Web (DeRiVE 2012). Boston, MA, USA. (CEUR Proceedings, 902)*.

Walther, Maximilian and Michael Kaisser (2013). "Geospatial event detection in the twitter stream". In: *Advances in Information Retrieval*. Springer, pp. 356–367.

App2Check: a Machine Learning-based system
for Sentiment Analysis of App Reviews in Italian Language

Emanuele Di Rosa, Alberto Durante

Head of Machine Learning & Semantics Lab at Finsa spa, Assistant Research Scientist at Finsa spa
Via XX Settembre 14, Genova, Italy
E-mail: emanuele.dirosa@finsa.it, alberto.durante@finsa.it

Abstract

Sentiment Analysis has nowadays a crucial role in social media analysis and, more generally, in analysing user opinions about general topics or user reviews about product/services, enabling a huge number of applications. Many methods and software implementing different approaches exist and there is not a clear best approach for Sentiment classification/quantification. We believe that performance reached by machine learning approaches is a key advantage to apply to sentiment analysis in order to reach a performance which is very close to the one obtained by group of humans, who evaluate subjective sentences such as user reviews. In this paper, we present the App2Check system, developed mainly applying supervised learning techniques, and the results of our experimental evaluation, showing that App2Check outperforms state-of-the-art research tools on user reviews in Italian language related to the evaluation of apps published to app stores.

Keywords: Sentiment Analysis, Machine Learning, User Reviews, Italian Language, App2Check, iFeel, SentiStrength.

1. Introduction

Sentiment Analysis has nowadays a crucial role in social media analysis and, more generally, in analysing user opinions about general topics or user reviews about product/services, enabling a huge number of applications. For instance, sentiment analysis can be applied to monitoring the reputation or opinion of a company or a brand with the analysis of reviews of consumer products or services [1]. Moreover, it can also provide analytical perspectives for financial investors who want to discover and respond to market opinions [2,3]. Another important set of applications is in politics, where marketing campaigns are interested in tracking sentiments expressed by voters associated with candidates [4]. Sentiment analysis can also be applied to social platforms to show in real-time what is the opinion of people about emerging events and, in general, named entities, and about the relationships with other events and sources of information. In [5] it is also shown that the growth on the number of searches on the topic according to Google Trends, appears mainly after the popularization of online social networks.

App stores can be seen as another, not yet well explored, field of application of sentiment analysis. Indeed, they are another social media where users can freely express their own opinion through app reviews about a product, i.e. the specific app under evaluation, or a service, to which the considered app is connecting the user (e.g., a mobile banking app connects users to mobile banking services). In addition, reading user reviews on app stores shows that people frequently talk about and evaluate also the brand associated to the app under review: thus, it is possible to extract people opinion about a brand or the sentiment about a company or the provided service quality.

In this paper, we focus on the app store as a social media platform and on the sentiment evaluation in app reviews, which are examples of reviews related to a product, or a service or the associated brand. App reviews are a very interesting application in our opinion because they have not been extensively explored yet [6], and also because the sentiment score detected in a comment can significantly differ from the score assigned by the user to the app under evaluation. For example, a user can assign his good score to the app (i.e. assigning 5 stars) but also express in natural language some suggestions or highlight some minor bugs that do not influence his overall app evaluation. For example, the comment "*Ottima app, per carità, ma effettuando i pagamenti bollettini premarcati si blocca con la scannerizzazione del codice, va in crash e si chiude. Fate qualcosa!!*" was rated 4 stars by the user, but the sentence contains overall a negative sentiment. Vice versa, the user can assign a low rating but highlight some good features. All of this non-structured information is fully missing by only superficially evaluating an app through a 1 to 5 overall score –or any other product evaluated by the user with both sentences and a score –.

About the methods of processing user reviews, many methods and software implementing different approaches exist and there is not a clear best approach for Sentiment classification/quantification [7,8,9]. In [5] it is also shown that more than 7,000 articles have been written about sentiment analysis applying different approaches or slightly different algorithms and various startups are developing tools and strategies to extract sentiments from text. From our side, we believe that performance reached by machine learning approaches is a key advantage to apply to sentiment analysis in order to reach a performance which is very close to the one obtained by group of humans evaluating subjective sentences such as user reviews.

In this paper, we present the App2Check system, developed mainly applying supervised learning techniques and focused – in this first release – on Italian language, and present the results of our experimental evaluation showing that App2Check (version 1.0) outperforms state-of-the-art research software on user reviews in Italian language related to apps (which are, at the moment, our main target application). We considered research tools for our experimental evaluation, since the current state-of-the-art

commercial tools recently included strict restrictions related to the possibility to run them for competitive analysis or benchmarking. In particular, since there are not so many research tools managing natively Italian language, we applied the approach already shown in [7] where the iFeel research platform has been presented. iFeel performs the promising approach to translate sentences into English before running 19 state-of-the-art research tools. In order to make a fair comparison, we also included in our comparison a research tool that natively manages Italian language: to the best of our knowledge, it is the only research tool with this feature that is available for download.

The structure of the paper is the following. After the current introduction about the main paper topics, in section 2, we report a description of the research tools we used to perform the comparison. In section 3, we briefly describe our system App2Check; in section 4, we present and discuss our experimental evaluation and, in section 5, we provide the paper conclusions.

2. State-of-the art Research Tools

In this section, we describe the research tools that will be mentioned in the following sections and included in our experimental evaluation: iFeel, a platform developed at Federal University of Minas Gerais and running 19 research tools, and SentiStrength version for Italian language.

2.1 iFeel

iFeel is a research web platform [10] allowing to run 19 state-of-the art research tools for sentiment analysis on the specified list of sentences. It allows to natively run tools supporting English and to first translate sentences from other languages into English and then run the underlying tools on the English translated sentences. It has been experimentally shown in [6] that well known language specific methods do not have a significant advantage over a simple machine translation approach.

The tools included in iFeel are the following (in alphabetical order): AFINN, Emolex, Emoticon DS, Emoticons, Happiness Index, NRC Hashtag, Opinion Finder, Opinion Lexicon, Panas-t, SANN, SASA, Senticnet, Sentiment140, SentiStrength, SentiWordNet, SO-CAL, Stanford Deep Learning, Umigon, Vader.

