Collective Intelligence, Overview

Chapter · July 2017

DOI: 10.1007/978-1-4614-7163-9_106-1

Ioannis (Yiannis) Kompatsiaris
The Centre for Research and Technology, Hellas

Symeon Papadopoulos
The Centre for Research and Technology, Hellas

Collective Intelligence: Overview

Yiannis Kompatsiaris, Sotiris Diplaris, and Symeon Papadopoulos
Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

Synonyms

emergent semantics, social media analysis.

Glossary

Graph: a set of nodes and edges connecting the nodes.

Network: a graph that assigns some semantics to the nodes and kind of interaction for the links.

SNA: Social Network Analysis is the study social network characteristics and dynamics.

Data mining: extracting implicit information from a domain.

Community detection: a class of network analysis algorithms that identify groups of nodes that are densely connected.

Multimodal: a kind of analysis involving more than one media or metadata types (e.g. text, image, geolocation).

UGC: User-Generated Content, multimedia content (image, text) that is created/captured by casual users and shared online.

Definition

Recent advances of Web technologies have effectively turned ordinary people into active members of the Web: casual users act as co-developers and their interactions and collaborations with each other have added a new social dimension on Web data. For example, Wikipedia[1] motivates thousands of volunteers around the world to create the world's largest encyclopedia. An image in Flickr[2] is annotated with descriptive tags, is associated with the users that like it and mark it as favorite, and often carries spatial and temporal information denoting its capture context. Even though all these facets of information are not combined naturally with each other, they still carry knowledge about the resource, each of them representing the resource in a different feature space.

The fact that users annotate and comment on content in the form of tags, preferences, ratings, etc. on a daily basis, gives this data source a dynamic nature reflecting events and the evolution of community focus. Although modern Web 2.0 applications encourage annotations and feedback by users, these are not sufficient for extracting such latent knowledge, because they lack clear semantics and it is the combination of multiple features that could enable more thorough understanding of social content. Therefore, there is need for scalable and distributed approaches capable of handling the massive amounts of available data and capturing the Collective Intelligence from User-Generated Content (UGC).

Introduction

This article focuses on approaches that aim to capture and expose the Collective Intelligence lying within large-scale UGC, and implemented by using media and users' connections, actions and interactions in social networking and sharing applications, commonly referred to as Web 2.0. Such approaches face numerous challenges due to the nature of social data, i.e. scalability, multimodality, noisy content, polysemy, lack of uniformity and spam. Therefore, the management, analysis and indexing of such content is becoming an increasingly important research topic.

Key Points

In Section Methodology a state-of-the-art survey is given by first defining the different features of existing applications dealing with Collective Intelligence extraction in terms of

[1] http://www.wikipedia.org
[2] http://www.flickr.com/

input sources, analysis dimensions and application domains. Then the applications are reviewed and categorised. Two use cases are presented that encompass many of the issues discussed here and showcasing the potential of Collective Intelligence for the delivery of enhanced services to end users in tourism and for social media indexing. Finally, we also discuss the identified issues and future challenges in developing large-scale Collective Intelligence extraction applications.

Historical Background

There already exist numerous approaches that mine UGC to provide useful results to various applications. For example, mobile location information and uploaded content is used to share nature experiences[3], discover travel patterns and provide travel advice[4,5], report problems in cities[6] and deal with climatic changes as in the Climate Collaboratorium project of the MIT Center for Collective Intelligence[7]. However, the main characteristic of such applications is their dependence on the collection of well-structured contributions through specific applications, and on simple statistical processing of the contributions. Very few focus on the analysis of implicit relations in UGC and on dealing with unstructured large-scale data.

Several other techniques deal with the tasks of tag refinement and automatic tagging, especially for user-contributed photos. For instance, the method in (Liu, et al., 2010) refines and enriches tags based on the visual and semantic consistency between social sites. Other algorithms perform automated tagging of images, either by building classifiers for individual semantic labels (Li, et al., 2003), or by learning relevance models between images and keywords (Jeon, et al., 2003). A widely used method for automatic tagging is the semi-supervised graph-based label propagation technique (Zhou, et al., 2004) based on the construction of a visual similarity graph. Although such approaches make use of online content collections and accompanying metadata, they are aiming to improve the quality of individual content items and therefore they do not directly address the collective nature of social media and the objective of extracting the emergent semantics of social content.

[3] http://www.ispot.org.uk/

[4] http://www.mobnotes.com/

[5] http://www.dopplr.com/

[6] http://www.fixmystreet.com/

[7] http://cci.mit.edu

Methodology

Social media dimensions

Understanding large-scale social media content involves the consideration of a multitude of features in the development of a suitable analysis framework.

A key issue to be addressed is the presence of *noisy and ambiguous data*. UGC is very noisy containing non-relevant contributions either intentionally (spam) or due to error. The lack of constraints on tagging is responsible for numerous annotation quality problems, such as spam, misspellings, and ambiguity of semantics. Moreover, the lack of semantic structure in the contributed information results in tag ambiguity, synonymy and granularity variation.

UGC can be viewed as a *rich multi-modal source of information* including attributes such as time, favorites and social connections. Social media analysis systems exploit different modalities of input, ranging from single visual, textual, or user information to fused sets of media sources, e.g. annotated images or geotagged images. In such approaches, the need to obtain a joint, unique representation of multimedia objects demands techniques that can handle the characteristics exhibited by different types of data. This is true both in terms of the nature of raw features (e.g., sparse, highdimensional word co-occurrence vectors extracted from text descriptions, compared to the usually dense and low-dimensional descriptors extracted from visual content), as well as in terms of their semantic level (e.g., while abstract concepts like “freedom” are more easily described with text, concrete concepts like “sun” are more easily grounded on visual information).

