Knowledge Discovery from Cohorts, Electronic Health Records and further Patient-related data

Myra Spiliopoulou

KDD 2018, London, August 19

FACULTY OF
COMPUTER SCIENCE

Cohorts

[Glenn, 2005]

The term “cohort”

Quoting [Glenn, 2005], page 2: “The term *cohort* originally referred to a group of warriors or soldiers, and the term is now sometimes used in a very general sense to refer to a number of individuals who have some characteristic in common.”

The term “cohort” in “cohort analysis”

Quoting [Glenn, 2005], page 2: “Here and in other literature on cohort analysis, however, the term is used in a more restricted sense to refer to those individuals (human or otherwise) who experienced a particular event during a specified period of time. The kind of cohort most often studied by social scientists is the human *birth cohort*, that is, those persons born during a given year, decade, or other period of time.”

Cohort Analysis

[Glenn, 2005]

The term "cohort analysis"

Quoting [Glenn, 2005], page 3: "The term *cohort analysis* is usually reserved for studies in which two or more cohorts are compared with regard to at least one dependent variable measured at two or more points in time."

Purposes of Cohort Analysis [Glenn, 2005], pages 1-2

- "Assessing the effects of aging"
- "Understand[ing] the sources and nature of social, cultural and political change."

Counter-examples – [Glenn, 2005], page 3

- *Cross-sectional study*: Comparison of different groups of individuals with respect to some characteristic/variable – such a study "is conducted with data collected at one point in time, or, more accurately, within a short period of time."
- *Panel study:* Comparison of the the attitudes of a group of individuals at two distinct timepoints – such a study "measures the characteristics of the same individuals at more than one point in time."

Turning from population-based studies towards studies on hospital data

- Example of a retrospective study on hospital data
- Building and refining a cohort with help of expert inputs
- Involving the experts for labeling
- Using cohorts for experiments
- Validation of the findings on another cohort

Disorders associated with Charcot Foot [Munson et al., 2014]

Charcot Foot is a rare disease: the bones/joints get brittle and disintegrate.

- Charcot Foot usually follows a bone injury.
- It often appears as followup of diabetes.
- Some risk factors are known, but the pathogenesis is not completely understood.

Goal of the study is to identify novel associations between Charcot Foot and other disorders/diseases,
paying particular emphasis on the temporal relationship in such associations.

The chase for Charcot Foot cases

- *Site of the study:* University of Michigan Health System (UMHS), encompassing three hospitals with six speciality centers (including a diabetes center with a podiatric clinic)
- *Complete dataset:* 1.6 million patients with 41.2 million ICD-9 codes (timestamped)
- *Candidates for Charcot Foot diagnosis:* “arthropathy associated with a neurological disorder” (ICD-9 code 713.5), amounting to 388 patients.

Codes associated with Charcot Foot [Munson et al., 2014]

Method

- Reviewing by Experts to separate among (1) well-known associations, (2) associations that were less known / had the potential to be novel, (3) uninformative associations – either because the ICD was unspecific [1] or because it was a misdiagnosis [2] that was later followed by the correct one, namely ”Charcot Foot”
- Investigation of the role of diabetes by separating between patients with Charcot Foot and diabetes (n=282), and those with Charcoot Foot but without diabetes (n=106) and investigating the dominant associations
- Ranking of the associations on p-values and odds ratio
- Testing the significance of the temporal relationship, i.e. when another diagnosis precedes the 713.5 diagnosis, using binomial test and $p < 0.001$ [3]

[1] unspecific ICD, e.g. ”viral infection, not otherwise specified”

[2] misdiagnosis like ”“gout, not otherwise specified”

[3] The test was on whether the one ICD-9 code preceded the other in a non-random way.