2.2 SentiStrength for Italian

SentiStrength was produced as part of the CyberEmotions project, supported by EU FP7. It estimates the strength of positive and negative sentiment in short texts, even for informal language. According to the authors, it has human-level accuracy for short social web texts in English, except political texts [11]. SentiStrength authors make available a version of the tool which natively manages Italian language. All tests have been carried out with both average emotion and strongest emotion options, but in this paper we only report the results obtained with the latter option turned on, due to better performance. Since the English version of SentiStrength is also included in iFeel, we will call the Italian version SentiStrengthIta.

3. App2Check system description

App2Check is our system using an approach in which supervised learning methods are applied in order to build a predictive model for sentiment *quantification*. The training of the model is performed by considering a huge variety of language domains and different kinds of user reviews. App2Check provides, as answer to a sentence in Italian language, a quantification of the sentiment polarity scored from 1 to 5, according to the most recent trend shown in the last sentiment evaluation SemEval [12], where tracks considering quantification have been introduced. Thus, we consider the following quantification: as "positive", sentences with score 4 (positive) or 5 (very positive); as "negative", sentences with score 1 (very negative) or 2 (negative); as "neutral", sentences with score 3. In order to compute the final answer, App2Check does not use just the prediction coming from the predictive model, but it applies also a set of algorithms which take into account some natural language processing techniques, allowing e.g. to also automatically perform topic/named entity extraction.
It is not possible to give more details about the engine due to non-disclosure restrictions.

App2Check is not only constituted by a web service providing access to the sentiment prediction of sentences, but it is also a full user-friendly web application allowing (more features in next release) in the current release 1.0 to:
a) Search for the app a user wants to monitor on the Apple App store, Google Play store or Microsoft Marketplace
b) Show the main topics discussed in user reviews which are both comment-specific, associated to a specific month or evaluated to overall the app life
c) Show the sentiment about the former extracted topics, including in the topics –if discussed in user comments– also the company brand and the provided service level
d) Show a sentiment comparison on the app time horizon between apps owned by different app publishers (even market competitors).

A demo of the App2Check is available after sending a request by email to the paper first author.

4. Experimental Evaluation

In our experimental evaluation we considered user reviews of apps from Apple App store and Google Play store. More specifically, we focused on two different sets of comments. Test set A is made of 10 thousands comments from 10 different very popular apps (one thousand comments per app). These comments are associated only to an overall score for the app, called app rating in the app stores. Test set B is made of 1 thousand comments from the famous Candy Crush Saga app: in this case, we performed a manual quantification (in the 1-5 range) of the sentiment (from now on called *human sentiment classification or **HSC***).

We ran App2Check, iFeel and SentiStrengthIta on these user reviews, in order to evaluate:
- on test set A, their relative performance using the app rating as a reference indicator, i.e. as an approximation of the user sentiment so that we avoid to manually classify the sentiment for 10 thousand comments. Of course, considering a

single comment, as already said, in general, the score/rating expressed by a user respect to an app can be substantially different respect to the sentiment expressed by a human. However, we experienced that the average score/rating of many (hundreds of) comments can be an approximation of the average sentiment expressed by a human on the same set. In Table 1 we show this phenomenon: human sentiment classification (performed by only one person trained with guidelines and examples) agrees with rating on 79.8% of cases with app rating.

- on test set B, the performance of the three systems is compared respect to the sentiment manually classified/quantified by a person on 1 thousand reviews of Candy Crush app (his classification is made publicly available). Thus, in this case we compare systems on a reference that is not approximated.

All of the user reviews are made available together with a limited demo access to the prediction web service, in order to make the experiments repeatable (from here http://demo.finsa.it/app2check/experiments/).

	Precision	Recall	F1
Negative	79.2	92.7	85.4
Neutral	14.6	21.9	17.5
Positive	96.6	78.6	86.7

Table 1: Comparing HSC vs App Rating on **Candy Crush Saga app**: accuracy is **79.8%**, other measures in table.

4.1 Systems comparison on Candy Crush Saga app Reviews

Tool	MF1	Acc	F1(-)	F1(x)	F1(+)
App2Check	**59.2**	**78.3**	**79.2**	14.0	**84.4**
Umigon	47.5	54.2	54.5	**16.7**	71.4
SentiWordNet	47.4	62.5	65.3	6.3	70.7
Sentiment140	47.1	63.6	72.2	3.6	65.6
SentiStrength	46.6	56.8	45.1	18.0	76.7
AFINN	46.5	56.7	51.7	14.0	73.7
Stanford DL	45.9	54.1	62.3	10.6	64.8
Op. Lexicon	45.7	51.6	53.8	15.5	67.7
NRC Hashtag	44.7	59.1	68.6	3.9	61.7
Emolex	39.8	44.6	44.6	13.4	61.4
SASA	37.9	44.7	38.9	13.0	61.8
Vader	37.8	41.1	32.4	16.7	64.1
Senticnet	37.2	53.1	35.9	7.8	67.8
SO-CAL	36.6	39.4	44.4	12.7	52.8
SentiStrengthIta	34.1	39.8	31.3	11.2	59.9
H. Index	29.8	40.1	16.7	11.7	61.1
Emoticon DS	23.8	50.8	2.1	1.7	67.5
Op. Finder	21.2	20.6	22.4	13.8	27.3
SANN	12.8	14.3	1.6	14.0	22.8
Panas-t	5.7	8.1	1.6	13.7	1.8
Emoticons	0.0	7.4	-	13.7	-

Table 2: Comparison of the tools respect to app rating.