A basic feature characterising the majority of Web 2.0 applications is their support for communication through social network structures. Information about the various types of social relations may be represented in *social graphs* depicting communication and friendship networks or organization charts. Social graphs do not only connect users but also their contributed content. Exploiting information about social relations between members of a community is also a task for emerging semantics discovery methods dealing with the *social network analysis* dimension.

Scalability is an important issue since the discovery of implicit information is based on massive amounts of data. Such huge volumes of UGC raise scalability issues that compromise the performance (in terms of accuracy) of algorithms operating on such data. The situation gets worse in cases where either the employed algorithms aim at extracting knowledge patterns that only become stable after a specific usage period or processing needs to address *(near) real-time data*. These cases pose challenging requirements in terms of algorithmic design, computational power, memory, storage, and computation distribution, which are considerations in the design of real-time Collective Intelligence discovery systems.

Another important aspect social media dimensions stems from the *spatial processing* of incoming data. Following established web and multimedia handling behavior, many online users upload, tag and localise their contributions. Thus, many recommendation, presentation or prediction techniques may exploit this information for enhancing location-based systems and services.

A different dimension, orthogonal to the location aspect is time. The possibility of processing *temporal* features endows such systems with event processing capabilities. When time aspects of UGC are considered, it is possible to derive information about situations or *events*. Based on massive user contributions, the Collective Intelligence extraction results could range from the single representation of events, to past/current events detection or even to future events prediction. More elaborate applications are able to capture knowledge in the combined *spatio-temporal* dimension, for instance to discover knowledge about routes or map areas with particular interest.

To summarise, analysis techniques dealing with Web 2.0 and social network content and structure can be categorised with respect to the different dimensions they embrace (multi-modality), their ability to process the underlying social network structure information, the different features they consider (visual, audio, textual or their fusion), the capability of scaling, as well as their potential for spatio-temporal events and running in real-time.

Exploiting the information carried by social media can then help tackle a variety of issues in different domains, e.g. tourism, culture, e-government and scientific disciplines including information retrieval, machine learning, data mining, and multimedia understanding, where for example, social media sites can be used as a rich source of weakly labeled data for solving large-scale artificial intelligence problems. Figure 1 depicts the various dimensions described along with possible early and late fusion combinations of the available features.

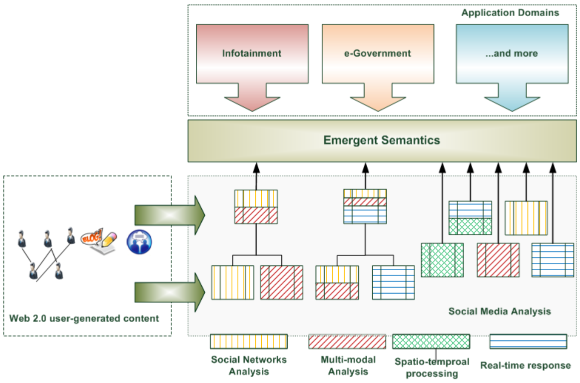

Figure 1: Dimensions and Applications of mining UGC.

Approaches for Collective Intelligence extraction

In this section we review some social media analysis applications that take into account large-scale semi-structured UGC and apply computational methods to discover "hidden" knowledge. We also attempt to categorise them according to the above discussion and to the features these techniques consider.

Multi-modal analysis approaches

City map extraction is a key task in different multi-modal analysis methods. In (Kennedy, et al, 2007), geo-location and tag information are used in order to generate representative city maps. In (Quack, et al. 2008) and (Papadopoulos, et al. 2011a) tags and visual information together with geo-location are used for object (e.g. monuments) and event extraction. Quack, et al. (2008) base their method on pairwise similarity calculations considering visual features and multi-view geometry, and employ hierarchical clustering for grouping images. Instead, Papadopoulos, et al. (2011a) devise hybrid graph-based image clustering and cluster classification.

Other works in the line of multi-modality-aided localisation include tourist travel pattern discovery (Girardin, et al., 2008) and landmark retrieval and localization (Kalantidis, et al., 2011). Tags from Flickr images and timestamp information are used in (Girardin, et al., 2008) to form a time-ordered set of geo-referenced photos to distinguish locals from tourists. The study included recording of photographing habits of around 1 million users thus depicting how localizable-data collection efforts can scale. VIRaL (Kalantidis, et al., 2011) is a web-based image search tool that identifies and localises similar images under different viewpoints.

Classification

A lot of research has been conducted towards classification with networked data from social media. The main application domain for such techniques is online advertising. Based on already known user interests extracted from their profiles or ad clicks, it is possible to infer preferences of the user's contacts in a social network in the form of known categories. This is mainly a multi-label classification problem, since a user may be classified in more than one categories, according to his different interests.

The Behavioral Targeting problem is tackled by different social media classification approaches (Tang and Liu, 2011; Macskassy & Provost, 2007). Tang and Liu (2011) use spectral clustering techniques to extract social dimensions, which then undergo classification with SVMs. The raw heterogeneous social connections become feature vectors that are used for networked classification. Another approach (Macskassy & Provost, 2007) employs weighted voting with classification based on collective inference in order to derive user preferences from social data. The simple approach achieves comparable performance to more elaborate methods, such as Gaussian fields.

Other techniques involve social relations analysis in defining proper classification models for online advertisement (Kendal and Zhou, 2009; Wen and Lin, 2010). Kendall and Zhou (2009) describe how they are able to infer missing profile information (age, gender, etc.) based on the user friends' preferences to classify them in suitable categories for advertising. Another approach for inferencing user preferences from social connections is given by Wen and Lin (2010) who use regression models and network correlation analysis for predicting and improving the quality of classification. The method is applied on both implicit user interests indicated by the content of communications or Web 2.0 activities, and on explicit ones derived from user profiles.

