Determining Sentiment about Venues on Social Media with non-linguistic cues

Annotation and Classification Experiments

Draft Documentation
v.1 - 28/04/2014
Contributors: Francesca Frontini, Clara Bacciu, Valeria Quochi, Davide Gazzè

TOC

1. Introduction

Social media have particular relevance and impact for the tourism domain and industry, as they enable tourists to share information, content (video, pictures), tips, and opinions about their experiences. Therefore, social media are used in marketing of tourism destination both to promote and to investigate users' needs/expectations.

Analyzing user opinions on social media has therefore a high potential impact on marketing strategies and decision making procedures especially for customer service, as it allows the management to react and respond to user feedback and thus increase user satisfaction.
Online consumer reviews are the most accessible and prevalent form of electronic word-of-mouth [3]. However, within social networks other additional mechanisms are available to express sentiment/support/satisfaction which, when adequately analyzed, may provide useful additional and complementary information to textual sentiment analysis.

Currently, sentiment analysis is typically performed on textual data or on explicit rating systems. Social media however display several other non-textual features that reflect user and stakeholder's behavior, which are currently used to compute metrics for influence, reputation and similar interesting metrics for business intelligence and marketing.

The hypothesis explored in the present experiments is that such "social" features can be used to compute also a metrics for social sentiment based on non-textual information. The basic idea is thus to obtain a general likeability score that may be calculated for single sources or for entire geographic areas, which provides end users with useful insights on their business.

In order to do that, we adopt a machine learning classification approach and build a Sentiment on Social Media Module (SSM) that classifies the "likeability" of venue pages on the basis of solely non-linguistic features. Specifically, it should take in input data concerning accommodations (such as hotels) collected from the APIs of social media such as Facebook or Foursquare and classify them based on quantitative measures according to the positive/negative sentiment that such pages convey.

The overall background and assumptions for the present tool can be found in OpeNER D4.21 "Definition of indicators and metrics for Sentiment Analysis"[1]. In this document the annotation process is outlined, the classification experiments are described and some preliminary evaluation is provided. For the purpose of this documentation, some general background knowledge is repeated here:

An **entity** is an abstract concept which represents something in the real world (e.g. resource, event, person, organization, brand, etc.), and is characterized by a set of attributes (e.g. id, name, country, etc.). The id specifies the unique identifier for the entity, the name specifies the entity name (e.g. "Hotel Bologna" or "Pizzeria Reginella"), the country specifies its country of origin. Finally, an entity is characterized by one or more tags. A tag (or category) specifies the category which the entity belongs to. An entity can belong to one or more categories. For example, the tags related to the entity "Hotel Bologna" could be "accommodation" or "tourism". With reference to the tourism domain, an entity is specifically a resource related to accommodation (i.e. hotel), i.e. a venue.

A **channel** is defined as a web site where each entity can perform one or more of the following operations: a) build its profile, b) share its profile with other individuals, c) communicate with other entities. Facebook and Foursquare are channels. The practical possible ways of communication vary according to the specific channel. In practice, once its own profile is built, an entity can establish relationships with other entities and interact with them. On a channel an entity may exploit many sources to express and spread information, opinions, (personal) news and thoughts.

[1] http://www.opener-project.org/deliverables/

A **source** is a web place providing a mechanism to express interests, publish news about themselves and establish relationships with other entities, etc.. A source is represented by an id, varying from source to source, and a type, which specifies the nature of the source. A source usually provides the collection of informative social metrics/indicators which can be used to establish the impact a given entity has on the channel. A page on Facebook is a source.

By **social metric** we denote a mechanisms giving statistical information about the entity associated to that source. An example of social metric could be the Facebook number of fans, which is the number of people who like the source associated to that entity. The higher the number of fans, the higher the probability that the entity can influence them. For example, the number of fans is commonly used as metric to assess the popularity of an entity on the channel.

The SSM classifier has been developed within the framework of OpeNER, a EU FP7 funded project that provides Cross-lingual NERC and Sentiment Analysis. The goal of the classifier is to provide a counterpart to the linguistic sentiment analysis pipeline; while the sentiment analysis tools can extract the overall sentiment for a given source (i.e. a venue on a certain channel) by aggregating the sentiment expressed in textual data of all the analysed reviews, the Sentiment on Social Media classifier attempts to derive the overall sentiment regarding a source by aggregating the social metrics (such as likes, number of posts, presence of photos, ...) of the source itself, as collected through the the channel’s API service.