In all of the following tables we show macro F1 (MF1), accuracy (Acc), F1 on the negative class (F1(-)), F1 on the neutral class (F1(x)), and F1 on the positive class (F1(+)). We highlight in bold the best value per column. In Table 2 we compare the tools on test set B (1 thousand user reviews from the popular Candy Crush app) with respect to the app rating. It shows that App2Check has the highest macro F1 (59.2%) and the highest accuracy (78.3%), calculated using app rating as a reference. The second and third accuracy is obtained by Sentiment140 and SentiWordNet, respectively. SentiStrengthIta produced a bad performance with respect to the English version of the same tool. In Table 3, we make a comparison with respect to the human sentiment classification. App2Check wins again here, showing the highest macro F1 (65.8%) and accuracy (81.8%); we see that it is even higher than the one calculated in Table 2 using app rating as a reference. This indicates that *App2Check is closer to the human sentiment classification (which is our goal) than to just the app rating*. In Table 3 we can also see that NRC Hashtag has the second accuracy; instead, Sentiment140 and SentiWordNet have the second and third macro F1, respectively. Almost all of the tools show the same pattern and we obtain almost the same chart, thus by confirming that, if we consider hundreds of comments, using app rating becomes –overall and on average– an approximation of the user sentiment. The latter result enables us to use app rating in the following experiments as a reference approximating the sentiment expressed by one single person on the test set.

Tool	MF1	Acc	F1(-)	F1(x)	F1(+)
App2Check	**65.8**	**81.8**	**85.9**	25.2	**86.4**
Sentiment140	58.1	71.6	80.4	21.4	72.5
SentiWordNet	57.2	67.3	73.5	**26.6**	71.4
Stanford DL	53.7	60.5	70.0	21.2	69.9
NRC Hashtag	52.9	81.0	76.4	16.4	66.0
Umigon	50.9	56.4	54.0	20.5	78.1
SentiStrength	50.4	57.8	47.5	25.6	78.0
Op. Lexicon	49.8	54.2	53.1	23.7	72.6
AFINN	49.7	57.4	52.4	21.8	74.9
Vader	40.9	42.8	31.9	22.8	68.1
Senticnet	40.2	50.6	37.4	19.7	63.6
Emolex	40.0	43.3	43.1	16.1	60.9
SASA	39.3	44.3	40.8	15.8	61.3
SO-CAL	39.0	40.8	45.2	17.0	54.8
SentiStrengthIta	38.2	41.5	34.3	20.5	59.7
H. Index	32.3	39.8	16.9	18.6	61.4
Op. Finder	23.6	22.9	24.9	18.1	27.9
Emoticon DS	21.5	41.6	1.8	4.1	58.7
SANN	16.5	17.8	1.8	19.9	27.7
Panas-t	7.4	11.0	1.3	18.7	2.2
Emoticons	0.0	10.3	-	18.6	-

Table 3: Comparison of the tools respect to HSC.

4.2 Systems comparison on 10 thousand user reviews from 10 different apps

In Table 4 we show the results of the systems on 10

thousand reviews, selected considering 1 thousand reviews per each of the following popular apps: Angry Birds, Banco Posta, Facebook, Fruit Ninja, Gmail, Mobile Banking Unicredit, My Vodafone, PayPal, Twitter, Whatsapp. Considering app rating as a reference, we clearly see that App2Check outperforms all of the other tools, reaching an accuracy of about 86%.

Figure 1: Comparison of 4 tools on 10 thousand chronologically sorted user reviews respect to app rating.

Tool	MF1	Acc	F1(-)	F1(x)	F1(+)
App2Check	**73.3**	**85.7**	**82.7**	**45.6**	**91.7**
SentiWordNet	47.9	65.9	60.4	6.2	77.1
AFINN	47.5	60.3	49.2	16.6	76.7
SentiStrength	47.5	59.7	46.3	19.3	76.8
Stanford DL	45.6	54.0	56.5	13.5	66.8
Op. Lexicon	44.9	55.3	45.1	17.5	72.2
Sentiment140	44.1	58.7	57.4	6.7	68.2
Umigon	42.8	50.1	47.8	14.6	66.2
SO-CAL	41.8	49.3	45.8	13.8	65.6
NRC Hashtag	41.2	65.7	53.6	8.3	61.7
Senticnet	40.9	63.1	36.6	9.1	76.9
Vader	38.5	46.2	29.5	19.7	66.3
Emolex	38.3	45.5	38.5	14.1	62.3
SASA	37.7	48.8	29.6	16.4	67.1
H. Index	31.1	39.0	21.9	13.4	57.9
Emoticon DS	27.8	63.6	3.0	2.5	77.8
Op. Finder	26.0	26.6	25.3	15.4	37.2
SentiStrengthIta	23.7	58.3	0.0	0.0	71.0
SANN	12.2	14.9	3.6	16.5	16.4
Panas-t	6.0	9.0	1.2	16.0	0.9
Emoticon	5.4	8.8	0.1	15.9	0.4

Table 4: Comparison of the tools on 10 thousand user reviews from 10 different apps respect to app rating.

In order to better analyze App2Check performance, in Figure 1 we show a plot of the average sentiment per month of all user reviews (1 for positive, 0 for neutral and -1 for negative sentiment). In the plot we include app rating as a reference, App2Check and SentiStrengthIta (SS. ITA in the plot), since they natively support Italian, and the two best tools according to accuracy from Table 4: SentiWordNet (SWN) and NRC Hashtag (NRC H.). It is clear that all of the tools follow quite well the trend of the rating plot. However, NRC Hashtag shows a behavior that is, on average, much closer to the neutral class (so far away from the reference plot). Both SentiStrengthIta and SentiWordNet, instead, are closer each other and to the app rating, but their evaluation is under the reference plot. App2Check is the closest to the app rating, but in certain areas it differs from the rating, especially when the score provided by the user is on average far away from the sentiment expressed in the review. In our opinion, App2Check would have even higher accuracy on these 10 thousand instances, considering as a reference the human sentiment classification: this is made clear while using the web application and evaluating the answer of the system on every single user comment.

5. Conclusion and future work

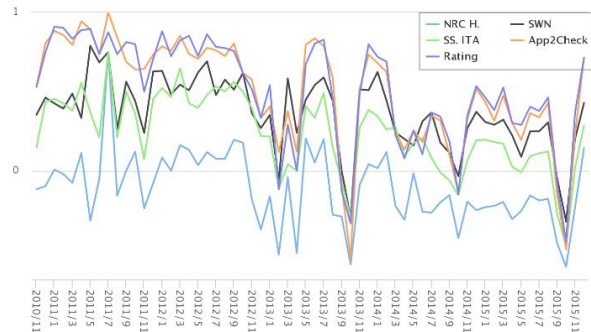In this paper we presented App2Check, a machine learning-based system performing sentiment classification/quantification on user reviews in Italian language. We evaluated it on 11 thousand user reviews related to apps published in app stores. Results show that App2Check outperforms state-of-the-art research tools on this test set. As future work, we want to extend the system to work on more languages and we want to extend the system evaluation on different kind of user reviews and on user feedbacks from Twitter.