Community, trend and event detection

There are also several applications exploiting the knowledge extracted from community detection in social networks. For instance, it is common to derive community-based views of networks, i.e. networks of which the nodes correspond to communities of the original networks and the edges to the relations between communities. Such views are more succinct and informative than the original networks. It is for this reason that community detection has been applied in recommendation systems (Specia & Motta, 2007; Schifanella, et al., 2010), as well as for representing user profiles (Gemmel, et al., 2008). Other applications making use of the knowledge extracted from tag communities include sense disambiguation (Yeung, et. al., 2009) and ontology engineering (Specia & Motta, 2007).

The patterns emerging often show deep interconnections with various real-world events (Sakaki, et al., 2010; Signorini, 2011) in a way that the evolving world reality is captured at each instant.

Recently research on emerging semantics analysis has progressed enough to achieve event prediction (Jin, et al., 2010) and automatic travel itinerary extraction (De Coudhury, et al., 2010). A leap in the exploitation of UGC research is the work by Jin, et al. (2010) which explores global trends and sentiments that can be drawn by analyzing the sharing patterns of social multimedia. Taking into account both spatial and temporal aspects of content item views and uploads in social media sites and aggregating them, the authors are trying to forecast future events impacting politics, economics and marketing. De Coudhury, et al. (2010) automatically derive travel itineraries for popular touristic cities from large-scale user contributed rich media repositories. To achieve this effect, timed paths are generated per user, by using EXIF temporal information from photos, which are then aggregated into a graph used for the approximate extraction of itineraries.

Furthermore, other work shows that actions of individual Web users can indicate macro trends, when properly pooled. There are studies using Search Engine queries for influenza surveillance over the Web, such as in (Gisnberg, et al., 2009), where Google search engine queries and data from the Centers for Disease Control (CDC) are used to monitor influenza rates 1-2 weeks ahead of the CDC reports.

Real-time applications

Finally, a separate class of applications involves the real-time aspect in the analysis. An early, but not quite scalable, tool (Hinze & Voisard, 2003) deals with the analysis of user profiles and query logs for extracting personalised tourist information (places and events) using a hierarchical semantic geospatial model as well as an event notification system. SCENT (Lin et al., 2010) is a framework for monitoring the evolution of multi-faceted social network data resulting from users' continuous online interactions. It enables large-scale data management and analysis by modeling social data as tensor (multi-dimensional array) streams, tracking changes in the spectral properties of the tensor over time. It has mainly been used in recommendation and monitoring applications.

Another tool (Yang, et al., 2008) was developed to improve search quality by reranking and to recommend supplementary information at query time (i.e., search-related tags and canonical images) analyzing visual content, high-level concept scores, time, and location metadata. In Wu et al. (2009) a real-time approach combining the contextual information from time duration, number of views, and thumbnail images with content analysis derived from color and local points is used in order to achieve near-duplicate elimination in video search. Finally, Henrich and Lüdecke, (2008) exploit georeferenced toponyms found in community websites to model vernacular places in cities. The system is capable of answering queries in real time.

Table 1 presents a categorisation of the above techniques and applications with respect to the kind of analysis they employ (i.e. recommendation, clustering, localisation, monitoring, event detection, etc.) and the parameters discussed in Section Social Media Dimensions.

Most applications do not employ the spatial or spatio-temporal dimension; instead they mostly make use of textual annotations to provide recommendations, or represent events and situations. However, more advanced systems (Kennedy, et al., 2007; Quack, et al., 2008, Girardin, et al., 2008; Sakaki, et al., 2010; Signorini, 2011; Jin, et al., 2010) have also been presented, capable of processing various types of modalities enabling the spatio-temporal and situational dimension. Scalability is addressed by a wide range of applications; however the amount of works enabling the real-time aspect is still very limited. It is notable that very few real-time applications (Yang, et al., 2010; Wu, et al., 2009) are also able to handle and analyze multi-modal information.

Illustrative examples

Emerging semantics in tourism: ClustTour

ClustTour[8] brings to surface Collective Intelligence by using image clustering to identify landmarks and events in large city-focused photo collections. In that way, the captured semantics contribute to a more efficient and pleasant online city exploration experience.

As a starting point, the framework uses the BIRCH clustering algorithm (Zhang, et al., 1996). to derive a set of clusters of geo-coordinates by clustering the set of input geotagged photos. By computing the convex hull of each such cluster with the use of the Graham scan algorithm, the framework produces a set of polygons corresponding to city areas. Therefore, the spatial dimension is encompassed; depending on the photo concentration across the city, areas can range from districts, to neighborhoods or even to famous landmarks.

Subsequently, the framework employs a graph-based scheme that groups the photos contained in the convex hull of each area into clusters of similar photos based on their visual content and text metadata. A hybrid similarity graph is constructed, containing the photos of the area under study as nodes, and their pairwise similarities as edges. Pairwise hybrid similarities are computed by considering two independent similarity components, thus incorporating the multi-modal dimension: (a) visual similarity, computed as the inverse of Euclidean distance for 500-bin histograms of SIFT visual words, (b) textual similarity, computed as the Jaccard coefficient between the terms associated with photos, after preprocessing their title and tags. Next, a community detection algorithm is applied on the similarity graph to extract sets of nodes that are densely connected to each other. For each image cluster, a title is automatically extracted by processing the individual image

[8] http://www.clusttour.gr/

titles to identify the most frequent consecutive sequence of terms.

