
Logisk analyse av medisinske databaser

DOKTORAVHANDLINGER

ALEKSANDER ØHRN

Email: aleks@idi.ntnu.no

Schønings gate 1

0356 Oslo

ØI skjæringspunktet mellom felter som kunstig intelligens, databaseteknikk, logikk og statistikk har det dukket opp metoder for automatisk å oppdage mønstre i datamengder. Slike mønstre kan for eksempel være i form av if-then-regler, og kan være med på å belyse sammenhenger i datamaterialet som man ellers ikke ville ha oppdaget. Mønstrene kan også danne en basis for å utvikle prediktive modeller, eller for klyngeanalyse. En slik logikkbasert analyse av data kan komplettere og fungere som et verdifullt supplement til mer tradisjonell statistisk dataanalyse.

Avhandlingen tar utgangspunkt i hvordan logiske modelleringsmetoder basert på betraktninger omkring skillbarhet kan brukes for å analysere medisinske data. For å muliggjøre denne typen analyse er et større programvaresystem, ROSETTA, blitt utviklet. Videre foreslås flere mulige anvendelser, som demonstreres med simuleringer på reelle pasientdatabaser.

Tre ulike klasser av anvendelser relevante for det medisinske domenet belyses. Anvendelsene er tilsynelatende ulike, men deler alle det samme matematiske grunnlaget:

- Enkelte medisinske tester med både høy sensitivitet og spesifisitet kan iblant være ønskelig å anvende sparsomt, for eksempel fordi de er kostbare eller invasive. En metode foreslås for å identifisere den undergruppen av pasienter i en database der en viss test er utslagsgivende med hensyn til å kunne foreta en diagnose eller prognose, og der det ikke eksisterer kombinasjoner av andre enklere og rimeligere biter av informasjon som gjør det mulig å kunne foreta den samme diagnosen eller prognosen. I simuleringer med pasientdata fra et prognostisk problem innen kardiologi var den identifiserte gruppen ikke spesielt stor, noe som antyder muligheten for ressursbesparelser.

– Medisinske opplysninger om individer kan være følsomme og ha et stort potensial for misbruk dersom de kommer i gale hender. Selv om umiddelbart identifiserende felter som navn eller fødselsnummer er blitt slettet fra en pasientdatabase, kan krysskoblinger mot eksterne registre ofte gjøre det enkelt å reidentifisere individer i en slik tilsynelatende anonymisert samling data. En algoritme for anonymisering presenteres som sikrer at det for enhver pasient i en database vil finnes et tilstrekkelig antall andre ”like” pasienter, slik at en reidentifisering vanskeliggjøres.

– Forskning innen informatikk har de siste tiårene frembrakt metoder som ”lærer” fra eksempler. Ved å anvende slike teknikker på data avledet fra pasientjournaler, kan man automatisk få frem minimale diagnostiske eller prognostiske modeller i form av if-then-regler. Slike regler kan for eksempel brukes i et system for beslutningsstøtte. Både regler for å diagnostisere akutt blindtarmsbetennelse og regler for å diagnostisere hjertefeil viser seg i simuleringer å fungere vel så bra som andre typer modeller referert til i litteraturen, og kommer også godt ut av det sammenliknet med legers diagnostiske nøyaktighet.

- *Avhandlingens tittel*
 - Discernibility and rough sets in medicine: tools and applications
 - *Utgår fra*
 - Institutt for datateknikk og informasjonsvitenskap
 - og
 - Harvard Medical School
 - Boston, USA
 - *Disputas* (dr.ing.) 18.2. 2000
 - Norges teknisk-naturvitenskapelige universitet
-

Publisert: 10. januar 2001. Tidsskr Nor Legeforen.

© Tidsskrift for Den norske legeforening 2026. Lastet ned fra tidsskriftet.no 24. februar 2026.