Main Findings

[Munson et al., 2014]

676 (of 710) associations with p-value $<$ 0.001; 603 with odds ratio $>$5.0

- Some were not reportedly linked to Charcot Foot but can be associated to it on the basis of existing etiology models. (e.g. bladder disorder; diseases/disorders associated with neurotrophic influences)
- Some diagnoses could be explained by diabetes, e.g. obesity, peripheral neuropathy.
- Associations that did not fit to etiology models but had very high odds ratio were: alkalosis, pulmonary eosinophilia [4], esophagean reflux [5]

111 ICD-9 codes with significant temporal relationship to Charcot Foot

- Four of them followed Charcot Foot (327.23 ”obstructive sleep apnea”, 786.7 ”abnormal chest sounds”, 353.6 ”phantom limb syndrome”, 786.9 ”nonspecific symptoms involving the chest and respiratory system”)
- Alkalosis preceded Charcot Foot 100% of the times; pulmonary eosinophilia also preceded it (significantly).

[4] Pulmonary eosinophilia may be treated with steroids; these may affect bone mineral density.

[5] Esophagean reflux might be associated to proton pump inhibitors; -//- -//- -//-

Turning from population-based studies towards studies on hospital data

- √ Example of a retrospective study on hospital data
- Building and refining a cohort with help of expert inputs [Zhang et al., 2014]
- Involving the experts for labeling
- Using cohorts for experiments
- Validation of the findings on another cohort

Iterative Cohort Analysis and Exploration [Zhang et al., 2014]

- **Goal:** get new insights about a population of patients (e.g. all patients of the cardiology unit who have hypertension)
- **Parties involved:** team of physicians + team of technologists
- **Data:** EHR of hospital patients (timeseries of patient recordings)

Conventional workflow – from [Zhang et al., 2014] with extensions

At the beginning, there is a question/observation – a concrete phenomenon that must be explained (cf. use cases in [Zhang et al., 2014]).

1. The (team of) physician(s) devise one or more hypotheses.
2. The physicians specify the cohort needed for the study of each hypothesis, possibly in interaction with a data analyst or DB expert.
3. The DB expert writes scripts to create the cohort and extract the data.
4. Data analysts build models according to the instructions of the physicians, e.g. on age and gender adjustment.
5. Physicians become a presentation/visualization of the model(s) and check whether their hypothesis is supported.
6. If necessary, GOTO 2.

Iterative Cohort Analysis and Exploration [Zhang et al., 2014]

Expanding the workflow to incorporate . . . [Zhang et al., 2014]

- *Early cohort definition:* The physicians must be able to define a cohort themselves in an ad hoc way, whenever they see fit (cf. steps 2 and 3 of the conventional workflow).
- *Flexible visualization:* The physicians must be able to inspect the cohort in different ways, without having to ask the technologists.
- *Flexible analysis:* The physicians must be able to invoke analytics modules and use them to perform analytics tasks without having to ask the technologists.
- *Cohort refinement and expansion:* The physicians must be able to modify themselves the cohort, i.e. the choice of patients and the choice of variables on them (cf. steps 6 and 1 of the conventional workflow).
- *Iterative analysis:* Cohort definition, visualization, analysis, refinement and expansion may need to be performed repeatedly, on the results of the previous iterations.

i.e. foster interaction between physician and system in a complete workflow, taking the technologists out of the workflow.

Iterative Cohort Analysis and Exploration [Zhang et al., 2014]

The elements of CAVA:

- **Cohorts:** Data construct

A cohort is a choice of individuals with their properties (feature space)

Inner feature space: set of properties shared by all cohort members
Outer feature space: set of all properties of the cohort members

- **Views:** Visualization components (library)

A view is a visualization component that

- presents a cohort to a user, and
- allows the user to modify the cohort interactively.

- **Analytics:** Computational elements (library)

An analytics component is

a piece of software that creates or modifies a cohort.