Once image clusters are extracted, the ClustTour framework classifies these clusters into landmarks and events (Papadopoulos, et al., 2011a). For each cluster, four features are extracted: (a) the duration of the cluster, i.e. the interval between the first and last of its photos (as ranked by their creation timestamp), (b) the cluster popularity, defined as the ratio of number of distinct photo owners to number of cluster photos, (c) the frequency of "landmark" terms, and (d) the frequency of "event" terms in the titles and tags of the cluster photos. The first two features are described in Quack, et al (2008). In our experiments, we found that they were not sufficiently discriminative for the classification problem at hand. For that reason, one term model was constructed for each class (landmarks, events). The term model is trained by annotating a small percentage (~10-20%) of the clusters of each city with the appropriate class. This constitutes a lightweight annotation process, since the total number of clusters for each city ranges from a few hundred to a couple of thousand at most, and in addition, the distinction between a landmark and an event is straightforward for human annotators. Once the two term models are constructed and the features are extracted for each cluster, an SVM classifier is trained to enable the classification of clusters into landmarks and events. In this stage both social network analysis and temporal dimensions are exploited.

In its final processing step along the spatio-temporal dimension, ClustTour processes the time intensity of image uploads in order to extract important time intervals. Time intensity series are derived for a set of images by considering sliding temporal windows over the active period for that set (defined as the interval between the first and the last image of the set), and counting, in each window position, the number of photos uploaded within the interval. Then, a simple burst detection scheme is applied that selects predefined intervals (days, day sequences, months, seasons) with exceptionally high intensity rates (two standard deviations above average). Intensity time series are constructed separately for each city and for each area in the city, and the corresponding burst intervals are extracted.

Ultimately, ClustTour leverages the captured semantics of places through a map-based interface that enables users to navigate through them and explore the corresponding image collections. The web application offers two modes of exploration: a city view depicting a high-level view of the most important areas in the city (Fig. 2) and an area view centered on the selected area and showing landmarks and events in its vicinity. ClustTour also exploits the temporal dimension in terms of presentation by providing a temporal content organisation layer on top of the detected areas and photo clusters (Papadopoulos, et al., 2011b). In that way, users obtain informative views over the interesting spots and areas of a city.

With its ability to analyse massive visual and textual information from UGC and to present both landmarks and events in a spatio-temporal dimension, ClustTour is a typical case of implicit knowledge extraction application enabling the analysis of multi-modal input at large scale.

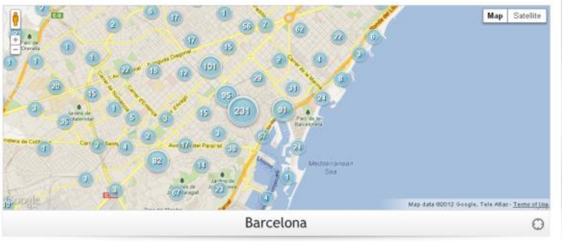

Figure 2: Areas detected by spatial clustering of Flickr photos in Barcelona.

Multi-modal indexing in social media

The diverse information carried by social media can be efficiently modeled with specialized multi-modal analysis techniques. For instance, the basic characteristic of tagged images is the coexistence of two heterogeneous information modalities, visual and tag, which refer to the same abstract meaning. This multi-modal nature makes efficient indexing of tagged images a challenging task that apart from dealing with the heterogeneity of modalities also has the opportunity to exploit their complementary information. Semantic image indexing has been recognized as a particularly valuable task for various applications of content consumption. Current literature has made considerable progress in this direction especially for uni-modal scenarios. However, multi-modal analysis has the potential to further improve this process. This is based on the fact that different modalities, when combined, encompass a higher amount of information. Web 2.0 and social networks have further motivated this idea by making available plentiful user tagged images.

Combining the low-level features at an early stage makes clear the need to obtain a joint representation of tagged images. The most trivial approach in this direction is to define a joint feature space by concatenating the individual uni-modal features extracted from both modalities, also known as early fusion. However, by indiscriminately placing features extracted from different modalities into a common feature vector, the resulting space is likely to be dominated by one of the combined modalities or lose its semantic consistency. To this end, in order to tackle the obstacle of data heterogeneity Nikolopoulos et al. (2011) developed a methodology focusing at optimizing the combination of low-level features. More specifically, the authors examine the use of probabilistic Latent Semantic Analysis (pLSA) based aspect models, as the means to define a latent semantic space where heterogeneous types of information can be effectively combined, thus enabling the semantic representation dimension. The method emphasizes the exploitation of relations across modalities, in cases where two or more of the modalities constitute different expressions of the same abstract meaning. The multi-modality dimension is exploited in order to build a second level pLSA model and to extract a hierarchy of topics. In this way the method manages to exploit the cross-modal relations between visual and tag features and to improve retrieval performance. Extending this work, the authors have examined the use of high-order pLSA to improve the semantic capacity of the derived latent topics. High-order pLSA is essentially the application of pLSA to more than two observable variables. The use of this method allowed the authors to integrate the cross-modal relations into the update rules of high-order pLSA, aiming to devise a feature space where the coexistence of two content items that are known from experience to co-occur rather frequently is more important in defining the latent topics, than the co-existence of two content items that rarely co-occur. The authors also proposed a distributed version of the algorithm, addressing the scalability dimension and enabling processing data at large scale. Table 2 shows the Mean Average Precision (MAP) scores achieved by seven feature spaces, evaluated on the NUS-WIDE dataset[9].

We notice that visual, being a more dense representation of image content, outperforms tags, which are typically very sparse. As expected, the straightforward combination of both modalities by their simple concatenation, fails to combine them efficiently and performs slightly better than the best of the uni-modal cases. When

[9] http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm

moving to the space of pLSA-based latent topics, we can see an increase of the retrieval performance for both uni-modal cases, which verifies the efficiency of aspect models to discover semantic relations between the images. Moreover, it is interesting to note that the relative improvement achieved by plsa-tags is considerably higher than the relative improvement of plsa-visual. This can be attributed to the ability of pLSA to more efficiently handle sparse data, since the image representation based on tags is much sparser than the corresponding representation based on visual. Additionally, the performance achieved by plsa-tags-plsa-visual introduces some improvement over the uni-modal cases, in contrast to the behavior of visual-tags. This verifies the ability of the latent space to more efficiently combine the heterogeneous modalities of tagged images, compared to the original space. Finally, the performance achieved by high-order-plsa verifies the usefulness of cross-modal relations in creating a semantics-sensitive feature space, since an improvement of approximately 1.8% MAP units is introduced over the best performing baseline. Although evaluated in an image retrieval setting, the method can be easily adjusted to address other tasks, such as clustering and recommendation.

