Beyond reported history: Strikes that never happened

Martha van den Hoven,
Antal van den Bosch, Kalliopi Zervanou

ILK Research Group
Tilburg center for Cognition and Communication
Tilburg University

Aim of the project

History of the social movement in Europe, 1850-1940

O1: Analysis of primary text sources
O2: Analysis of secondary text sources
O3: Linking all of this information
O4: Visualising/Associating info on timeline

Events in time

Event

Time

Daily narrative

conflict strike occupation

resolution agreement

negotiation threat

Person
Occupation
Organisation
Location
Time
Computational Counterfactual History

Strike database
- Designed and created by Sjaak van der Velden
- Contains over 16,000 strikes in the Netherlands

Newspaper archives
- Koninklijke Bibliotheek in The Hague
  - Newspapers from 1618 to 1995
  - Over 1 million pages online, target 8 million
  - Searchable on title, date, words
Searching for articles on strikes

- Create query (model / classification rule)
- Most strikes in 1910-1940
- Search terms: staking, staken
  - Earlier: strike, grève, bollejeije, laveij
  - Later: werkonderbreking, stiptheidsactie
- From strike database record:
  - date and length
  - names of companies, unions, locations, occupations, sectors

Example query

```sql
SELECT all articles
CONTAINING THE WORDS
stak?n* AND
(blokband OR Amsterdam OR
taxichauffeur)
AND artikeldatum BETWEEN 6 apr 1937 - 7
AND 9 apr 1937 + 3
```
**Precision versus Recall**

- **Precision**: how many of the retrieved articles are relevant?
- **Recall**: how many of all relevant articles were found?

Trade-off between precision en recall

**Exploratory test**

- Sample based on “big semi-short” strikes
  - over 250 strikers
  - lasting 2 to 6 days
- 27 strikes in database
- Retrieved 333 articles
  160 (partly) relevant = 48%

**Ranking**

- Automatic ranking on relevance
- Compare all articles found with record; which is more similar?
- Strike record expanded with 84 “strike words” staking, arbeiders, werk, vergadering, bestuur, etc.

- Result: Ranking slightly improves results

** Strikes that never happened**

- Many strikes have a lead-in “prelude” period (often associated with unrest)
- Prelude does not necessarily lead to strike
- By finding prelude periods typical for strikes, can we find strikes that never happened?
Machine learning

- Computer can learn to discriminate between concepts
- Discriminate prelude articles from other
- Computer needs examples of both
- Finding prelude articles of known strikes

Manually found pre-strike articles

<table>
<thead>
<tr>
<th>Date</th>
<th>Days</th>
<th>Who, where</th>
<th>articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-6-1917</td>
<td>10</td>
<td>Mijnwerkers, Limburg</td>
<td>33</td>
</tr>
<tr>
<td>22-9-1919</td>
<td>41</td>
<td>Beurtvaarders, landelijk</td>
<td>5</td>
</tr>
<tr>
<td>25-5-1928</td>
<td>123</td>
<td>Scheepsbouwers, De Schelde, Vlissingen</td>
<td>22</td>
</tr>
<tr>
<td>8-4-1929</td>
<td>180</td>
<td>Zagerij/schaverij, Zaandam</td>
<td>11</td>
</tr>
<tr>
<td>16-11-1931</td>
<td>140</td>
<td>Katoenwevers, Jannink, Enschede</td>
<td>22</td>
</tr>
<tr>
<td>11-5-1938</td>
<td>58</td>
<td>Zeevissers, Katwijk, Scheveningen, Vlaardingen</td>
<td>4</td>
</tr>
<tr>
<td>misc</td>
<td></td>
<td>Articles found by browsing</td>
<td>11</td>
</tr>
</tbody>
</table>

Features

- Machine discriminates on basis of features
- Words from prelude articles
  - only non-auxiliary verbs, nouns, and adjectives
  - occurring more than 10 times
- Resulting in list of 208 indicator words

Word frequency
Discriminatory power

Finding strikes that never happened

- 53 most discriminative words
  - Gain ratio > 0.1, no names
- where to look?
- as a first approximate test, select “unrestful” weeks
- between 1-1-1910 and 31-12-1939
- measure fluctuations in degree of strike news

Found articles in hot weeks

<table>
<thead>
<tr>
<th>Week</th>
<th>Total found</th>
<th>Prelude</th>
<th>Strike</th>
<th>Foreign</th>
<th>Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>24-feb-1912</td>
<td>273</td>
<td>12</td>
<td>20</td>
<td>18</td>
<td>223</td>
</tr>
<tr>
<td>11-sep-1926</td>
<td>133</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>127</td>
</tr>
<tr>
<td>16-apr-1938</td>
<td>208</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>201</td>
</tr>
</tbody>
</table>

Found 14 prelude articles
Did some of them not happen?

Check with database:
1912: policemen in Amsterdam, tram conductors
    The Hague (2 articles), and teachers (3
    articles)
1926: Lion’s butchery in Boxmeer; transport
    workers of Steenkolen Handelsvereeniging
1938: Braat metal factory in Delft

Mean average precision with ranking: 0.43

Discussion

• Straightforward, limited, semi-manual test
  produces potentially usable results
• Up to historians to value the outcome and the
  procedure
• To what other research questions does this
  generalize?
• Underlying model: the narrative of an event
  over time
• Computational counterfactual history