Iterative Cohort Analysis and Exploration [Zhang et al., 2014]

High-level architecture of CAVA (fig. 3, page 9)
Figure removed

Data provenance

- *Population database:* contains all information about all individuals in the population; is expanded by new information (derived via analytics or views)
- *Cohort database:* contains the description of each cohort (as defined by the user) and the IDs of the cohort members

Iterative Cohort Analysis and Exploration [Zhang et al., 2014]

Placing the CAVA elements into a workflow f(fig. 5, page 11)

Figure removed

Analytics components in CAVA

- *Batch analytics modules*, including a ”demographics module” and a ”risk stratification module”
- *On-demand analytics modules*, including a ”patient similarity component” (published in AMIA 2010), a ”utilization analysis component” (published in AMIA 2012) and a ”heart failure risk assessment component” (published in AMIA 2012)

Iterative Cohort Analysis and Exploration [Zhang et al., 2014]

Evaluation by a domain expert - a very experienced emergency room physician, also having long experience in hospital management

Usability and design

- Ease-of-use and speed in comparison to the typical procedure: only a couple of days would be needed to build a cohort, in comparison to at least two weeks for answering basic questions
- More statistics are needed, next to the graphical views e.g. to conclude whether there were enough patients (in support of some finding)

Applicability to the challenges of healthcare

- Appropriate for quick and easy experimentation on patient groups
- Patient similarity function is a very promising aid:
  - \+ for finding similar patients, if the cohort being built is too small
  - \+ in combination with on-demand-analytics, which can show trends of interest to the physicians
- CAVA workflow agrees with the way things are being done
- Limited amount of patient detail – physicians need also texts etc

Turning from population-based studies towards studies on hospital data

- √ Example of a retrospective study on hospital data
- √ Building and refining a cohort with help of expert inputs [Zhang et al., 2014]
- Involving the experts for labeling [Nissim et al., 2016, Nissim et al., 2017]
- Using cohorts for experiments
- Validation of the findings on another cohort

Active learning with multiple labelers on hospital data

[Nissim et al., 2017] and earlier works

Application area: Classification of condition severity (severe vs mild)
Input: SNOMED-CT and EHR

- CAESAR: Classification Approach for Extracting Severity Automatically from Electronic Health Records (earlier work)
  Labels are delivered by medical experts who inspect the conditions and decide between *severe* and *mild*.
- CAESAR-ALE: CAESAR with Active Learning Enhancement
  [Nissim et al., 2016]
- CAESAR-ALE followup: exploit inputs from labelers who vary in their expertise [Nissim et al., 2017]

Figure removed

Workflow of CAEASAR-ALE – from [Nissim et al., 2017]

CAESAR-ALE procedure

[Nissim et al., 2016]

AL core for an SVM-classifier

Method that returns for each condition: (i) the confidence of the classifier to the label and (ii) the distance of the condition to the separating hyperplane.

AL methods

- SVM-Simple-Margin, proposed by Tong & Koller (JMLR 2000-2001)
- Method “Exploitation”: selects conditions which
  - are at the “severe” side
  - are far from the separating hyperplane
  - are far from each other
- Method “Combination_XA” that exchanges between the two other methods (one trial each) for a total of n trials

The dataset used in the CAESAR-ALE experiments

[Nissim et al., 2016]

Figures removed

CAESAR-ALE to reduce labeling effort [Nissim et al., 2016]

Experiment 1 on 516 conditions (144 severe, 372 mild) in 10 randomly selected datasets for 10-fold cross validation

- Learning an initial classification model on 6 conditions
- Active learning on a pool of $n =$310 conditions:
  AL method chooses $k =$5 conditions at a time
  until the whole pool is processed
  and presents them to the expert
- Test on $m =$200 conditions
- Estimate savings as

$$Reduction\_In\_Labeling\_Effort * Cost\_of\_Labeling(10529\ conditions) * 3$$

using 3 physician labelers (120 USD per hour)

CAESAR-ALE: Classification performance

[Nissim et al., 2016]