New Developments

This section covers developments in the Collective Intelligence field since the first edition of the encyclopedia and presents how the literature has advanced in the meantime. New approaches have emerged in all aspects of Collective Intelligence extraction from social media, with several new approaches for multi-modal analysis, novel classification schemes for recommendations, advanced topic and social event detection methods, whilst real-time applications that leverage on social media content have become more prominent.

Multi-modal indexing

In the realm of multi-modal indexing in social media, a novel approach for clustering multi-modal data (Petkos et. al, 2016) was introduced. The approach differs to previous methods in that it attempts to take into account an existing example clustering. In particular, it utilizes the example clustering in order to learn a model that predicts whether a pair of items belongs to the same cluster or not. This model is effectively a learned similarity measure and uses as input the set of per modality distances between the pair of items. Subsequently, the approach computes the predictions of this model for pairs of items in the collection to be clustered and the items are organized in a graph. This graph has a node for each item of the collection and an edge between two nodes indicates that the prediction of the model for the corresponding pair of items is positive. Finally, a community detection algorithm is applied on the resulting graph. The approach is applied on two problems that involve the clustering of multi-modal data retrieved from social media, namely the discovery of social events in collections of images, and the the detection of different views of landmarks, again based on images retrieved from social media.

Another approach to tackle the problem of multi-modal clustering (Cai et. al, 2011) combines modalities at an intermediate level between early and late fusion. In this method, similarities according to the different modalities are combined by summing the inverses of regularized individual Laplacian matrices in an aggregate Laplacian matrix. This aggregate Laplacian is subsequently used to perform clustering exactly like in the common spectral clustering scenario. In a final step, the algorithm computes the final assignment of items to clusters, so that a cost function that measures the disagreement between the final clustering and the individual clusterings for each modality is minimized.

It is also worth noting the connection between multi-modal clustering and ensemble clustering, in which the goal is to combine a set of clusterings (produced e.g. from a set of different algorithms) in some optimal manner. A multi-modal clustering problem can be cast into an ensemble clustering problem by performing a set of clusterings - e.g. one per modality - and then combining the individual clusterings. This is also one of the proposed approaches for social event detection in Becker et. al, (2010).

Classification

Sentiment analysis has become of particular importance in microblogging platforms, such as Twitter: brands can learn what people think of their products, politicians can learn the users' opinions on them, people can aggregate opinions on topics of their interest and so on. Given the huge amounts of broadcasted content, automating the process of opinion mining becomes a rather crucial task for creating real-time insights. The two mostly studied sentiment analysis tasks are subjectivity (given a set of documents, find the subjective ones) and polarity detection (discrimination of positive/negative documents). The most common approach to deal with these tasks is to train a classifier on a labeled corpus and apply the learned model on the desired test set. A novel approach (Tsakalidis et. al., 2014) tackles the problem of the domain-dependent nature of the polarity detection task in Twitter. They train different classifiers on various sets of features and combine them in an ensemble model with a resulting accuracy highly competitive against traditional in-domain training methods.

A key question for social media platforms is how users navigate inside and between various photostreams. A novel method (Chiarandini et. al., 2013) analyzes a large sample of navigation logs to gain insights into how users navigate between different photostreams. They examine user navigation logs containing several millions pageviews in order to create a photo-stream transition graph to analyze frequent topic transitions and consequently implement two photostream recommender systems: a collaborative filtering approach based on transitions between photostreams and a content-based method using tag-similarity of photostreams.

Another fascinating development has been the idea is to analyse in large scale the sentiment of geotagged photos (based on their titles and tags) and to derive paths in a city that are more pleasant compared to the shortest paths. Quercia et. al., (2013) consider votes from more than 3.3K individuals and translate them into quantitative measures of location perceptions. They arrange those locations into a graph upon which they learn pleasant routes. Based on a quantitative validation, the recommended routes add just a few extra walking minutes, compared to the shortest routes, and are indeed perceived to be more beautiful, quiet, and happy.

Community and event detection

Latest developments in the field of community detection in social media try to identify the different types of groups that emerge spontaneously in online social media and how they differ. To this end, Martin-Borregon et. al., (2014) provide a categorization of online groups along three axes: spatial, temporal and socio-topical. For each dimension they propose a set of general metrics that capture quantitatively the different facets of groups. Specifically, they describe groups with respect of the geographic scattering of their members, the temporal footprint of the members' activity in terms of dispersion, skeweness, and burstiness, and the tendency of the group to aggregate members on a topical or social basis. With respect to the last dimension, they rely on a longstanding theory about the creation of social communities. The theory states that people join groups driven by either pre-existing social ties with other members or by the interest in the topical focus of the group as a whole, so they build metrics to quantify this tendency. They show that their metric well reflects the cardinal points of the theory, being good predictors of the group type. Their metrics are tested on a large-scale corpus of public, online, user-generated groups.

In event detection, the event-centric annotation problem has recently attracted interest facing even harder annotation challenges. Most annotation tools are based on a single type of information (e.g. temporal) to group content, which is restrictive given that there are additional dimensions associating content with an event (location, participants, etc.). Zigkolis et. al., (2012) provide support to the process of event-centric annotation by proposing the CrEve framework, which offers users a set of query facilities covering different aspects of data (i.e. location, time, textual, ownership, etc.) for exploring the collection and associating photo content from social media with real-life events.