The tool thus attempts to replicate the judgment of a human, a potential customer, when assessing the likeability of the entity or venue based on the information presented in the source. This can potentially allow hotel owners whose venue is represented in different sources on different channels to assess how well they are represented on each channel.

In order to obtain such a result, a machine learning approach was chosen, and a human annotation was performed in order to train models for the different sources on a human assessed set of sources.

In Section 2 of this document we describe how we performed the annotation in order to obtain the training set;
In Section 3 we describe the collection of the data from the channels, the training experiment and the test results;
In Section 4 the prototype is briefly described;
In Section 5 a link to the test application is provided.

2. Annotation interface and results

An annotation experiment was carried out on sources from two channels, Foursquare and Facebook, using two human annotators. The goal of this first experiment was to establish an annotation protocol and guidelines for the creation of adequate training sets for the classifier, and to assess the human agreement on the task.

2.1 Annotation criteria

The annotator is instructed to assume the **point of view** of a potential customer who is organizing a holiday, assuming that (s)he targets the location and accomodation type of the venue to be annotated. Would he/she take this hotel into consideration or recommend it (*recommendation*), and to which degree (*liking*)?

The **objective** is to obtain a manually annotated dataset of venues from social media assessed on the basis of two criteria: the first criterion is binary, the second is an appreciation judgement on a 5 point scale (with the middle value representing a neuter judgement). Such dataset is then used to train the classifiers for the social media in question.

Thus, two types of annotation have been performed at the same time:

1. Recommendation (*Would you choose this venue, or recommend it to a friend? Yes /No*)
2. Liking (appreciation rating)
   1 = very bad, would never go or recommend
   2= bad, would not go/recommend
   3 = ok, but nothing extraordinary (neutral opinion)
   4 = good, I like this place and would recommend it
   5= very good, I would definitely go here and certainly recommend it

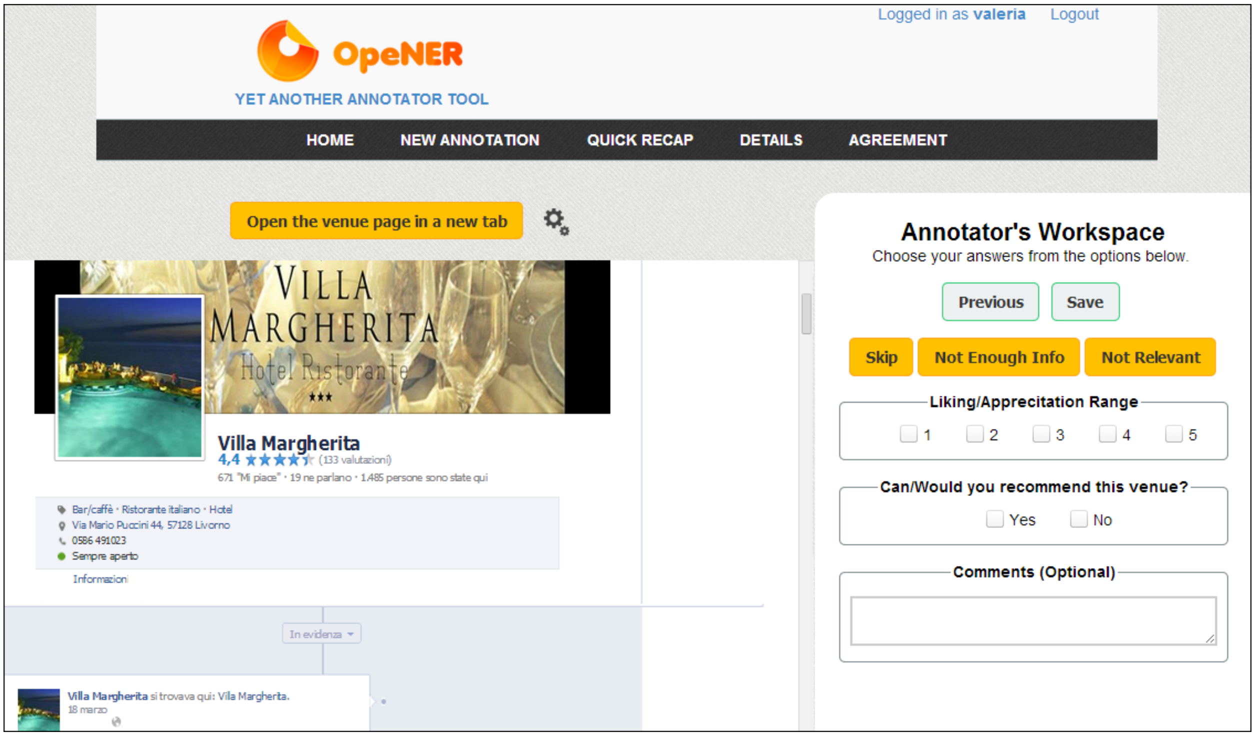
Before annotating/rating the venue, the annotator is asked to skim sources that have been incorrectly collected (anything that is not a hotel, for this experiment) and to mark sources that have not not enough info (e.g. zero or very few visitors or check-ins on Foursquare, zero or very few likes, check-ins, or posts on Facebook). The annotator is also asked to judge the venue on the basis of non-textual cues on the pages.