6. Bibliographical References

[1] Hu, M., Liu, B. (2004): Mining and summarizing customer reviews. *In Proc. of KDD 2004*, pp. 168–177.

[2] Oliveira, N., Cortez, P., Areal, N. (2013). On the predictability of stock market behaviour using stocktwits sentiment and posting volume. *In Proc. of LNCS 2013*, vol. 8154, pp. 355–365.

[3] Bollen, J., Mao, H., Zeng, X.-J. (2010). Twitter Mood Predicts the Stock Market. *CoRR* abs/1010.3003

[4] Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *In Proc. of ICWSM 2010*.

[5] Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., Benevenuto, F. (2015). A Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods.

[6] Guzman, E., Maalej, W. (2014). How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews. *In Proc. of RE 2014*, pp. 153–162.

[7] Araújo, M., Reis, J. C. S., Pereira, A. C. M., Benevenuto, F. (2016). An Evaluation of Machine Translation for Multilingual Sentence-level Sentiment Analysis. *In Proc. of ACM SAC 2016.* To appear.

[8] Esuli, A., Sebastiani, F., (2010). AI and Opinion Mining, Part 2. IEEE Computer Society, pp. 153–162.

[9] Gao, W., Sebastiani, F. (2015). Tweet Sentiment: From Classification to Quantification. *In Proc. of IEEE/ACM ASONAM 2015*, pp. 97–104.

[10] Araújo, M., Gonçalves, P., Cha, M., Benevenuto, F. (2014). ifeel: A system that compares and combines sentiment analysis methods. *In Proc. of ACM WWW 2014*, pp. 75–78.

[11] Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., Kappas, A. (2010). Sentiment strength detection in short informal text. *ASIST journal*, 61(12), pp. 2544–2558.

[12] SemEval-2016 Task 4: Sentiment Analysis in Twitter.

Distrusting Science on Communication Platforms: Socio-anthropological Aspects of the Science-Society Dialectic within a Phytosanitary Emergency

Christian Colella
christian.colella@edu.unito.it

Abstract

The work aims to investigate the conspiracy-like and pseudo-scientific beliefs arose in Salento during a spread of a plant disease that affected olive tree crops known as "OQDS" and, more generally, tries to analyze through a socio-anthropological perspective the communication biases into the dialectical relationship between scientific research and general public and how social media platforms act like a conceptual container for pseudo-scientific belief and distrust sentiments toward science research.

Keywords: Pseudo-science, Conspiracy Theory, Cultural Identity, Cultural Analysis of Social Media, Social Media as Social Sensor

1. Introduction

This case study takes place in Salento, a peninsula located in the south-east part of Italy. This region is crowded with 20,000,000 olive trees. More than 60% of this "peninsula inside a peninsula" is covered by these ancient fruit trees, like sculptures that time and nature have forged during the centuries.

As often happens, some ecological elements belonging to the domain of nature are charged with deep cultural and symbolic meanings. That is the case of the olive tree in Salento.

In this ecological and semantic context, the olive tree is not just a plant and it is not just an important source of revenue in the center of the traditional food production of the region. Here, olive tree crops are also a part of the human-made panorama, a useful compass to navigate into the aesthetic landscape, a statement of cultural identity and cultural integrity.

It is not surprising that the reactions following the decision to eradicate the infected trees as necessary and functional solution to halt the spread of the bacterium "Xylella Fastidiosa" believed to be the cause of the olive quick decline syndrome (OQDS) and responsible for the death of some olive trees has been particularly hated by the public opinion.

The phenomenon started to appear in the hinterland of Gallipoli, the western part of the Salento peninsula, during the late 2008. Since then, outbreaks of this highly virulent plant disease were reported on several olive trees in the Salento aggravated by the particular mild climatic conditions of the winter 2013-2014. From then on, scientific research did anything possible to tackle the spread of this plague, but the scientific process, like other human matters, is not instant and straightaway as it takes time.

The sudden awareness of the perceived disaster, together with the widespread belief that it was all a staged sham concocted by unknown and high powers, allowed this discourse on the phenomenon to find a privileged place on the social platforms, causing conspiracy-like and pseudo-scientific distortion of the actual reality of the situation, with consequences that impacted even outside of the social media informative sphere.

2. OQDS and Online Social Networks

The distrust and the general suspicion toward scientific research institutions linked to the diffusion of this plant disease begins in March 2015 when Sabina Guzzanti, an Italian satirical comedian wrote on her Facebook page about the "strange correlation" in the explanation of this plant disease (Fig.1).

The famous television personality, in her post, argued that "Xylella Fastidiosa" was nothing more than a clever fabrication, beginning to connect various causes that could have been behind the spread of the "fake" pathogen. She immediately refers to the American multinational agro-chemical and agricultural biotechnology corporation Monsanto as the main culprit. Her belief is that the bacterium "Xylella fastidiosa" has been created in a laboratory by the same American company with the aim of selling insecticides and, as a last resort, to sell GMOs olive trees immune to the bacterium.

The post suggest that the actual name of the pathogen "Xylella" was actually the anagram of "Allelyx", a Brazilian multinational belonging to Monsanto guilty of having created the bacterium in laboratory, all with the blessing and cooperation of the Italian and Salentinian scientific research institutes.