Real-time applications

Applications that leverage on UGC streams from social media have recently started to emerge. The task of detecting topics in (near) real time from multiple social networks has found application in real-world scenarios, in which an expert of some domain has to monitor specific topics or events being discussed in social media. For instance, this is the case for computational journalism, in which the media inquirer is supposed to have enough knowledge of the domain of interest to provide initial keywords to perform an initial filtering of the data stream. To monitor and detect all these aspects in real time, we need to extract relevant information from the continuous stream of data originating from such online sources. Determining which are the topics being discussed by the crowd is the first step towards a high-level, human-understandable description of the social data stream. The task of Topic Detection and Tracking in a social media context is demanding since there are many additional factors to consider such as the fragmentation and noise of the user generated content, the real-time requirement, the burstiness of events and their time resolution. Aiello et. al., (2013) explore how much these factors affect the topic detection results by exploring two orthogonal dimensions: a) the effect of the nature of input data, including the pre-processing phase, on the topic detection outcome; and b) the behaviour of different topic detection algorithms themselves. Their methods cover three different classes: probabilistic models (Latent DirichletAllocation), classical Topic Detection and Tracking (a common document-pivot approach) and feature-pivot methods. Along this series of methods, they develop four novel approaches, including methods that use the concept of frequent item set mining.

Taking a step further, Ahmed et. al., (2013) try to predict the time evolution of popularity in UGC without making too many assumptions regarding the underlying phenomena. They define an application agnostic feature space to capture the patterns of behaviour that contents display overtime. They first categorise the behaviour of content overtime to reveal distinct patterns of popularity growth (over time) that generalise the variety of different behaviours displayed by large numbers of content. Second, based on this result they predict the future popularity of content by tracing the types of popularity growth behaviour that contents display overtime.

However, the problem of predicting content popularity form social media in a cold-start prediction scenario is proved difficult. Arapakis et. al., (2014) measured the popularity of articles in terms of their tweet counts as well as page views. Using the content of news articles and external sources, they engineered a large number of features that may indicate the future popularity of news articles. This work revealed that predicting the news popularity at cold start is not a solved problem yet.

Key Applications

The implicit knowledge derived through social media analysis methods can be consumed in many different application domains.

Travel-related web and tourist location-aware services are an obvious market for leveraging multimedia search and context-centric delivery capabilities. ClustTour is such application. Also, in Emergency Response cases, mining UGC coming from citizens and enabling real-time discovery of events is crucial for effective planning and action. An Emergency Response platform featuring emerging semantics capabilities was built within the WeKnowIt FP7 project.

The infotainment sector, and more specifically the cultural event organisation, constitutes a crucial market that can benefit from social media analysis capabilities, e.g. by improving the way that events are experienced from the audience, and offering new kinds of services to event attendants. Infotainment applications that currently enable a minimal degree of semantics analysis include mobile apps for sports (e.g. Wimbledon tennis tournament app[10]) and entertainment (Cannes Film Festival app[11]).

Brand or product monitoring is a further market where the capabilities of emerging semantics are of great value. Such products can help marketing professionals as well as individual customers. The technological innovations stemming from social media analysis could help create new business opportunities in the areas of event organisation, event-related and location-based marketing. The work by Jin, et al. (2010) is a step towards this direction.

A domain that traditionally uses analysis techniques for inferring consumer profiles for marketing purposes is the online advertising sector. Through the so-called Behavioral Targeting (BT) methods it is able to communicate proper messages to target groups, taking into account the online behavior of users. Due to the rise of social media sites, this domain becomes significantly appropriate for the application of emerging semantics techniques (Tang and Liu, 2011).

Another application domain for emerging semantics extraction is e-government. Strengthening citizens' participation in governmental and local communities' decision making processes is not only a cornerstone for the notion of democracy, but also the only way for people to influence the decisions and policies affecting their daily lives. Web 2.0 tools enable citizens to voice their opinions and take part in decision making. Emergent semantics extraction methods can enable citizens to track legal procedures, understand technical documents, express views, and elected representatives to better handle the gathered information and transfer it to other bodies as well.

[10] http://www.wimbledon.com/en_GB/news/mobile.html
[11] http://www.festival-cannes.com/en/apps.html

Table 1: Categorisation of Collective Intelligence extraction applications.

		Dimensions					
		Scalability	**Spatio-temporal**	**SNA**	**Multi-modal**	**Semantic represent-ation**	**Real-time**
Analysis Methods	**Recommendation**	[1-8]	[1], [9]	[1], [3-7], [9-12]	[1-4], [8]	[8], [12]	[2-3]
	Clustering	[8], [13]	[13-14]	[13-15]	[8], [13], [16-17]	[8], [14]	
	Localisation	[18-20]	[18], [20]	[18-21]	[18],[21]	[21]	[19]
	Trend/event detection	[13], [23]	[13], [23-25]	[10], [13], [23-25]	[13], [16],[21], [24-25]	[21], [24-25]	

Index			
[1]	Lin et al. (2010)	[14]	Girardin et al. (2008)
[2]	Yang et al. (2010)	[15]	Au Yeung et al. (2009)
[3]	Wu et al. (2009)	[16]	Quack et al. (2008)
[4]	Tang & Liu (2011)	[17]	Papadopoulos et al. (2011a)
[5]	Macskassy & Provost (2007)	[18]	Kalantidis et al. (2011)
[6]	Kendall & Zhou (2009)	[19]	Henrich & Lüdecke (2008)
[7]	Wen & Lin (2010)	[20]	De Choudhury et al. (2010)
[8]	Nikolopoulos et al. (2011)	[21]	Kennedy et al. (2007)
[9]	Hinze & Voisard (2003)	[22]	Ginsberg et al. (2009)
[10]	Specia & Motta (2007)	[23]	Sakaki et al. (2010)
[11]	Schifanella et al. (2010)	[24]	Signorini (2011)
[12]	Gemmell et al. (2008);	[25]	Jin et al. (2010)
[13]	Papadopoulos et al. (2011b)		

