FashionBrain: Understanding Europe’s Fashion Data Universe

Gianluca Demartini
Senior Lecturer in Data Science
University of Sheffield, UK
FashionBrain Consortium

Project Duration 2017-2019. Funded under the H2020-ICT-14-2016 topic Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation. Total cost: 2.9M EUR.
Data Challenges in the Fashion Industry

• **Data is not used in an integrated fashion**
  - Data is stored in silos
  - Each data source owns data in different formats and platforms

• **Search engines dominate** customer behaviors
  - Search engines and social network companies are in a strong position to settle a new global monopoly on customer understanding thanks to AI technology
  - Retailers have limited access to customer data

• **Fashion bloggers influence** customer choices
  - Limited understanding on bottom-up changes in fashion
  - Shift from traditional stylist-driven fashion
FashionBrain Objectives

• Novel Shopping Experience: **Make Images Searchable**
  • Product search and recommendation
  • Moving away from the ’real human shop assistants’

• Shift Traffic away from Web Search Engines to **Retailer's Mobile Apps**
  • By providing custom shopping experiences and advanced search tools

• Detect Influencers and **Predict Fashion Trends**
  • Time Series Analysis
  • Social Media data

• **Share Insights** with Cross Industry Partner Network
  • Data Integration infrastructure based on HDFS and column stores
Exploitation of Business Data Assets

- Data integration and value creation across
  - Shop Inventory of around 100 online shops
  - Social Media (Fashion bloggers)
  - User reviews about fashion products in 11 languages
  - Product catalogs
  - OEM manufacturer information
  - Sales data
  - Web crawls
  - Crowd-generated data
Measurement of data exploitation

• Demos; Algorithms; Impact on sales;

<table>
<thead>
<tr>
<th>Data asset</th>
<th>Exploited by</th>
<th>Impact on business partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shop Inventory of around 100 online shops</td>
<td>Fashwell</td>
<td>Used for Photo-based product search demos</td>
</tr>
<tr>
<td>Social Media (Fashion bloggers)</td>
<td>Fashwell</td>
<td>Used for Photo-based product search demos</td>
</tr>
<tr>
<td>User reviews about fashion products in 11 languages</td>
<td>Zalando</td>
<td>Demonstrator of text join and relation extraction methods</td>
</tr>
<tr>
<td>Product catalogs</td>
<td>Zalando</td>
<td>Used for Photo-based product search demos</td>
</tr>
<tr>
<td>OEM manufacturer information</td>
<td>Zalando</td>
<td>Supply chain optimization</td>
</tr>
<tr>
<td>Sales data</td>
<td>Zalando</td>
<td>Time Series analysis; Trend detection demo</td>
</tr>
<tr>
<td>Web crawls</td>
<td>Beuth</td>
<td>Identification of fashion-related content; Demonstrator of text join and relation extraction methods</td>
</tr>
<tr>
<td>Crowd-generated data</td>
<td>All</td>
<td>Used for Photo-based product search demos; Trend detection demo</td>
</tr>
</tbody>
</table>
FashionBrain - Data Value Chain

• **New data** entity-centric **integration methods**
  • Named Entity Recognition and Linking across sources

• New Data Technology
  • **Time Series Analysis** for Fashion Trend Detection
  • **Deep Learning** based Text Mining

• Data-driven **demos**
  • Search by Image
  • Product Recommendation
FashionBrain - Searching and shopping query workflows
FashionBrain - Photo-based product search
FashionBrain Project Roadmap

- **Year 1 - Foundations**
  - Requirement Analysis
  - FashionBrain Schema
  - Datasets

- **Year 2 – Core techniques, data, demos**
  - Named Entity Recognition and Linking
  - Time Series Analysis for MonetDB
  - Crowd Data Aggregation
  - Social Media Annotations
  - Demos

- **Year 3 – Advanced applications**
  - Relation Extraction w/ Deep Learning
  - Final Demos (Trend Prediction, Search)
  - Business Plan
FashionBrain: Understanding Europe’s Fashion Data Universe

Gianluca Demartini, University of Sheffield, UK