This dystopian-sci-fi-like scenario evoked in the before mentioned Facebook post, is considered to be a crucial starting point for analyzing how the pseudo-scientific and anti-scientific beliefs started and became widespread in Salento.
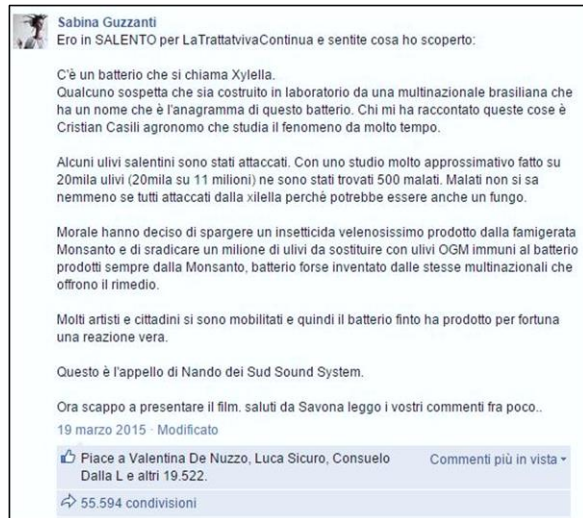

Sabina Guzzanti
Ero in SALENTO per LaTrattatvivaContinua e sentite cosa ho scoperto:

C'è un batterio che si chiama Xylella.
Qualcuno sospetta che sia costruito in laboratorio da una multinazionale brasiliana che ha un nome che è l'anagramma di questo batterio. Chi mi ha raccontato queste cose è Cristian Casili agronomo che studia il fenomeno da molto tempo.

Alcuni ulivi salentini sono stati attaccati. Con uno studio molto approssimativo fatto su 20mila ulivi (20mila su 11 milioni) ne sono stati trovati 500 malati. Malati non si sa nemmeno se tutti attaccati dalla xilella perchè potrebbe essere anche un fungo.

Morale hanno deciso di spargere un insetticida velenosissimo prodotto dalla famigerata Monsanto e di sradicare un milione di ulivi da sostituire con ulivi OGM immuni al batterio prodotti sempre dalla Monsanto, batterio forse inventato dalle stesse multinazionali che offrono il rimedio.

Molti artisti e cittadini si sono mobilitati e quindi il batterio finto ha prodotto per fortuna una reazione vera.

Questo è l'appello di Nando dei Sud Sound System.

Ora scappo a presentare il film. saluti da Savona leggo i vostri commenti fra poco..
19 marzo 2015 · Modificato

Piace a Valentina De Nuzzo, Luca Sicuro, Consuelo   Commenti più in vista ▾
Dalla L e altri 19.522.

55.594 condivisioni

Fig. 1 Sabina Guzzanti's Facebook Post.

With almost twenty thousand "likes" and more than fifty-five thousand shares, the idea of a malevolent intention ahead of the diffusion of this plant pathology spreads among Facebook users and, at a local level, it resulted in a general suspicion towards the work of scientific institutions engaged in the study of the bacterium and the grounds of olive tree desiccation.

This malicious association and suspicion became established due to the presence of two different plains of discourse concerning the particular socio-anthropological context of Salento and the peculiarities of the social medium in which these discourses take place.

The first one is the mythologizing exaltation of local identity within the peculiarity of the local cultural heritage and its public narratives. The rural origins in the identity-making discourse in this region is a very important matter. An idea carried out mainly by local music performer and cultural media (which find a privileged place on the most popular social media platforms) which have over time idealized and transformed the meaning of representation taking it from a mere geographical belonging to a more

comprehensive, pervasive and masterfully defined cultural identity. In such romantically portrayed idea of wholeness of nature and culture, as closely related to one another, the perceived danger of losing the most significant representing symbolic element, the olive tree, in a land full of identity exaltation tied to the idea of unspoiled nature, is obviously seen as a disaster of epic proportions. In addition, recent revisionist drifts about the historical process of "Risorgimento", the Italian political unification process, have brought forward an idea of a supposed hidden quasi-enslavement project carried out by the neo-industrial northern region against the still rural south.

It seems superfluous to mention that such ideas have quickly gained popularity in some socio-cultural circles that meet on social media platforms.

The second aspect to be considered in the public misunderstanding of the role, as well as the aims of scientific research in the case of olive quick decline syndrome in Salento, is the unique social media behavior of "local" conspiracy theorist and their pseudo-scientific approach observed. The same social subjects that showed more distrust and suspicions regarding the causes of this plant pathology are usually pseudoscientific and conspiracy-like Facebook pages and groups members. Themes like the preservation of the authenticity of nature from the presumed dangers of modernity are central to this view and, in this case, the threatening scientific work as the executive arm of corporations and a higher political power.

During the research we noticed that the users engaged in the debate about the causes of olive tree decline syndrome tent also to be, through their Facebook activity, leaned towards other pseudoscientific and conspiracy-like themes, for example the skepticism about GMOs, climate change denialism, "chemtrails conspiracy" up to the "Illuminati conspiracy".

The particular behavior on social media platforms like Facebook of these subjects, prone to a brisk production of useful material to analyze from a socio-anthropological point of view, makes in some way easier the difficult and prolonged action of collecting human data through a face to face ethnographic fieldwork. Their activity is all-encompassing, commenting on the news and the latest developments on the Xylella/OQDS case arguing explanations perfectly suited to their conspiracy theorist habitus (Bourdieu, 1979); accusing the scientific field of inherent evaluation errors, putting the discourse on a "clash" level with two different *modus operandi*, a clash between two different knowledge. "Our" science, validated by time and culture, against "Their" science, fueled by the lust for power, in a strictly dichotomous semantic frame. (fig.2) In their particular views, the work of scientific institutions on OQDS is inevitably compromised by the

corrupt nature of the scientists, who are seen as servants of political and economic power, directly accusing individual scientists even on a personal level, not as a category then, but also as individuals. As Fredric Jameson has argued, the association science-corporate power made by conspiracy theorists is the poor person's cognitive mapping in the postmodern age, it is the degraded figures of the total logic of late capital. (Jameson, 1990).

Fig. 2 "These 'scientist' talk about Xylella epidemic in Salento without ever presenting evidence!" "They condemned to death our olive trees WITHOUT EVIDENCE! This is war!" Image taken from "Terroni di Pino Aprile" Facebook page.

In this particular view, the concept of discovery itself, inherently social, is something to be frightened of. Both the glorification of cultural heritage end cultural identity discourse and the pseudoscientific and conspiracy argumentative logic and social behavior, find their respective and privileged place on social media platforms.

