

# Neuroinformatik: Maskinforståelse af hjernen?

*Finn Årup Nielsen, Neurobiologisk Forskningsenhed, Københavns Universitetshospital Rigshospitalet og Informatik og Matematisk Modellering, Danmarks Tekniske Universitet.*

Hjernevidenskaben skaber en lind strøm af data og information, og særligt indenfor den del af hjernevidenskaben der anvender hjerneskantere baseret på *magnetisk resonans* og *positron-emissions-tomografi* kan datamængden være stor: Et enkelt studie fra en moderne skanner producerer data i gigabyte størrelsen. Håndtering og analyse af sådanne data er utænkelig uden brug af computere.

Data indenfor genetikken har længe været så omfattende og kompleks at det har ført til udvikling af en række databaser og værktøjer, hvoraf mange findes frit tilgængelige på Internettet. Disse services gør det muligt at søge og analysere data på tværs af de enkelte studier. Lignende er ved at ske indenfor hjernevidenskaben: Den vision man har er en “maskine”, hvor man lægger al kendt viden om hjernen ind, trykker på Den Store Knap, og ud kommer svarene på hjernens store spørgsmål. Der er lang vej til dette mål kan nås, men man er ved at komme i gang med at oprette hjernevidenskabelige databaser og computerbaserede metoder til at *datamine* og opdage interessant viden ved at sammenkoble og analysere informationen i databaserne. Det felt, der beskæftiger sig med at udvikle de computerbaserede værktøjer til dette, kaldes *neuroinformatik*.

Der har i nogen tid eksisteret neuroinformatiske programmer til billedanalyse af hjerneskan. Størstedelen af programmerne er frit tilgængelige på Internettet, og de forskellige forskergrupper konkurrerer om at få andre forskere til at bruge netop deres program. Hvad der savnes, er programmer til at sætte det enkelte studie i forhold til andre studier.

Data fra hjerneskantere er ikke den eneste type hjernedata. Generelt kan hjernevidenskabelig data være ganske blandet: De kan for eksempel komme fra kemiske analyser, elektriske målinger og psykologiske test. Den neuroinformatiske opgave bliver derfor at håndtere og analysere store datamængder, samkøre de forskelligartede data og præsentere resultatet på en forståelig måde.

## Tekstmining: Maskinforståelse af tekst

Langt det meste hjernevidenskabelige information bliver stadig kommunikeret i form af tekst — tekst i videnskabelige artikler. Web-baseret informationssøgning i databaser, der indekserer disse artikler, er blevet en uundværlig

del af en hjerneforskers måde at opnå viden på. Den amerikanske PubMed Database indekserer 15 millioner videnskabelige artikler, hvoraf over 800.000 returneres hvis man søger på ordet “hjerne”, og Google Scholar databasen returnerer over 2 millioner videnskabelige kilder på samme søgning. Google Scholar indekserer også de videnskabelige henvisninger, der findes bagerst i stort set alle artikler. Har man fundet en artikel, kan man således finde tilbage til de artikler som citerer den. PubMed gør det muligt at finde lignende artikler, og søgemaskinen Vivísimo kan organisere PubMeds søgeresultater i klynger efter temaer.

Sådanne metoder, der søger efter temaer i et sæt af tekster, kaldes “tekstmining”, og kan for eksempel også anvendes til at finde de mest dominerende hjernefunktioner for et bestemt hjerneområde: Et hjerneområde udvælges, og alle artikler, der refererer til hjerneområdet, hentes fra PubMed. Teksten i artiklerne konverteres til en form, der kan forstås af en algoritme, som grupperer teksterne, og ud kommer klynger af artikler. Algoritmen behøver ikke at få fortalt, hvordan den skal gruppere tekstene. Det gøres ud fra, hvor hyppigt hvert ord forekommer i artiklerne. Algoritmen lærer så at sige konteksterne automatisk.

Ved at undersøge hvem der citerer hvem, er det med en lignende tekstminingsmetode muligt, automatisk at udpege nøglepersoner indenfor et bestemt forskningsområde. Når en forsker har været en årrække i feltet, opbygger hun en viden om, hvem der er særlige eksperter indenfor hendes felt — måske uden at lægge specielt mærke til det. Computere, der analyserer citeringsdata, opbygger også en model for autoriteterne i feltet.