2.2 Application for the annotation

To carry out the annotation task, an ad-hoc web application was developed. It lets the user log in and see the details of his/hers previous activity: the number of annotated sources, their distribution over the available channels, statistics on the number of discarded sources and so on. Users can choose a channel and the location of the venue to which the source to be annotated must refer. A new source is then chosen randomly among the ones satisfying user's criteria, so that different users see all the sources in a different order.

The annotation interface is shown in the figure below. On the left, the user can see the sources as they are shown on the chosen channel. On the right, there is instead the annotation form. Besides the liking rating and the recommendation, each source can be skipped and left for a later annotation, marked as not having enough information to decide, marked as not relevant for the task (i.e. it refers to a bar, or a restaurant instead of an accommodation). A comment can be left by the user, that can be useful later when revising the results.

We plan to make the annotation tool available, so that similar training sets may be created for other channels.


Logged in as valeria
Logout
OpeNER
YET ANOTHER ANNOTATOR TOOL
HOME
NEW ANNOTATION
QUICK RECAP
DETAILS
AGREEMENT
Open the venue page in a new tab
VILLA
MARGHERITA
Hotel Ristorante
★★★
Villa Margherita
4,4 ★★★★★ (133 valutazioni)
671 "Mi piace" · 19 ne parlano · 1.485 persone sono state qui
Bar/caffè · Ristorante italiano · Hotel
Via Mario Puccini 44, 57128 Livorno
0586 491023
Sempre aperto
Informazioni
In evidenza
Villa Margherita si trovava qui: Vila Margherita.
18 marzo
Annotator's Workspace
Choose your answers from the options below.
Previous
Save
Skip
Not Enough Info
Not Relevant
Liking/Appreciation Range
□ 1 □ 2 □ 3 □ 4 □ 5
Can/Would you recommend this venue?
□ Yes □ No
Comments (Optional)

2.3 Annotation Results

We first performed an experiment with two annotators on 200 sources to assess the inter-annotator agreement. The sources were chosen by querying the channels' APIs and retrieving those falling under the "hotel" or "b&b" category.
The annotation procedure ran as follows: first the two annotators independently assessed each venue in the dataset through the interface; then they discussed and resolved disagreement where possible, leading to a refinement of the guidelines accordingly. Finally, agreement was calculated both as percent agreement and with Cohen's Kappa measure (for the binary assessment) and with Krippensdorf Alpha (for the 5 point scale ratings).

The results on inter-annotator agreement are briefly detailed here.

The table below shows the percent agreement on the total dataset (discarded venues as not relevant or not showing enough info are included).