This idiosyncratic theme of a plant disease that has struck a crucial identity and identifying natural element, along with the presence of the agency of scientific research, has allowed these two conceptual universes to come together in a common cause.

Their effective presence, pervasiveness and hyperactivity on social platforms like Facebook has meant that the case itself reached a large audience causing an incontrollable exaggeration of the real issue. Soon people start to think of an inevitable and unstoppable environmental disaster of enormous proportions when, in reality of data, the abatement plan with the aim to contain the bacteria causing the OQDS touched only a 0,09% of the total olive tree population. The "attack on the forest" portrayed by these social groups was apparently unfounded.

The instances and fears of these groups, gathered on social media, reached escalated levels when the issue was addressed by mainstream media and particularly, when a popular Infotainment television show in Italy well known for its distinctive media populism and audience-making sensationalism, endorsed an idea of a compromised, if not powerless and ultimately ineffective, science research process, instilling in the broad general public the belief of a scientific underestimation error at the expense of the local farmers and land owners.

The last, and perhaps the most paradoxical implication of the public negative opinion toward the scientific and institutional efforts deployed in the "Xylella emergency" took place on a judicial level, when the regional administrative court (TAR) accepted the legal recourse brought by environmental groups (as an almost "natural" conclusion of this "gish gallop") on the supposed origins of the spreading of this plant disease. Their allegation, upheld by a local judge, bring us back to an event that took place in Bari during October 2010, when the Mediterranean Agronomic Institute (IAM) along with the National Research Council (CNR) held a workshop entitled "Phytosanitary Workshop on the Quarantine Pathogen Xylella fastidiosa" as part of the research activities and cooperation of the action program "EU-COST 873 project Bacterial Diseases of stone fruit and nuts", a short course on the most advanced methods for the diagnosis of quarantine pathogen Xylella fastidiosa. The allegations proposed by these social media-coordinated environmental groups and supported by the court, accused the entire local scientific research apparatus of intentionally spreading the bacterium, leading a biological warfare and having created the harmful bacteria with the pernicious complicity of Monsanto. As stated earlier in this paper, a convincing sci-fi-like dystopian scenario.

Up to this point we have tried to describe the subjects involved in the case of OQDS/Xylella Fastidiosa public reactions, how these reactions that bordered collective paranoia started on and found a privileged place on social media platforms, and how the medium of Facebook magnified the issue even bringing it on the judges' desks unlashing a collective phobia on scientific research aims and methods.

At this point, we would like to proceed with the description and explanation of the peculiar qualitative methodology used in this work in progress.

3. Methodology

The majority of qualitative research carried out in this work consisted in the usual methods of investigation belonging to the demographical, sociological and anthropological disciplines, proceeding with a (physical) fieldwork lasted about five months, during which we attended public events, public demonstrations and conferences held by the protagonist of this "resistance" and proceeding with ethnographic interviews with several individual components.

But, for what concerns the focus of this paper, we tried to perform a methodological shifting of the very physical nature of the anthropological notion of "field". Social media platforms behave as actual, tangible containers of cultures. The "anthropopoiesis" process (From ancient Greek *Ánthrōpos*, "man"; and *Poïesis*, "to make") (Herder, 1968; Geertz, 1964; Remotti 2013) as the self-building practice of social and individual cultural identity, is in some ways, clearly revealed by the showcase of our Facebook profiles, at least depending to what we want to show, or better, depending on what we plan to show.

On social media platforms, our cultural identities are forged constantly and meticulously through a regular and astonishingly easy work, our constructive nature and disposition made explicit by the simple act of clicking on a like button.

The central issue addressed here is that, through a socio-anthropological lens, a lot can be understood about the single individual just observing his/her Facebook activity.

Referring to the "naturalistic observation method", a study method that involves watching subjects' behaviors in their natural environment, without intervention (Adler & Adler, 1994), and taking as a "natural environment" (and therefore as a field) Facebook social network, I interpreted the social media platform as a "public realm" (Goffman, 1971).

But Facebook is a public and available space up to a certain point.

In order to engage entirely into the social bubble where this kind of cultural discourse is produced and diffused and given the distinct nature of the pseudoscientific and conspiracy individual behavior, we proceeded with interweaving a network of Facebook contacts with these subjects by creating a fake account with the same "Anthropopoietic characteristics".

Furthermore, with the process of "hiding" the researcher's anthropoietic characteristic, we proceed with what we could define as "hided participant observation" (Delamont, 2004; Humphreys, 1975), creating not only a new Facebook identity, but providing to Facebook algorithm new directives, allowing our brand new filter bubble (Pariser, 2011) to see us as "conspiracy theorist".

This allowed us to fully enter in this particular "echo chamber" (Quattrociocchi et al., 2015) observing and occasionally interacting with these same social subjects from an unusual and privileged position.

4. Conclusions

This singular social research method, essentially an "ethnographic lurking" allowed us to conduct the research from a privileged non-physical location.

Creating from scratch an ideal anthropological habitus we were able to join the community, it did not take long to become legitimate conspiracy theorists and have direct access to a substantial amount of useful qualitative data.

Moreover, once started the process, Facebook algorithm tents to suggest and direct every single user toward anthropopoietical choices that perfectly suits the latter's interests, aims and purposes, thus strengthening the fictitious boundaries of the personal social bubble. The peculiar social media *modus operandi* of these subjects, a conspiracy theorist and pseudoscientific *habitus*, in which tends to be central the constant and spasmodic social media activity (posts, notes, other type of information sharing) almost take the form and emerge as a mission, "to wake people up", "to enlighten", allow the social scientist to grasp onto their deep cultural meanings and social behaviors with less efforts. The particular methodology that has characterized this part of the work led us to rethink the epistemological concept of fieldwork, in an anthropological sense, as an ethnographic tool.

The ethnographic field, as a physical place in which identity and culture are produced and shaped, has shifted and changed in both space and time.

Cultural identities, as in the case of this public misunderstanding of science, are expressed and then delivered, their examination, immediate.

5. References

Adler, Patricia & Adler, Peter (1994): *"Observational Techniques"*, in: Denzin e Lincoln [1994, 377-392].

