Semantically-Enabled Optimization of Digital Marketing Campaigns

Vincenzo Cutrona¹, Flavio De Paoli¹, Aljaž Košmerlj², Nikolay Nikolov³, Matteo Palmonari¹, Fernando Perales⁴, and Dumitru Roman³

¹University of Milano - Bicocca
²Jožef Stefan Institute
³SINTEF DIGITAL
⁴JOT Internet Media
Digital Marketing

Unique ad efficiency & high-quality traffic generation company

- Impressions
- Clicks
- Conversion rates
- Locations
- Dates
- And > 40 additional variables more
Digital Marketing

- In marketing numbers:
  - Manages campaigns in **74 countries and 17 languages**
  - More than **3200 accounts** daily managed
  - About **250 million impressions & 8.7 million 'clicks' daily**
  - Over **3,800 million keywords** managed

Massive scale
Multiple traffic sources
Multiple languages
Performance driven
Business Motivation: Intuition & Desired Insights

Keyword: Burger at home

Keyword: Shaving Machines

Weather and other events affect the behavior of users in search engine platforms

WEATHER & EVENT-BASED DIGITAL MARKETING CAMPAIGN OPTIMIZATION
Business Motivation: Intuition & Desired Insights

Searching for advanced data insights

- Correlation between performance indicators and external variables
  - Temperature, probability of rain, light hours
- New trends and patterns
- Information for adjusting bids for the affected keywords

Keyword: Burger at home

Weather and other events affect the behavior of users in search engine platforms
Pilot Service: Weather-based Campaign Scheduler

New services for campaign optimization:

- Pilot service: **weather-based campaign scheduler**
  - Select the best dates in the upcoming week and for each region to launch the campaign with weather-sensitive keywords

- Additional services:
  - Advise clients about key keywords enabling higher impact depending on external factors
  - Analyse impact depending on campaign properties (country, topic, and timing)
Supporting the **Weather-based Campaign Scheduler**

- **Semantics highlights:**
  - Enrichment based on semantic table annotation approach
  - Reconciliation (Google GeoTargets to Geonames)
  - Graph-based weather data modeling (avoid replication)

- Machine learning for prediction
Modeling Data Enrichment with DataGraft + ASIA (#1)

1. Google GeoTarget \(\rightarrow\) GeoNames ids
2. Geonames ids \(\rightarrow\) GeoNames coordinates (via query)
3. GeoNames coordinates & time \(\rightarrow\) weather data (via APIs)

<table>
<thead>
<tr>
<th>KEYWORD</th>
<th>#im</th>
<th>REGION</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>194906</td>
<td>64</td>
<td>Thuringia</td>
<td>11/03/2017</td>
</tr>
<tr>
<td>517827</td>
<td>50</td>
<td>Bavaria</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>459143</td>
<td>42</td>
<td>Berlin</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>891139</td>
<td>36</td>
<td>Bavaria</td>
<td>11/03/2017</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEYWORD</th>
<th>#im</th>
<th>REGION</th>
<th>geol.</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>194906</td>
<td>64</td>
<td>Thuringia</td>
<td>50.91, 11</td>
<td>11/03/2017</td>
</tr>
<tr>
<td>517827</td>
<td>50</td>
<td>Bavaria</td>
<td>49, 11.5</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>459143</td>
<td>42</td>
<td>Berlin</td>
<td>52.5, 13.41</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>891139</td>
<td>36</td>
<td>Bavaria</td>
<td>49, 11.5</td>
<td>11/03/2017</td>
</tr>
</tbody>
</table>

DataGraft (Roman&al.SWJ2018) + ASIA (Cutrona&al.ISWCdemo2019)
Modeling Data Enrichment with DataGraft + ASIA (#2)

Optimization via graph-based weather data modeling:
- pre-fetch and storage in ArangoDB
- GeoNames ids added to weather data