Tekstminingsmetoder er generelle og ikke specielle for hjernevidenskab. For at få maskiner til at opbygge en dybere forståelse af hjernen, er det nødvendigt at bygge computer-metoder, der specifikt retter sig mod hjernedata.

## **Hjernemining: Maskinforståelse af hjernedata**

En af de første databaser, der var rettet specifikt mod hjernevidenskabelige data, var BrainMap fra starten af 1990'erne. I 2005 er de nået op på at have indtastet information fra over 500 videnskabelige artikler. Udover tekst fra artiklerne indeholder den hjerneaktiveringer: Aktiveringerne fremkommer ved en speciel type eksperimenter, hvor en gruppe forsøgspersoner skal udføre en opgave, der får deres hjerner til at arbejde: Det kan være at mærke kuldesmerte, lytte til sætninger der fremkalder erindringer, eller se på billeder af ansigter der udtrykker frygt eller glæde. Sådanne opgaver vil ofte hver især aktivere specifikke hjerneområder, og disse kan opspores med hjerneskanere. Billedanalyse af hjerneskanerne fører til en liste af aktiverede områder. De

bliver rapporteret i et standardiseret referencesystem, oprindeligt udviklet til hjernekirurgi, så den anatomiske position af en persons aktiveringer kan sammenlignes med en anden persons aktiveringer. Resultaterne er dog ikke altid lige overbevisende, da forsøgspersonernes tankeaktivitet ikke kan styres præcist, og hjerneskanterne kun måler tankeaktiviteten indirekte. Oftest er det først når adskillige studier viser det samme, at man begynder at tale om noget der nærmer sig "viden". Det er derfor nødvendigt at sammenholde studier, og oftest foregår det ved at forskeren læser artikler og integrerer informationen "i hovedet". Ved at lægge hjerneaktiveringerne ind i en tilpas struktureret database, er det muligt for en computer at udtrække konsensus automatisk. Det er netop, hvad der er muligt med data fra BrainMap. Har computeren opbygget en tilstrækkelig god model for konsensus, er det også muligt automatisk at vurdere i hvor høj grad et nyt studie skiller sig ud: Om der for eksempel er en nyhedsværdi, eller om det blot føjer sig til den gængse viden.

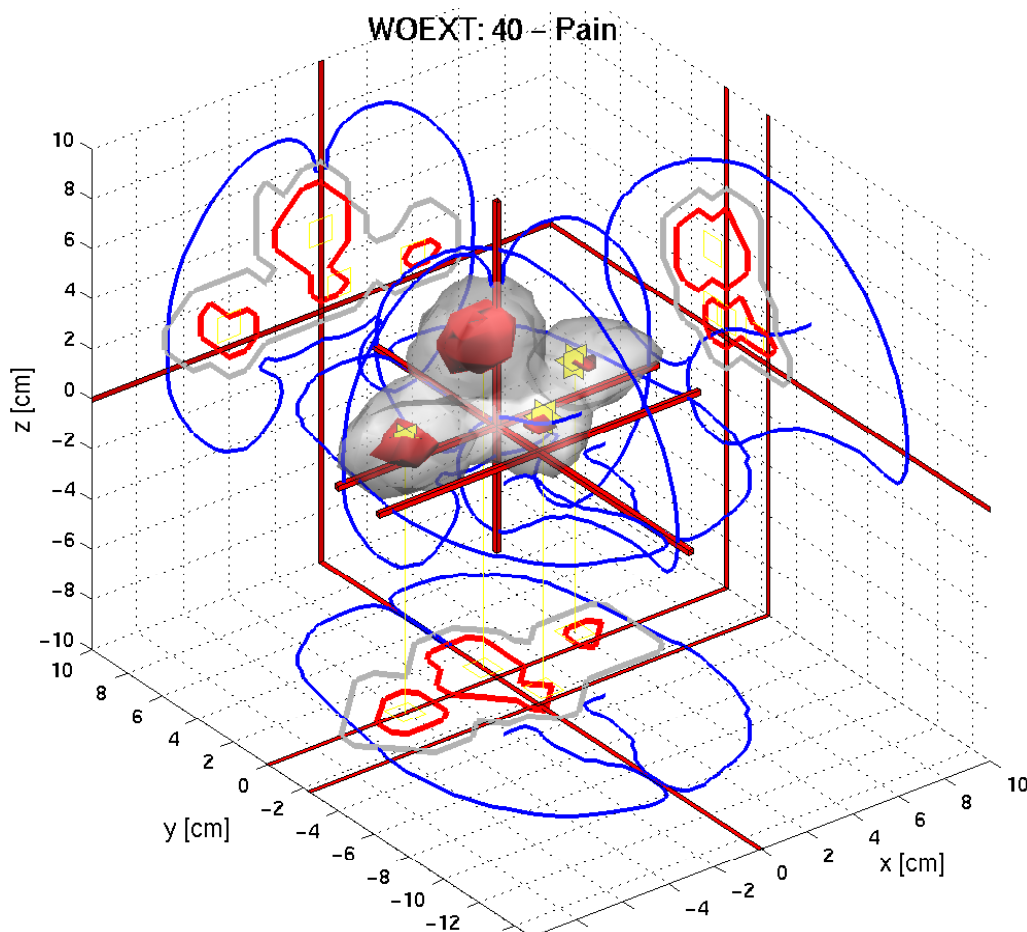
En ekstra støtte for de computerbaserede metoder i forståelsen af hjerne-data er ontologier. I computer-sammenhæng er en ontologi en database af begreber og sammenhænge mellem begreberne. Et eksempel er BrainInfo, der hierarkisk beskriver hjernestrukturer samtidig med at den katalogiserer variationen i benævnelserne. Med dette ontologi er det for eksempel muligt at fortælle computeren at hjernestrukturen "caudate" er en del af "basal ganglia" og at den også benævnes "nucleus caudatus". Ontologier hjælper med at sætte begrænsninger på antallet af mulige fortolkninger, som computeren ville skulle gennemse for at forstå data.