Table 2: Mean Average Precision (MAP) scores achieved by 7 different feature spaces in an image retrieval scenario

Feature Space	#dims	MAP (%)
tags	1000	29.45
visual	500	31.07
visual-tags	1500	31.08
plsa-tags	30	35.674
plsa-visual	30	31.728
plsa-tags-plsa-visual	60	35.906
high-order-plsa	30	37.76

Future Directions

The presented Collective Intelligence extraction techniques attempt to harness one or more forms of online user contributions in order to benefit end-users and organisations by employing large-scale recommendation, prediction, detection, presentation and summarisation techniques.

Such approaches can be applied to domains such as tourism, culture, social sciences, politics, economics, and marketing. In tourism and culture, uploaded media can reveal "off-the-beaten-path" points of interest and events, otherwise difficult to discover through usual Web 2.0 sources. In economics, marketing and brand monitoring, the number of related media uploaded online can reflect the number and locations of products sold in the market.

However, such advanced efforts are still limited despite the widespread usage of social media. Several aspects are currently covered by these methodologies, addressing issues such as scalability, efficiency, fusion and integration of multi-modal sources, as well as real-time analysis. Additional issues and dimensions relevant to social media analysis and data mining include:

Aggregation of multiple input sources. Analysis of multiple Web 2.0 sources by semantics extraction techniques is a step beyond the existing approaches that mostly consider single sources (e.g. Flickr, Twitter). The establishment of connections between content items, which refer to the same topic, among different applications remains a challenge, e.g. the connection of a photo, a tweet and a blog related to the same topic. The need for such methods is particularly motivated by the constantly increasing level of information flow between different social networks.

Linking with Open Data. Exposing, sharing, and connecting pieces of information extracted from social media, using URIs and RDF, with existing Open Data is a crucial step towards comprehensive semantic description and representation of the extracted information.

User Interfaces and Visualisation must be developed in order to support user interaction and understanding of the results of social media analysis.

Trust, security, privacy is a concern for users who contribute content, especially when it is further analyzed by the application. Users should be guaranteed that their contribution remains anonymous and/or is used for an objective they consent to.

Social data sampling, modeling and representation. In order to deal with the enormity of real-world social media datasets, appropriate sampling and data reduction techniques are needed, that are capable of preserving the structure, topology and social context of the original data. There are still many open questions when dealing with social media-related sampling strategies.

Acknowledgments

The work presented in this article was supported by the European Commission under contracts FP7-215453 WeKnowIt and FP7-287975 SocialSensor.

References

1. Ahmed, M., Spagna, S., Huici, F., Niccolini, F. (2013). A peek into the future: Predicting the evolution of popularity in user generated content. In: Proc. 6th ACM Int'l Conf. Web Search and Data Mining, pp. 607–616 (2013)
2. Aiello, L.M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., Jaimes, A. (2013). Sensing trending topics in Twitter. In Transactions on Multimedia 15(6). pp. 1268-1282, October 2013, DOI: 10.1109/TMM.2013.2265080

3. Arapakis, I., Barla Cambazoglu, B., Lalmas, M. (2014). On the Feasibility of Predicting News Popularity at Cold Start. In Prpc. 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014, pp. 290-299.
4. Au Yeung, C.M., Gibbins, N., & Shadbolt, N. (2009). Contextualising tags in collaborative tagging systems. In: HT '09: Proceedings of 20th ACM conference on Hypertext and hypermedia, 251–260.
5. Becker H, Naaman M, Gravano L (2010). Learning similarity metrics for event identification in social media. In: Proceedings of the third ACM international conference on Web search and data mining, WSDM '10. ACM, New York, pp 291–300.
6. Cai X, Nie F, Huang H, Kamangar F (2011) Heterogeneous image feature integration via multi-modal spectral clustering. In: 2011 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1977–1984. doi:10.1109/CVPR.2011.5995740
7. Chiarandini, L., Grabowicz, P.A., Trevisiol, M., Jaimes, A. (2013). Leveraging Browsing Patterns for Topic Discovery and Photostream Recommendation. In ICWSM'13: 7th International AAAI Conference on Weblogs and Social Media, Boston, US.
8. De Choudhury, M., Feldman, M., Amer, S., Golbandi, N., Lempel, R., & Yu, C. (2011). Automatic Construction of Travel Itineraries using Social Breadcrumbs. Proceedings of 21st ACM conference on Hypertext and hypermedia, 35-44.
9. Gemmell, J., Shepitsen, A., Mobasher, B., & Burke, R. (2008). Personalizing navigation in folksonomies using hierarchical tag clustering. In: DaWaK '08: Proceedings of 10th Intern. Conference on Data Warehousing and Knowledge Discovery, 196–205.
10. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457, 1012-1014.
11. Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. IEEE Pervasive Computing, 7(4), 36-43.
12. Henrich, A., & Lüdecke, V. (2008). Determining geographic representations for arbitrary concepts at query time. In Proceedings of first Intern. workshop on Location and the web, 17-24.
13. Hinze, A., & Voisard, A. (2003). Location and time-based information delivery in tourism, Advances in spatial and temporal databases. Lect Notes Comput Sci, 2750, 489–507.
14. Jeon, J., Lavrenko, V., & Manmatha, R. (2003). Automatic image annotation and retrieval using crossmedia relevance models. Proceedings of 26th annual international ACM SIGIR conference on Research and development in information retrieval, 119-126.
15. Jin, X., Gallagher, A., Cao, L., Luo, J., & Han, J. (2010)._The Wisdom of Social Multimedia: Using Flickr For Prediction and Forecast, MM '10 Proceedings of International Conference on Multimedia, 1235-1244.
16. Kalantidis, Y., Tolias, G., Avrithis, Y., Phinikettos, M., Spyrou, E., Mylonas, P., & Kollias, S. (2011) VIRaL: Visual Image Retrieval and Localization. Multimedia Tools Appl. 51(2), 555-592.
17. Kendall, T., & Zhou, D. (2009). Leveraging information in a social network for inferential targeting of advertisements, US Patent App. 12/419,958.
18. Kennedy, L.S., Naaman, M., Ahern, S., Nair, R., & Rattenbury, T. (2007). How Flickr helps us make sense of the

