Mining Careflow Patterns in Data Warehouses of Breast Cancer Patients

Lucia Sacchi, Daniele Segagni, Arianna Dagliati, Alberto Zambelli and Riccardo Bellazzi

Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy
IRCCS Fondazione Salvatore Mugeri, Pavia, Italy
Collecting a Data Warehouse of Oncology Patients

Collaboration between University of Pavia and the IRCCS Fondazione S. Maugeri (FSM) hospital (Pavia, Italy)

**ONCO-i2b2**

i2b2 DW infrastructure to integrate FSM hospital information system (HIS) and a cancer biobank that manages both plasma and cancer tissues

Currently data related to 28,838 oncology patients

Knowledge discovery to find frequent, meaningful paths and build structured scenarios
The Data Warehouse

Information on:
Admission, Day Hospital, Procedures, Drugs, Blood tests

For each patient: integrated temporal clinical history
How to Exploit this Data for Research?

Objective: extract the most frequent careflows from clinical histories

PROCESS MINING

- Mines workflows from event logs
- Doesn’t take into account quantitative clinical data (only “process” data)

TEMPORAL DATA MINING

- Able to extract frequent patterns from clinical data with temporal features (e.g. TARs)
- Cannot generate complex temporal histories (e.g. chains of events)
How to Exploit this Data for Research?

Objective: extract the most frequent careflows from clinical histories

**PROCESS MINING**
- Mines workflows from event logs
- Doesn’t take into account quantitiative clinical data (only “process” data)

**TEMPORAL DATA MINING**
- Able to extract frequent patterns from clinical data with temporal features (e.g. TARs)
- Cannot generate complex temporal histories (e.g. chains of events)
Methods

Use process mining techniques to extract frequent flows of care in the data warehouse

Enrich the mined careflows with clinical information
Methods

Use process mining techniques to extract frequent flows of care in the data warehouse

Enrich the mined careflows with clinical information
Our Process Mining Approach

Currently works on sequences of in-hospital and day hospital (DH) accesses to the different FSM wards

![Diagram showing sequences of hospitalizations and DH accesses](image)

- Hospitalization - Breast Surgery
- DH Oncology
- Hospitalization - Breast Surgery

Allows extracting the most frequent clinical paths on the basis of threshold parameters

Each step of a clinical pattern is enriched with temporal information
Clinical Histories Characterization

Admission BREAST SURGICAL ONCOLOGY
- 2397/4533 pts
- median: 8 interquartile range: 4 - 171 days

End

738 pts

N° patients

Event type

Temporal details

Day Hospital ONCOLOGY
- 1257 pts
- median: 15 interquartile range: 1 - 165 days
Methods

Use process mining techniques to extract frequent flows of care in the data warehouse.

Enrich the mined careflows with clinical information.
Enriching Clinical Histories

Event 1

Event 2

Event 3

Event 4

End

End
Enriching Clinical Histories

Event 1

Event 2

Event 3

Event 4

End

End

Procedures (ICD9-CM codes)
Information on drugs
Blood test results
Enriching Clinical Histories

Event 1

Event 2

Event 3

Event 4

End

End

Procedures (ICD9-CM codes)
Information on drugs
Blood test results

Procedures (ICD9-CM codes)
Information on drugs
Blood test results
Results – A Case Study on Breast Cancer

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with a diagnosis of malignant neoplasm of female breast (ICD9-CM code 174.0-174.9) or personal history of malignant neoplasm (ICD9-CM V10.3)</td>
<td>8.727</td>
</tr>
<tr>
<td>Number of patients with a single hospital admission</td>
<td>2.713</td>
</tr>
<tr>
<td>Number of patients with a single procedure</td>
<td>420</td>
</tr>
<tr>
<td>Number of patients with a single DH</td>
<td>227</td>
</tr>
<tr>
<td>Number of patients with at least one event before 2000</td>
<td>1446</td>
</tr>
</tbody>
</table>

4.568 subjects
Results

First access

admission_BREAST_SURGICAL_ONCOLOGY (complete)
2397/4533pts-median:8.25**%ile:4.75**%ile:171days-738out