	Agr. Recomm	**Agr. Liking**	**Agr. liking groupings**
tot agr	195	157	173
% agr	97,5	78,5	86,5

Percent Agreement

	Agr. Recomm	**Agr. Liking**
Krippendorff's ∝	0.897	0.556
Cohen's K	0.890	---

∝ and k Agreement

As the agreement results are considered satisfactory (at least for the binary judgement[2]), the rest of the the annotation was carried out by a single annotator on another 100 Facebook sources.

The overall composition of the annotated set is given in tables below.

Facebook Location	**Nr. of sources**
Tuscany	30
Spain	51
Amsterdam	24
London	5
Paris	56
Dubai	5

[2] ∝ for the 5-point ratings is instead quite low, which is not surprising also given the relatively small annotated sample. However, as the percent agreement shows, we assume that grouping the ratings into 3 sets also increased **∝.**

Rome	1
Berlin	22
Total	*203*

Foursquare Location	**Nr. of sources**
Tuscany	26
Spain	70
Amsterdam	4
Total	*100*

The Facebook dataset is the used for running the experiments for building the SSM module.

3. Data acquisition and training

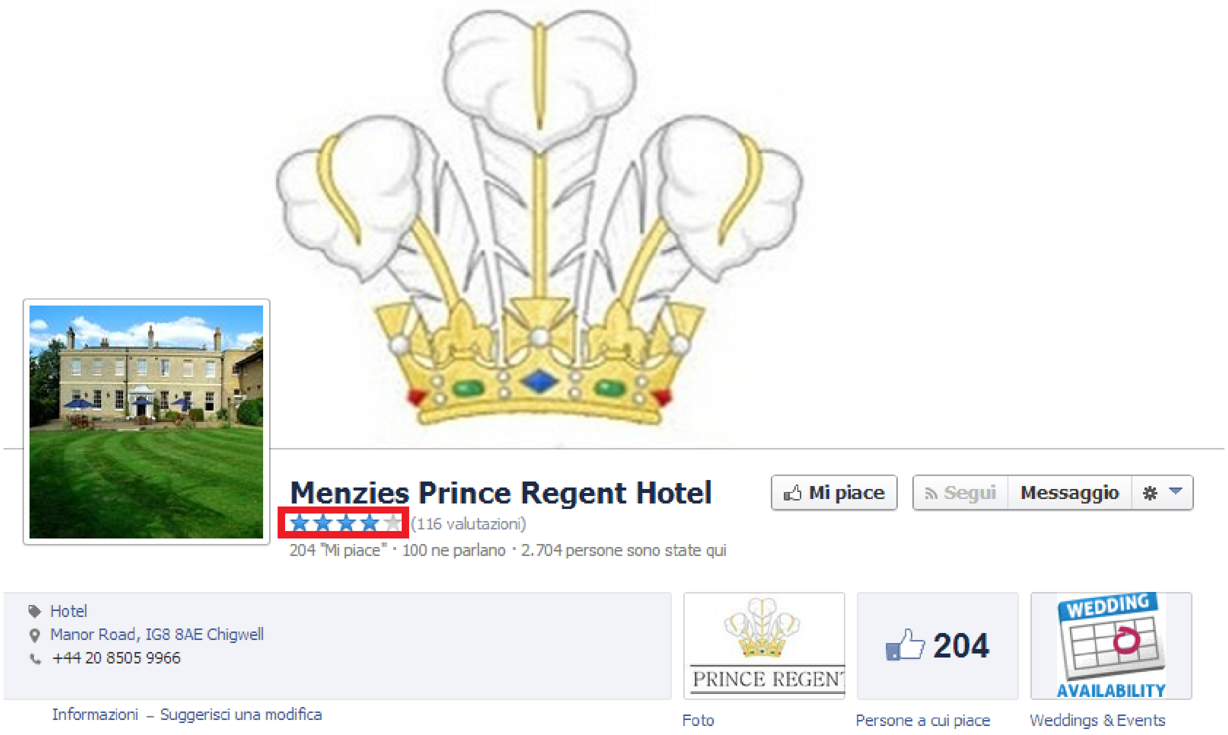
After the annotation phase, the actual metrics for the annotated sources, except for those classified as “not relevant”, have been retrieved by querying the channel’s APIs for each source. The action of taking data using API is different from the act of crawling a single web site. In fact, on a normal web site most of the data is represented by text. On social media, in addition to text, different metadata are present, which represent the interaction between users. The collection of the data was performed at a time as close to the annotation as possible, in order to avoid a mismatch: the downloading of data regarding a source was issued as soon as both the annotators finished annotating that source. Moreover, only the last 2 years were taken into account, as humans are presented with recent activity first when viewing a source, and recent activity is clearly more relevant for humans when assessing a source.