Bessi, Alessandro, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. (2015) "*Viral Misinformation: The Role of Homophily and Polarization*." Proceedings of the 24th International

Conference on World Wide Web - WWW '15 Companion.

Bessi, Alessandro, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. "*Science vs Conspiracy: Collective Narratives in the Age of Misinformation*." PLOS ONE 10.2 (2015): n. pag. Web.

Bourdieu, Pierre, and Richard Nice (1979): "*Outline of a Theory of Practice*". Cambridge Etc.: Cambridge UP.

Delamont, Sara (2004): "*Ethnography and Participant Observation*", in C.Seale, G.Gobo, J.F.Gubrium & D.Silverman. "Qualitative Research Practice, London, Sage, p.p. 217-229.

Herder, Johann Gottfried, and Frank Edward Manuel. (1968): "*Reflections on the Philosophy of the History of Mankind.*" Chicago: University of Chicago.

Jameson, Fredric (1990): "*Cognitive Mapping*". In: Nelson, C. Grossberg, L. [ed]. "Marxism and the Interpretation of Culture" University of Illinois Press

Geertz, Clifford (1963) "*The Transition to Humanity*". Washington, D.C.: Voice of America, U.S. Information Agency.

Mocanu, Delia, Luca Rossi, Qian Zhang, Marton Karsai, and Walter Quattrociocchi. (2015) "*Collective Attention in the Age of (mis)information*." Computers in Human Behavior 51: 1198-204. Web.

Pariser, Eli (2011) "*The Filter Bubble: What the Internet Is Hiding from You*". New York: Penguin.

Quattrociocchi Walter, Fabiana Zollo, Alessandro Bessi, Michela Vicario, Antonio Scala, Guido Cardarelli, Luis Shakhtman, and Shlomo Havlin. (2015): "*Debunking in a World of Tribes*."

Remotti, Francesco (2010): "*L'ossessione Identitaria*". Roma; Bari: Laterza.

Remotti, Francesco (2010): "*Fare Umanità: I Drammi Dell'antropo-poiesi*". Bari: Laterza.

Tracking and Analyzing the "Second Life" of TV Content: a Media and Social-driven Framework

L. Vignaroli[1], C. Schifanella[1], K. S. Candan[2], R. Pensa[3], M.L. Sapino[3]

[1]RAI – Centre for Research and Technological Innovation, Turin, Italy
{luca.vignaroli, claudio.schifanella}@rai.it

[2]Arizona State University
candan@asu.edu

[3]Department of Compute Science, University of Turin
{pensa, mlsapino}@di.unito.it

Abstract

People on the Web talk about television. TV users' social activities implicitly connect the concepts referred to by videos, news, comments, and posts. The strength of such connections may change as the perception of users on the Web changes over time. With the goal of leveraging users' social activities to better understand how TV programs are perceived by the TV public and how the users' interests evolve in time, in this work, a framework that allows to manage, explore and analyze the heterogeneous and dynamic data coming from different information sources which play a role in what we call the "second life" of TV content will be exposed

Keywords: social network analysis, social TV, second screen applications

1. Introduction

In recent years, the way users watch television has radically changed. With the introduction of digital television and the growing number of generic and thematic channels the final user tends to use new forms of navigation in the television content space. To help users' navigation, broadcasters provide new enriched metadata services such as EPGs (Electronic Program Guides) which accurately describe scheduled programs. More in general, traditional broadcasters are moving to become "digital media companies" making their content available not only on the traditional TV channels, but, in a multi-platform perspective, the introduction of OTT (Over The Top) services for these new companies is a must to bring its content offering to the needs and expectation of the audience. Also the home environment is changing since many smart users watch television while using a notebook, smartphone or a tablet as secondary screen more or less related to the broadcast programmes. At the same time, social networks allow the final user to be immersed in a collaborative environment and to talk about television. TV users' social activities implicitly make connections between concepts by means of videos, news, comments, and posts. The strength of such connections may change as the perception of users on the Web changes over time. Moreover, user-generated contents (UGC) are revolutionizing all phases of the content production value chain, in particular it can be observed that a very large number of UGCs include significant portions of content already broadcast by the TV networks. In this context a number of Social TV applications are emerging, providing to the final user tools for social interaction while watching television or media content related to a particular TV program. If properly leveraged, these collaborative social environments can be seen as rich information data sources, indirectly returning to broadcasters and content producers some form of implicit feedback from the final users.

In general, television content evolves in time and its life undergoes a number of different steps. Firstly, a content is typically produced and put on air by (e.g.) the broadcaster. In addition, a copy of it, enriched with its description together with a collection of related metadata is (statically) stored in the TV archive to be reused if needed, and the broadcaster puts the description in his EPG. Big broadcasters also make their TV contents available in the Internet site or as OTT service. Secondly, after the on-air time, the broadcaster is interested in estimating to which degree users are satisfied with the broadcasted content by means of quantifiable data, such as TV audience measurements, or the number of user views in the online resource.

This data collection concludes the "on-air phase" of the content, whose life most likely spans much beyond that point.

In fact, successful TV programs will be probably commented on Twitter and Facebook and published (either entirely, or more often in part) online by users, for example on YouTube. The online posting of (a fragment of) a TV program starts a second phase for it, which can be called "on-line phase".