Experiment 1 on 516 conditions (144 severe, 372 mild) in 10 randomly selected datasets for 10-fold cross validation Figures removed
Accuracy (left) and True Positive Rate (right), where Positive $\rightarrow$ "severe"

Differences among labelers [Nissim et al., 2016]

Experiment 2 with 3 physicians (have completed residency training) & 4 informatics experts (have at least a master degree)

- Learning an initial classification model on 6 conditions
- Active learning on a pool of $n =$100 conditions:
  AL method chooses $k =$5 conditions at a time
  until the whole pool is processed
  and presents them to the expert
- Test on $m =$410 of the 516 conditions in the gold standard

Followup on labeler performance [Nissim et al., 2017]

Looking deeper into: the behaviour of 3+4 labelers with different levels of expertise (cf. protocol of [Nissim et al., 2016])

1. The labeling process of each labeler corresponds to a learning curve. Do the learning curves observed when answering to AL queries show different variability than the curves observed under passive labeling? Inter-labeler and Intra-labeler variability
2. Given a group of labelers and a majority consensus scheme for them, how does it affect the quality of the model learned actively vs passively?

Findings:

> *"The use of AL methods, (a) reduces intra-labeler variability in the performance of the induced models during the training phase, . . .*
> *and (b) reduces Inter-labeler performance variance, and thus reduces the dependence on the use of a particular labeler.*
> *In addition, the use of a consensus label, agreed upon by a rather uneven group of labelers, might be at least as good as using the gold standard labeler, who might not be available, and certainly better than randomly selecting one of the group's individual labelers."*

CAESAR-ALE: Finding severe conditions

[Nissim et al., 2016]

Figures removed

The two CAESAR-ALE samplers

- find the severe conditions faster; after 62 trials, they have found more than 80 of the 144 severe conditions.
- lead to lower variance among the labelers than the SVM-based sampler

CAESAR-ALE: Finding severe conditions

[Nissim et al., 2016]

The two CAESAR-ALE samplers

- find the severe conditions faster; after 62 trials, they have found more than 80 of the 144 severe conditions.
- lead to lower variance among the labelers than the SVM-based sampler

Remarks

- The SVM-based sampler aimed to maximize separation, rather than to oversample from the class “severe”.
- The CAESAR-ALE samplers favoured sampling from the class “severe”. If the labelers noticed that, would this have influenced their behaviour (labeling speed and choice of label)?
- Severe conditions are found faster, but what would be the cost of finding all severe conditions?

Turning from population-based studies towards studies on hospital data

- √ Example of a retrospective study on hospital data
- √ Building and refining a cohort with help of expert inputs [Zhang et al., 2014]
- √ Involving the experts for labeling [Nissim et al., 2016, Nissim et al., 2017]
- Using cohorts for experiments [Deschamps et al., 2013, Niemann et al., 2016]
- Validation of the findings on another cohort [Hielscher et al., 2018]

The validation issue

- √ Model validation
- ? Validation of the findings

on a dataset drawn independently from the same population

Constraint-based learning and Validation on Cohorts

[Hielscher et al., 2018]

The DIVA framework:

- Discovery: Given cohort dataset D and a set of ML/NL constraints, find groups of participants
  within subspaces
  which best describe the concept, as reflected in the constraints,
  where “best” refers to participant similarity/separation and constraint satisfaction.
- Inspection: Given these groups (subpopulations),
  provide ways to identify and analyze the most distinct ones w.r.t. to the medical outcome.
- Validation: Enable experts to investigate whether discovered supopulations are generalizable or not.