3.1 Facebook / Foursquare APIs

A module that retrieves data from Facebook was implemented. Facebook allows to access its data with two instruments. The first one is the *graph API* and *FQL* (Facebook Query Language). Both of them allow to get data using a particular syntax.
In particular the crawler uses the GraphAPI and in particular the API for page and feed.

For this work, the entities taken into account are the *pages*, with geographical information regarding the *location* of the venue to which the page is related.

The act of crawling public available data from Facebook has different issues. In fact, not all the data are publicly available. In accordance with the developer Roadmap web site of Facebook, the crawling module must use an Application Access Token for retrieving data via the API. Unfortunately, not all the relevant aspects of a source could be represented as quantifiable metrics to be later used as features for the machine learning algorithm, as some **mismatch** exists for each source between what is visualised in the browser and what is obtainable via the API. Most notably for our case, Facebook's API does not provide information on the place rating (stars) using the Application Access Token. This might translate into a limitation for the performance of our classifier, as the human annotator reported that such information was valuable for them.



Menzies Prince Regent Hotel

★★★★☆ (116 valutazioni)

204 "Mi piace" · 100 ne parlano · 2.704 persone sono state qui

Hotel
Manor Road, IG8 8AE Chigwell
+44 20 8505 9966

Informazioni – Suggerisci una modifica

Unfortunately, as of today, this information is not available via Facebook API because it is necessary to use a Page Access Token that is not publicly available.

3.2 Raw measures

Facebook APIs (as described in Section 3.1) can provide the following data

Type	Raw data	Description

Info		
	about	The About section of the page
	description	Description of a page
	fan_count	Number of users that liked the page (likes)
	talking_about_count	Number of people talking about this page
	were_here_count	Number of visits of the page
	checkins_count	Number of checkins at the venue the page refers to
	cover	The cover image
Post		
	post message	Text of a post
	post created time	Time of creation of a post
	post update time	Time of last update of a post
	admin or user	Information if the post is created by admin or an external user
	type	Type of post (photo, status, note)
	post_likes_count	Number of likes of the post
	comments count	Number of comments to the post
	shares_count	Number of shares of the post
Comment		
	comment message	Text of comment
	comment created time	Time when the comment was created
	comment_likes_count	Number of likes of the comment

3.3 Derived metrics /features

From such raw measures, and following discussions with the annotators, some complex metrics have been derived. Such metrics, along with some row ones, represent different aspects of the source, and were judged by human annotators as important in issuing their assessment.

These metrics are the features on which the machine learning algorithm is later trained.

They can be represented as integers (counts) or booleans (presence or absence of a certain feature).

For Facebook we chose to use:

Feature	Type
Is the cover present	Complex - bool
Fan count	Raw - numeric

Talking about count	Raw - numeric
Were here count	Raw - numeric
Checkins count	Raw - numeric
Total Posts	Raw - numeric
Total Likes	Raw - numeric
Total Comments	Raw - numeric
Total Shares	Raw - numeric
Number of posts by page admin	Complex - numeric
Number of photos in the page	Complex - numeric
Number of words in the “about” section	Complex - numeric
Average number of words on posts	Complex - numeric
Average number of words on comments	Complex - numeric
Average number of posts for month	Complex - numeric
Average number of comments for month	Complex - numeric

3.4 Training and cross-fold validation

The classifier training was performed using Weka [1].
Training data is available at this stage for 2 channels (FB and FS) and for 2 types of assessment (liking - recommendation).

Thus, 2 classifiers can be trained:

- FB classifier for liking
- FB classifier for recommendation

Using the Weka cross-validation feature, we were able to experiment and test with the training set in order to identify the best classification algorithm, which was identified to be the `weka.classifiers.trees.RandomForest`.

The data crawled from the annotated sources was then transformed into an arff file, containing the aforementioned features and the classification result from the annotation.

