SocialSensor: Surfacing Real-Time Trends and Insights from Multiple Social Networks

Sotiris Diplaris, Giorgos Petkos, Symeon Papadopoulos, Ioannis Kompatsiaris, Nikos Sarris, Carlos Martin, Ayse Goker, David Corney, Joost Geurts, Yaning Liu, Jean-Charles Point

Dr. Nikos Sarris, ATC
2012 NEM Summit Istanbul, 16-18 Oct 2012
What is SocialSensor?

- 3-year FP7 European Integrated Project
  - http://www.socialsensor.eu
- 1 year into the project (Development of user requirements and first research implementations)
- Consortium: 11 partners
Motivation: Social Networks as Sensors

- Social Networks is a data source with an extremely dynamic nature that reflects events and the evolution of community focus (user’s interests)

- We need to:
  - Transform social media to meaningful topics, entities, points of interest, emotional states, social connections and events
  - Mine the data and their relations and exploit them in the right context
  - Provide scalable mining and indexing approaches taking into account the content and social context of social networks
The SocialSensor Objective

Analyse – Organise – Deliver

DySCO

content

location

behaviour

usage

time

social context

Massive social media and unstructured web

Dynamic Social COnainers

News - Infotainment
The SocialSensor Vision

SocialSensor quickly surfaces trusted and relevant material from social media – with context.

• “quickly”: in real time
• “surfaces”: automatically discovers, clusters and searches
• “trusted”: automatic support in verification process
• “relevant”: to the users, personalized
• “material”: any material (text, image, audio, video), aggregated with other sources (e.g. web)
• “social media”: across all relevant social media platforms
• “with context”: location, time, sentiment, influence
The SocialSensor Challenges

- **Multi-modality**: e.g. image + tags
- **Rich social context**: spatio-temporal, social connections, relations and social graph
- **Inconsistent quality**: noise, spam, ambiguity
- **Huge volume**: Massively produced and disseminated
- **Multi-source**: may be generated by different applications and user communities
  - Also connected to other sources (e.g. LOD, web)
- **Dynamic**: Fast updates, real-time
Conceptual Architecture and Main components

- Real time dynamic topic and event clustering
- Trend, popularity and sentiment analysis
- Calculate trust/influence scores around people
- Personalized search, access & presentation based on social network interactions
- Semantic enrichment and discovery of services
DySCO Lifecycle

Stream of online data

Data Collection

Content Analysis & Mining

DySCOs

Feature Extraction, Indexing & Source Aggregation

Query

Delivery

Presentation

creation & maintenance

search, delivery & presentation
On-going results

• Social Data Collection and Analytics
• Multimodal Event Detection
• Efficient Content Delivery via CCN
• The News pilot application
• The Infotainment pilot application
Social data collection

- **Social media content** on Journalism and Infotainment

- **Modules:**
  - Social data analytics ➔ Context-aware focused crawling
  - Indexing low-level multimedia content and links between content items
  - Indexing data from different sources ➔ Aggregation
  - Multimedia retrieval and filtering

- **Task challenges:** filtering, sensing, analysis, cross-platform

- **Ongoing work:** *US elections* (6 November)
Social data analytics

• Pilot study: *Super Tuesday*
  – *Crawling:* Hashtags and mentions from Twitter
  – *Dates:* 5-8 March
  – *Dataset:* 284,732 tweets (with 100,832 distinct users)
  – *Analysis* of the frequency of mentions of candidates and states - *Temporal dimension* of the collection
  – *Real-world events* associated with Twitter activity – *Automatic* process
Social data analytics

- Newt Gingrich → Oklahoma, Alaska and Georgia states
Multimodal Event Detection

• Develop novel multimodal clustering techniques for grouping multimedia content that is associated with multiple modalities

• Encompass the social dimension of the content
  – take into account the profile and social relationships between the users that are associated to a multimedia item

• Extend to incremental and stream-based processing approach (on-line)
Multimodal Event Detection: progress

- Multimodal clustering integrating “supervisory signals”
- A known clustering is used, in order to supervise the multimodal fusion and clustering procedure
Multimodal Event Detection: first results

Tested on MediaEval 2011 Social Event Detection challenge data
Best NMI achieved by proposed method comparing to a standard multimodal clustering method