<table>
<thead>
<tr>
<th>KEYWORD</th>
<th>#im</th>
<th>REGION</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>194906</td>
<td>64</td>
<td>Thuringia</td>
<td>11/03/2017</td>
</tr>
<tr>
<td>517827</td>
<td>50</td>
<td>Bavaria</td>
<td>-12/03/2017</td>
</tr>
<tr>
<td>459143</td>
<td>42</td>
<td>Berlin</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>891139</td>
<td>36</td>
<td>Bavaria</td>
<td>11/03/2017</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEYWORD</th>
<th>#im</th>
<th>REGION</th>
<th>geoid</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>194906</td>
<td>64</td>
<td>Thuringia</td>
<td>gn:2822542</td>
<td>11/03/2017</td>
</tr>
<tr>
<td>517827</td>
<td>50</td>
<td>Bavaria</td>
<td>gn:2951839</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>459143</td>
<td>42</td>
<td>Berlin</td>
<td>gn:2951839</td>
<td>12/03/2017</td>
</tr>
<tr>
<td>891139</td>
<td>36</td>
<td>Bavaria</td>
<td>gn:2951839</td>
<td>11/03/2017</td>
</tr>
</tbody>
</table>

DataGraft (Roman & al.SWJ2018) + ASIA (Cutrona & al.ISWCdemo2019)
Data Enrichment: Batch Execution

- Data transformations designed on samples using a UI
- Need a strategy to apply them to Big Data sources
  - Applied to the full dataset using a **scalable platform** in batch mode
  - Repeatable on new datasets that **hold the same structure** (e.g., campaigns in the same country for a different time period)

More info about architecture and tools at: [https://www.ew-shopp.eu/toolkit/](https://www.ew-shopp.eu/toolkit/)
Data Analytics

- **Regression problem**: predict the number of impressions for a keyword on a target day from weather features
- **Model**: random forest
- **Test**: data on regions in Germany over 15.5. - 15.7. 2017
- **Result**: RMSE of 0.77
Discussion: Business Value & Analytics

Summary of results

- **Enrichment pipeline**
  - Enriched data for 2016/2017: ~22M keywords from ~47K campaigns
  - In use to enrich data for 2018/2019

- **Analytics**
  - Promising keywords filtered by a data scientist
  - On a smaller number of keywords, **best results** (RMSE) judged **valuable for usage in production**
    - Finding peaks is enough for monetization purposes
    - Campaigns include many keywords when the analyses are scaled up

Limitations

- One model per keyword does not scale
  - too many, too noisy and too scattered data signals for keyword-level analytics

- Further exploration of the model parameters space
Discussion: Semantics

PROS

1. **Google GeoTargets to GeoNames reconciliation** relevant to enrichment of marketing data in real-world scenarios

2. Semantic enrichment is a valuable application for **table annotation**; so far focus on large scale and cloud-ready table annotation

3. Semantics and **graph-based modeling** can be useful in the background even when linked publication is not the application objective

CONS

1. Still limited spatial coverage of computed links: need for **better reconciliation methods** for tables

2. Work to be done to seamlessly integrate **smarter table interpretation approaches** to reduce users’ effort

3. Some ad-hoc data preprocessing still needed vs full approach based on table-annotation: more work needed to to understand **trade-offs between tabular vs graph-based representations**
Future Work

- **Deploy** the data enrichment & analytics framework on JOT premises (ongoing)
- **Cluster the keywords** based on their semantics using representation learning methodology and build the models at the cluster level (done: more robust models and efficient processing)
- Extend **Google GeoTargets to Geonames links** to support reconciliation
- Add context features regarding world events collected from news media by the **Event Registry** platform (ongoing)
Questions?

vincenzo.cutrona@unimib.it
matteo.palonari@unimib.it
dumitru.roman@sintef.no

More on services with event and weather-based data analytics and semantic enrichment tools at

ew-shopp.eu
ew-shopp.eu/data-blog

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreements No 732590 (EW-Shopp) and No 732003 (euBusinessGraph).