For what concerns the “not enough info” instances, they have been treated as not classifiable, both for liking and recommendation.

The possible classes for recommendation are:

- 0 = not recommended or not classified
- 1 = recommended

For what concerns liking, we experimented with the classes and decided to group some values together (similarly to what was done for the agreement calculation), obtaining 4 classes:

- 0 = not classified
- 1 = bad (former classes 1 and 2)
- 2 = ok (former class 3)
- 3 = good (former classes 4 and 5)

The tables below report on the 10-fold cross validation results obtained in weka.

Recommendation

```
Correctly Classified Instances          99               69.2308 %
Incorrectly Classified Instances        44               30.7692 %
Kappa statistic                          0.3795
Mean absolute error                      0.3594
Root mean squared error                  0.4795
Relative absolute error                 72.4539 %
Root relative squared error             96.2664 %
Total Number of Instances              143

=== Detailed Accuracy By Class ===

              TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                0.662     0.282      0.662     0.662     0.662      0.739    0
                0.718     0.338      0.718     0.718     0.718      0.739    1
Weighted Avg.   0.692     0.313      0.692     0.692     0.692      0.739
```

Liking

```
Correctly Classified Instances          80               55.9441 %
Incorrectly Classified Instances        63               44.0559 %
Kappa statistic                          0.2771
Mean absolute error                      0.2392
Root mean squared error                  0.3887
Relative absolute error                 75.9655 %
Root relative squared error             98.2329 %
Total Number of Instances              143

=== Detailed Accuracy By Class ===

              TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                0         0.014      0         0         0          0.554    0
                0.136     0.124      0.167     0.136     0.15       0.58     1
                0.62      0.458      0.571     0.62      0.595      0.623    2
                0.717     0.134      0.717     0.717     0.717      0.863    3
```

Weighted Avg.	0.559	0.29	0.54	0.559	0.549	0.692

As it can be seen, while the absolute results for the recommendation classification are acceptably good (especially considering that we used a small training set), those for liking are not so good. This is however consistent with the manual annotation agreement results.

4. The Sentiment on Social Media OpeNER module

The final version of the tool thus uses two models generated from the training data (one for liking and one for recommendation) applying the aforementioned `RandomForest` algorithm as implemented in Weka, and runs it on a Weka engine to classify new instances.

The prototype has been built as a module using the OpeNER specifications in D7.11 Definition of the project development environment.
The source files and the installation guide can be found in: https://github.com/opener-project/social-media-classifier

Upon installation on a new machine the Weka models are automatically generated from the enclosed training sets; this way the model is always generated by the same version of Weka that is going to be used to classify, thus avoiding conflicts.

The current version of the module contains training sets for Facebook.

Expert users may change the generation of the model by changing the algorithm; see line 31 of https://github.com/opener-project/social-media-classifier/blob/master/Rakefile

Expert users can also replace the training data, by changing the contents of https://github.com/opener-project/social-media-classifier/blob/master/core/target/training_liking.arff and https://github.com/opener-project/social-media-classifier/blob/master/core/target/training_recommendation.arff and running a new bundle exec rake generate.

This file https://github.com/opener-project/social-media-classifier/blob/master/lib/opener/social_media_classifier.rb contains the command that launches the classifier with the previously built model, in order to classify new entries.

Further details can be found in the OpeNER document D4.22 and in the readme contained in the code distribution.

5. Test interface

To facilitate the visualization of results a simple test interface has been prepared, which aggregates the two classification results (i.e. both for likings and recommendation) and the list of source venues with url to the original page on FB.

The interface can be found here: http://wafi.iit.cnr.it/openervm/weka/results/FB.php

References

[1] Clara Bacciu, Francesca Frontini, Davide Gazzé, Angelica Lo Duca, Valeria Quochi, Irene Russo, Maurizio Tesconi (2013) Definition of indicators and metrics for Sentiment Analysis. OpeNER Deliverable D4.21.

[2] Francesca Frontini, Clara Bacciu, Davide Gazzè, Valeria Quochi, Maurizio Tesconi (2014) Sentiment analyser on Social Media. OpeNER Deliverable D4.22

[3] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.