<table>
<thead>
<tr>
<th>Run</th>
<th>Baseline</th>
<th>With supervisory signal</th>
<th>Difference (abs. and %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8586</td>
<td>0.9469</td>
<td>0.0883 (10.28%)</td>
</tr>
<tr>
<td>2</td>
<td>0.7843</td>
<td>0.9197</td>
<td>0.1354 (17.26%)</td>
</tr>
<tr>
<td>3</td>
<td>0.7867</td>
<td>0.9164</td>
<td>0.1297 (16.48%)</td>
</tr>
<tr>
<td>4</td>
<td>0.8021</td>
<td>0.9062</td>
<td>0.1041 (12.97%)</td>
</tr>
<tr>
<td>5</td>
<td>0.7559</td>
<td>0.8939</td>
<td>0.138 (18.25%)</td>
</tr>
<tr>
<td>6</td>
<td>0.8410</td>
<td>0.8994</td>
<td>0.0584 (6.94%)</td>
</tr>
<tr>
<td>7</td>
<td>0.8153</td>
<td>0.9538</td>
<td>0.1385 (16.98%)</td>
</tr>
<tr>
<td>8</td>
<td>0.7097</td>
<td>0.8798</td>
<td>0.1701 (23.96%)</td>
</tr>
<tr>
<td>9</td>
<td>0.8058</td>
<td>0.9406</td>
<td>0.1348 (16.72%)</td>
</tr>
<tr>
<td>10</td>
<td>0.8223</td>
<td>0.8935</td>
<td>0.0712 (8.65%)</td>
</tr>
<tr>
<td>All</td>
<td>0.858</td>
<td>0.9538</td>
<td>0.0958 (11.16%)</td>
</tr>
</tbody>
</table>
Content Delivery via CCN

- **DASH**: Dynamic Adaptive Streaming over HTTP
  - Partition multimedia files into segments. Each segment is available in multiple quality levels
  - Select most appropriate segments based on network conditions

- **CCN**: Content-Centric Networking
  - A future Internet architecture to optimize content distribution
  - Content is identified by its name instead of location
  - Content is cached in the network close to end users, and is retrieved by name-based routing
  - Overlay over IP currently, and a clean state network mean to replace IP

- **DASC**: DASH over CCN
  - A CCN use case for real-time social media applications
  - Use CCN instead of HTTP
  - Implemented as a VLC Plugin
DASH over CCN – Preliminary results

• QoS analysis
  – CCN improves QoS, through caching content close to clients
  – DASH adaptive algorithm is based on end-to-end connections. Need to be re-designed for hop-by-hop architecture of CCN.

• Overhead analysis
  – Generates more overhead than DASH with HTTP
  – Constant overhead ratio when streaming from one node
  – Streaming from multiple sources introduces more overhead with small DASH segments, which needs to be further investigated.
1. discover emerging trends & topics

Tapping into social sources, journalists understand what the internet audience talks about…

2. aggregate content

Discover and combine relevant media content from different sources…

3. gain insights

Interact with the masses of multimedia data and understand social phenomena…

4. engage readers

Understand what is interesting to readers and recommend relevant stories enriched with rich media…

The News Pilot
Content verification is critical...

• Can we tell if a hot lead in social media corresponds to an actual event?
• Can we distinguish a hoax from a true story?
• Can we filter out propaganda?

How can we teach a machine to distinguish truth from lies?
By using the journalists’ ways...

**Contributor**
What can we find out about the source of information? How can we calculate trust, reputation and influence?

**Content**
Does the posted content look reliable? Can we find out if it has been posted before or manipulated?

**Context**
Does the 'what', 'when' and 'where' contextualise together?
Initial interface design prototypes...
Initial interface design prototypes...
Infotainment: Discovery and presentation of relevant content

Use Case sites

- Thessaloniki International Film Festival (TIFF) (www.filmfestival.gr)
  - 80,000 attendants / 100,000 visitors in 10 days
  - 150 films / 350 screenings
  - Film and Documentary festivals

- Fete de la Musique Berlin
  - 100,000 visitors every year

User Requirements: Most wanted basic features

- Browse schedule by category, date
- Personalised schedule
- Map
- Share and interact in Social Media

<table>
<thead>
<tr>
<th>Basic Features</th>
<th>Schedule</th>
<th>Alerts</th>
<th>Map</th>
<th>Friends</th>
<th>Ticket</th>
<th>My Schedule</th>
<th>Social Media share</th>
<th>Check-in/Who’s here</th>
<th>Music Video</th>
<th>Photos</th>
<th>Export</th>
<th>Social Media links/feeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDF14</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SocialSensor</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Infotainment

**User Requirements:**
*Most wanted advanced features*

- ✔ Social media-based film recommendations
- ✔ Sentiment analysis for screenings
- ✔ Real-time check-in heatmaps
- ✔ AR 3D interfaces with RT information layers
- ✔ Robust media sharing
- ✔ Smart location-based recommendations

<table>
<thead>
<tr>
<th>Advanced features</th>
<th>Rate</th>
<th>Interactive map</th>
<th>Media share</th>
<th>Recommendations</th>
<th>Arrange meetings &amp; events</th>
<th>What’s on near you</th>
<th>Now playing</th>
<th>Selling fast shows</th>
<th>AR for POIs around you</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDF14</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialSensor</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Infotainment - results

• Initial prototyping
  – 14th Thessaloniki Documentary Festival ThessFest application for iPhone

• Usage statistics
  ✓ 405 downloads
  ✓ 6,673 sessions
  ✓ App usage per day per user:
    ✓ 14 times for search
    ✓ 11 times for film browsing
    ✓ 15 times for personal schedule

Feature assessment questionnaire showed:
Thank you!

http://www.socialsensor.eu