738pts
median:25-25**%ile:20-75**%ile:34 days

1257pts
median:15-25**%ile:1-75**%ile:185days-684out

1257pts-median:15-25**%ile:1-75**%ile:108days-130out

194pts-median:1-25**%ile:1-75**%ile:1days-136out

194pts-median:1-25**%ile:1-75**%ile:369 days

194pts-median:1-25**%ile:1-75**%ile:369 days

136pts
median:36-25**%ile:16-75**%ile:136 days

58pts
median:20-25**%ile:7-75**%ile:22 days

58pts-median:1-25**%ile:1-75**%ile:108days-22out

363pts-median:136-25**%ile:74-75**%ile:216 days

98pts-median:7-25**%ile:3-75**%ile:23days-30out

98pts-median:7-25**%ile:3-75**%ile:230out

684pts
median:136-25**%ile:74-75**%ile:216 days

363pts-median:5-25**%ile:4-75**%ile:10days-141out

OUT (complete) 738pts

OUT (complete) 684pts

OUT (complete) 141pts

OUT (complete) 130pts

OUT (complete) 141pts

OUT (complete) 30pts

OUT (complete) 30pts

OUT (complete) 22pts

OUT (complete) 22pts

OUT (complete) 50pts
Most Frequent Clinical Pattern

Admission BREAST SURGICAL ONCOLOGY
2397/4568 pts
8 [4 – 171] days

Day Hospital ONCOLOGY
1257 pts
15 [1 – 165] days

738 pts

End

1257 pts

684 pts

End

End
Enriching the History - Procedures

Admission BREAST SURGICAL ONCOLOGY
2397/4533pts
8 [4 – 171] days

1257pts (1257)
Operations on the breast  n_pts:1237/1257
Radioisotope scan and functional study  n_pts:560/1257

Day Hospital ONCOLOGY
1257pts
15 [1 – 165] days

684pts (684)
Injection infusion of chemotherapeutic substances  n_pts:512/684
Microscopic examination  n_pts:299/684

End
Enriching the History – Blood Tests

Admission BREAST SURGICAL ONCOLOGY
2397/4533pts
8 [4 – 171] days

1257pts (30)
- CA15.3 normal n_pts:15/23
- CA15.3 normal stable n_pts:7/23
- CEA normal n_pts:3/4
- Leukocytes low n_pts:3/8
- Leukocytes normal n_pts:3/8
- Platelets low n_pts:3/7
- Platelets low stable n_pts:3/7

Day Hospital ONCOLOGY
1257pts
15 [1 – 165] days

684pts (86)
- CA15.3 normal n_pts:44/46
- CEA normal n_pts:55/56
- Leukocytes low n_pts:10/12
- Platelet normal n_pts:9/13
- Platelet low n_pts:4/13

End
Patients with Complications

- Admission BREAST SURGICAL ONCOLOGY
  - 2397/4533pts 8 [4 – 171] days

- 1257pts 8 [4 – 171] days

- Day Hospital ONCOLOGY
  - 1257pts 15 [1 –165] days

  - 363pts 47 [26 – 109] days
    - Admission BREAST SURGICAL ONCOLOGY
      - 363pts 5 [4 - 10] days

  - 98pts 21 [5 – 124] days
    - Admission ONCOLOGY
      - 98pts 7 [3 – 23] days

  - 164pts 36 [16 – 136] days
    - Day Hospital ONCOLOGY
      - 164pts 97 [15 – 351] days

  - 58pts 20 [7 – 22] days
    - Day Hospital ONCOLOGY
      - 58pts 102 [49 – 353] days
Conclusions

- Heterogeneous information collected in the ONCO-i2b2 Data Warehouse used for research.
- A methods able to extract frequent careflows composed by clinical and administrative events.
- Data Mining approaches exploitation to create an enriched careflow model.

- Add more drugs to the analyzed spectrum (only oncology drugs up to now)
- Stratify patients on the cancer morphological features to check if different types of tumors lead to different careflows.
Thank you