Introduction

1.1 Project Main Goal

The main project purpose is to follow patient's trajectories and find possible pathways for different diagnosis.

1.2 Available Data

According to the Norwegian Data Inspectorate's regulation, anonymous data[1] should be used for research purposes. Anonymization is not subject of this thesis; another student is responsible for that matter. But because of this policy I could not have access to the main data.

First data format I had received was text format file with 10 columns of entities and each row represented an event. Each column was separated by ";". The Entities were DAGB, POLK, INNL, LOS, DRG, HOVEDDIAG, BIDIAG 1, PROSEDYRE, KJØNN and ALDER.

DAGB: the date that patient has visited physician

POLK: patient was not hospitalized, out patient

INNL: patient was hospitalized, in patient

LOS: long of stay

DRG: Diagnosis Related Group. It includes diagnosis and what hospital could provide for the patient.

HOVEDDIAG: The main diagnosis ICD-10

BIDIAG 1: first possible diagnosis

PROSEDYRE: procedure that the patient had (NCMP/ NCSP codes)

INNDATO: when patient has arrived at hospital

KJØNN: 1 for mail, 2 for femail

ALDER: age

Sample of a row was like: 0 ; 1 ; 0 ; 0 ; 4620 ; R298 ; ; ; 3/25/2012 ; 2 ; 48

The main problem with this group of entities was, patient's trajectory was not possible to follow. But all the information that was needed for analyzing trajectories was given.

---

[1] Anonymous data: data from which the name, personal identity number and other characteristics serving to identify a person have been removed, so that the data can no longer be linked to a natural person [Datatilsynet12].
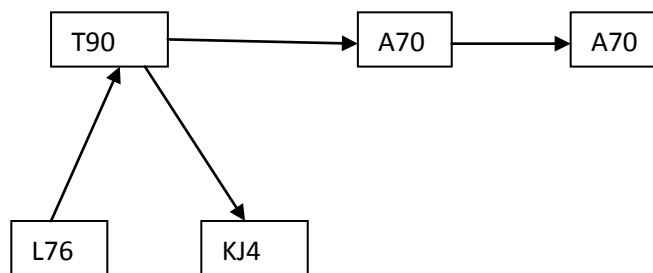
The second group of entities that were given was LØPENRNY, DATEDIFF, ICD-10 PROSEDYRE and DRG. In this group Løpenrny stands for patient id number, which is not real. Datediff stands for the time patient has visited a physician.

Sample of a row is like: 1 ; 0 ; S63 ; TND30 ; 808Y

In this new format there is no possible way to track the patients, but also a lot of information is omitted.

1.3 Our approach to the problem

First I need to transform rows (records) to event stream. For this matter I have thought of also making events more generalized. For instance patient number 21 has been in hospital 5 times in a row for same diagnosis (ICD-10: G80) and has received same treatment (DRG: 901O). I want to make one event stream out of these 5 rows; the format could be like PatientId # time # Diagnosis (ICD-10) # Procedure (DRG and NCMP/ NCSP). For Time entity because I do not have a clear description of it, I have decided to make a format like: first arrival-last arrival/ times of arrival (0 – 251 / 5).

I have the idea of making diagnosis like the diagram above, from another student's project [N05]. In this diagram it is easy to follow trajectories of diagnosis and know what diagnosis may end in another one. It also can help physician to know what he should expect to happen next.

From each diagnosis I can show different procedures that have happen to it, For instance in case of G80 in ICD--10, patient 21 had 901O, 997O and 998O in DRG codes, and 42KJ in NCMP/ NCSP codes.

I also have decided to make database tables for patient, time/ diagnosis and procedures (DRG, NCMP/ NCSP). I think it must be easier and faster to work with database.

References:

[Datatilsynet12] Act of 18 May 2001 No. 24 on Personal Health Data Filing Systems and Processing of Person Health Data (Personal Health Data Filing System Act). Last update July 10, 2010.

[N05] stein Jakob Nordbø. Visualising Collections of event sequences from general practice patient records. Decempber 20, 2005.