world: context and content in community-contributed media collections. In proc. ACM Multimedia ‘07, 631-640.
19. Li, J., & Wang, J.Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans. Pattern Anal. Mach. Intell., 25, 1075–1088.
20. Lin, Y.R., Candan, K.S., Sundaram, H., & Xie, L. (2011). Scent: Scalable compressed monitoring of evolving multi-relational social networks. ACM Trans. On Multimedia Computing, Communications and Applications. In press.
21. Liu, D., Hua, X.S., Wang, M., & Zhang, H.J. (2010). Retagging social images based on visual and semantic consistency. In: Proceedings of 19th international conference on World wide web,WWW’10, 1149–1150.
22. Macskassy, S.A., & Provost, F. (2007). Classification in networked data: a toolkit and a univariate case study. J Mach Learn Res, 8, 935–983.
23. Martin-Borregon, D., Aiello, L.M., Grabowicz, P., Jaimes, A., Baeza-Yates, R., (2014). Characterization of online groups along space, time, and social dimensions. EPJ Data Science 2014, 2014:8.
24. Nikolopoulos, S., Giannakidou, E., Kompatsiaris, I., Patras, I., & Vakali, A. (2011). Combining multi-modal features for social media analysis. In Hoi, S., Luo, J., Boll, S., Xu, D., Jin, R., & King, I. (Eds.), Social Media Modeling and Computing (in Press), Berlin: Springer.
25. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., & Vakali, A. (2011a). Cluster-based Landmark and Event Detection on Tagged Photo Collections. IEEE Multimedia, 18(1), 52-63.
26. Papadopoulos, S., Zigkolis, C., Kapiris, S., Kompatsiaris, Y., & Vakali, A. (2011b). City exploration by use of spatio-temporal analysis and clustering of user contributed photos. Demo paper in ACM International Conference on Multimedia Retrieval (ICMR), 65:1-65:2.
27. Petkos, G., Schinas, M., Papadopoulos, S., Kompatsiaris, I. (2016). Graph-based multimodal clustering for social multimedia. In Multimedia Tools and Applications. doi:10.1007/s11042-016-3378-2
28. Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In World Wide Web Conference, 851-860.
29. Schifanella, R., Barrat, A., Cattuto, C., Markines, B., & Menczer, F. (2010). Folks in folksonomies: social link prediction from shared metadata. In: WSDM ’10: Proc. 3rd ACM int. conference on Web search and data mining, 271–280.
30. Signorini, A. (2011). Swine Flu monitoring using twitter. In http: //compepi.cs.uiowa.edu/ alessio/twitter − monitor − swine − flu/. Last accessed: 25 October 2011.
31. Specia, L., & Motta, E. (2007). Integrating folksonomies with the semantic web. In: ESWC ’07: Proceedings of 4th European conference on The SemanticWeb, 624–639.
32. Tang, L., & Liu, H. (2011). Leveraging social media networks for classification. Data Min Knowl Disc, 23, 447-478.
33. Tsakalidis, A., Papadopoulos, S., Kompatsiaris, I., (2014). An Ensemble Model for Cross-Domain Polarity Classification on Twitter. In Proceedings of Web Information Systems Engineering – WISE 2014, pp. 168-177, Springer (2014).
34. Quack, T., Leibe, B., & Van Gool, L. (2008). World-scale mining of objects and events from community photo collections.

In Proc. Int. Conf. on Content-based image and video retrieval, 47-56.

35. Quercia, D., Schifanella, R., Aiello L.M. (2014), The shortest path to happiness: Recommending beautiful, quiet and happy routes in the city. In Proceedings of the 25th ACM conference on Hypertext and social media (HT '14). ACM, New York, NY, USA, pp. 116-125. DOI=http://dx.doi.org/10.1145/2631775.2631799.

36. Wen, Z., & Lin, C.Y. (2010). On the quality of inferring interests from social neighbors. In Proceedings of 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10), 373-382.

37. Wu, X., Ngo, C.W., Hauptmann, A.G., & Tan, H.K. (2009). Real-Time Near-Duplicate Elimination for Web Video Search With Content and Context. Multimedia, IEEE Transactions on, 11(2), 196-207.

38. Yang, Y.H., Wu, P.T., Lee, C.W., Lin, K.H., Hsu, W.H., & Chen, H.H. (2008). ContextSeer: context search and recommendation at query time for shared consumer photos. In Proc. 16th ACM int. conf. on Multimedia (MM '08), 199-208.

39. Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. SIGMOD Rec. 25(2), 103-114.

40. Zhou, D., Bousquet, O., Lal, T.N., Weston, & J., Schölkopf, B. (2004). Learning with local and global consistency. In Advances in NIPS, 16, 321-328.

41. Zigkolis, C., Papadopoulos, S., Filippou, G. et al. Multimed Tools Appl (2014) 70: 89. doi:10.1007/s11042-012-1154-5