DIVA components

[Hielscher et al., 2018]

Figure removed

Validation Procedure in DIVA

[Hielscher et al., 2018]

Figure removed

Validation example using DIVA [Hielscher et al., 2018]

Learning and Validation with two cohorts:

SHIP-CORE (3rd wave) & SHIP-TREND (1st wave)

Figure removed

Validation example using DIVA [Hielscher et al., 2018]

Learning and Validation with two SHIP cohorts: Discovery of subpopulations with increased prevalence of hepatic steatosis

CORE (3rd wave) for learning & and TREND (1st wave) for validation

Figure removed

Closing on **Cohorts from hospital data**

- a wealth of data
- many people, many recordings
- urgent scientific questions

State of affairs:

- $\surd$ ML methods for the analysis of hospital data
- $\surd$ Interactive methods helping a medical expert to build and refine a cohort
- $\surd$ Interactive methods for label acquisition
- $\surd$ Methods for validation on another, comparable cohort

Open issues

- $\rightarrow$ The data are collected for patient treatment, not for analytics
  $\Rightarrow$ Understand the data origin and original purpose
- $\rightarrow$ Insights are needed to improve diagnostics and therapy
  $\Rightarrow$ Distinguish between clinical decision support and clinical research
- $\rightarrow$ The medical expert reigns
  $\Rightarrow$ Involve the expert
  $\Rightarrow$ Distinguish between the role “medical researcher” and the role “treating physician”

VISIT THE KMD LAB:

- http://www.kmd.ovgu.de/
- Faculty of Computer Science, Otto-von-Guericke-University Magdeburg
- Sendmail at: myra@ovgu.de

- Thank you!

Acknowledgements: German Research Foundation project OSCAR
“Opinion Stream Classification with Ensembles and Active Learners”

Bibliography

[Deschamps et al., 2013] Deschamps, K., Matricali, G. A., Roosen, P., Desloovere, K., Bruyninckx, H., Spaepen, P., Nobels, F., Tits, J., Flour, M., and Staes, F. (2013).
Classification of forefoot plantar pressure distribution in persons with diabetes: A novel perspective for the mechanical management of diabetic foot?
PLOS ONE, 8(11):e79924.

[Glenn, 2005] Glenn, N. D. (2005).
Cohort Analysis.
Quantitative Applications in the Social Sciences. SAGE, 2nd edition.

[Hielscher et al., 2018] Hielscher, T., Niemann, U., Preim, B., Völzke, H., Ittermann, T., and Spiliopoulou, M. (2018).
A framework for expert-driven subpopulation discovery and evaluation using subspace clustering for epidemiological data.
Expert Systems with Applications, 113:147 – 160.

[Munson et al., 2014] Munson, M. E., Wrobel, J. S., Holmes, C. M., and Hanauer, D. A. (2014).
Data mining for identifying novel associations and temporal relationships with charcot foot.
Journal of Diabetes Research, 2014.

Bibliography

[Niemann et al., 2016] Niemann, U., Spiliopoulou, M., Szczepanski, T., Samland, F., Grützner, J., Senk, D., Ming, A., Kellersmann, J., Malanowski, J., Klose, S., and Mertens, P. R. (2016).
Comparative clustering of plantar pressure distributions in diabetics with polyneuropathy may be applied to reveal inappropriate biomechanical stress.
PLOS ONE.
accepted in August 2016.

[Nissim et al., 2016] Nissim, N., Boland, M. R., Tatonetti, N. P., Elovici, Y., Hripcsak, G., Shahar, Y., and Moskovitch, R. (2016).
Improving condition severity classification with an efficient active learning based framework.
Journal of Biomedical Informatics, 61:44–54.

[Nissim et al., 2017] Nissim, N., Shahar, Y., Elovici, Y., Hripcsak, G., and Moskovitch, R. (2017).
Inter-labeler and intra-labeler variability of condition severity classification models using active and passive learning methods.
Artificial Intelligence in Medicine, 81:12 – 32.
Artificial Intelligence in Medicine AIME 2015.

[Zhang et al., 2014] Zhang, Z., Gotz, D., and Perer, A. (2014).
Iterative cohort analysis and exploration.
Information Visualization (Info Vis), pages 1–19.