During this on-line phase the content will be watched, tagged, liked, commented and shared again and again by users in the network. The television content turns to be a "magnet" for users in the network attracting other users, and it becomes a "Social Object".
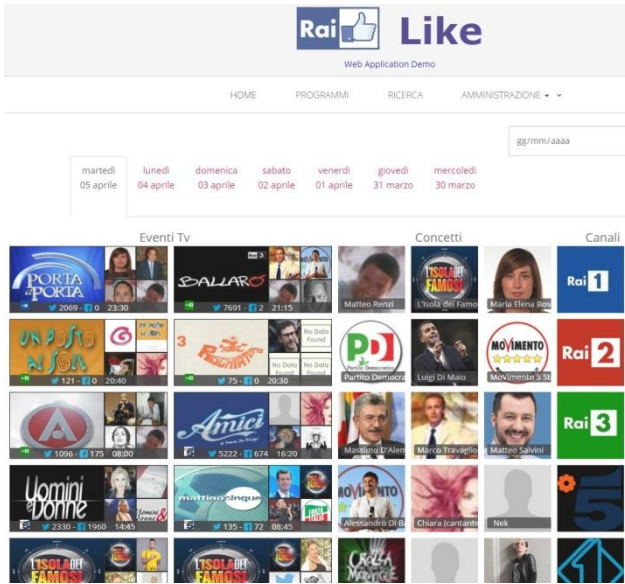
Figure 1: home page of the application demo

Figure 2: Analyzing interaction graph communities

Figure 3: Migration of users between different TV programs

Typically, YouTube is the first place where people come to look for a television content which they missed and despite the copyright issues this might raise, this is a fact. Instead of contrasting this tendency, content producers can try to leverage it to their benefit. In fact, users that post TV content in YouTube are implicitly disseminating, describing and publishing it for free.

As an example of the "extended" life of television content, it can be observed that it is very easy to find in YouTube segments of TV programs that have been uploaded long time ago, even years ago, and are still nowadays very often watched, commented and liked. Without video-sharing platforms, the TV content would otherwise be just stored in the archives of the broadcaster and it would be inaccessible to users. As time passes and the users' social context changes, the way any specific television content is perceived also evolves. For example, a content can attract a new community of users interested in it, or it might change its own meaning because of a new fact happened in the world. If timely discovered these phenomena could be leveraged by the broadcaster and some of the contents already available in the archives could be considered for the production of new programmes based on new interests of the public.

In this context, one of the main objectives of the project is to capture how can the TV content evolve and detecting which phenomena can emerge from the contents' evolution.

2.   The second life of TV content

As a result, the Rai Research Centre has been developed a framework for enabling the integration of heterogeneous data coming from the knowledge sources (broadcasters' archives, EPGs, collected audience data, social networks, etc.) which play a role in the "second life" of TV content, starting from its production phase, going through the on-air phase, and continuing with the on-line phase. The system is designed to provide powerful tools helping to highlight the tight interactions between the Web world and the TV world.
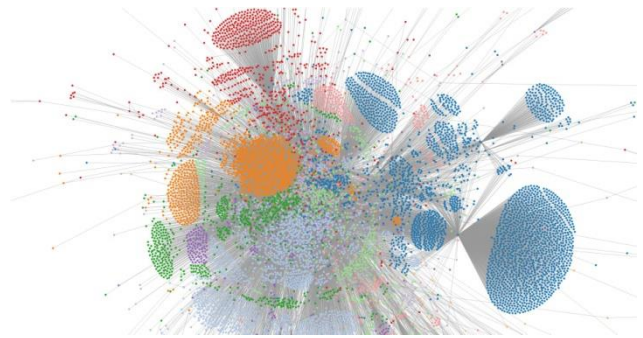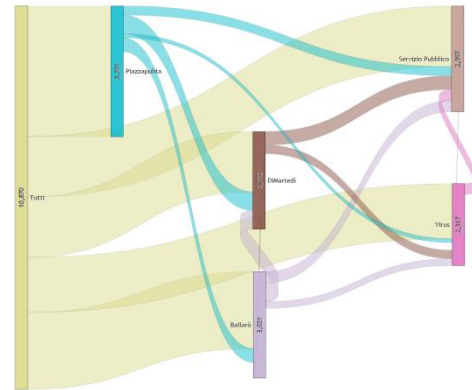
From a more technical point of view the framework enables the integration of various information sources into a unique "knowledge base", modelled as a knowledge graph. Integrating domain and general purpose ontologies, as well as social interactions among users and social media, the knowledge base can be queried and analyzed as a whole, enabling the discovery of new and interesting cross-domain patterns.

The integration framework consists of three main layers: a source processing layer, a knowledge graph layer and a knowledge query and analysis layer.

The "source processing layer" has the role of collecting all the data which will be conveyed in the model. It accesses a number of predefined web/social/media sources (e.g., broadcasters official web sites, social networks, TV channels, etc) and processes them in order to extract those information units which will be represented as nodes in the knowledge graph, as well as information that support the existence of relationships among them.

The "knowledge layer" is the core of the system and represents the result of public actions of users in social environments. Furthermore, relationships are introduced between subjects and social objects and between social objects and concepts. Other relationships involve entities of the same type and these are called structural dependencies. Moreover, social objects evolve in time, hence, as a special case of representation relationship, we consider the temporal representation of a social object against a special type of concept called time objects.

18

Finally, the "knowledge query and analysis layer" consists in a set of components for querying, browsing and analyzing the knowledge graph. A query module extracts sub-graphs from the knowledge layer based on user's requests and constraints. An analysis module provides a set of analysis tools to obtain insights of the data.

The development of the system had to face some challenges in order to make manageable the analysis over the huge amount of data which is being collected from the various knowledge sources. The optimisation of algorithms and the definition of the data model has been done in close cooperation with the Department of Computer Science at the University of Torino.

The interactive demo (Figure 1), thanks to powerful and flexible analysis tools (Figure 2,3) of the social network data flow, allows to make available to Rai a number of strategic services such as user behaviour profiling, brand reputation, community detection and recommendation systems for contents and advertisements.

3. Bibliographical References

A. Antonini, L. Vignaroli, C. Schifanella, R. G. Pensa, and M. L. Sapino, "Mesoontv: a media and social-driven ontology-based TV knowledge management system," in Proceedings of ACM HT '13. ACM, 2013, pp. 208–213.

A. Antonini, R. G. Pensa, M. L. Sapino, C. Schifanella, R. T. Prioletti, and L. Vignaroli, "Tracking and analyzing TV content on the web through social and ontological knowledge," in Proceedings of EuroITV '13. ACM, 2013, pp. 13–22.

X. Han, W. Wei, C. Miao, J. Mei, and H. Song, "Context-aware personal information retrieval from multiple social networks," IEEE Comp. Int. Mag., vol. 9, no. 2, pp. 18–28, 2014. [Online]. Available: http://dx.doi.org/10.1109/MCI.2014.2307222.