Kun visse funktioner er tilgængelige med BrainMap. En dansk pendant til BrainMap er Brede databasen, og den er mere fleksibel. Den indeholder ontologier for hjernestrukturer og hjernefunktioner, og ved at kombinere dem med tekst og hjerneaktiveringsdata, kan man *datamine* sig frem til interessant information og vurdere de enkelte informations relevans i forhold til hinanden. Den automatiske måde, hvorpå den interessante information fremdrages, benytter metoder der ligner dem fra tekstmining. Blot er inputtet til algoritmen ikke kun tekst men også hjerneaktiveringsdata og ontologier.

I første omgang vil neuroinformatikken lette adgangen til og skabe nye måder at tilgå information på. Både BrainMap, BrainInfo og Brede repræsenterer deres information grafisk. Alle tre databaser er tilgængelige via Internettet, og alle tre tillader at man ikke blot kan søge på ord, men også anvende billedorienterede søgninger: I Brede kan man for eksempel indtaste en position i hjernen, givet ved koordinater, og derefter få returneret de hjerneaktiveringer der ligger nærmest. Men der er stadig et godt stykke igen til en egentlig maskineforståelse. For eksempel har algoritmen i de fleste typer tekstmining ingen forståelse for sætningsopbygningen, og behandler

blot teksten som “en sæk ord”. Det er endnu ikke helt klart, hvilke hjernevidenskabelige data det er relevant at lægge i databaser, og hvordan man sammenkobler de forskellige databaser bedst muligt.

Der er et interessant aspekt ved neuroinformatik: Hjernen er en generel maskine der kan bruges til at forstå hjernen med. Indenfor neuroinformatikken forsøger man at opbygge en maskine der kan forstå hjernen. Hvis maskinen forstår hjernen, vil den i en eller anden grad indeholde elementer som også karakteriserer den rigtige hjerne.



Figur 1: En visualisering af områder i den menneskelige hjerne der er associeret med smerteopfattelsen. De røde områder angiver de steder, der er specielt involveret, og visualisering er automatisk genereret ud fra informationen i Brede databasen.

## Henvisninger

- Jackson Beatty, *What's New About Neuroinformatics?*  
<http://www.ni.ucla.edu/library/whatsnew/>  
En introducerende artikel om neuroinformatik.
- The Global Science Forum Neuroinformatics Working Group of the Organisation for Economic Co-operation and Development, *Report on Neuroinformatics*.  
<http://www.oecd.org/pdf/M00033000/M00033112.pdf>  
En rapport om neuroinformatik fra en arbejdsgruppe under OECD.
- Google Scholar. <http://scholar.google.com/>  
En søgemaskine for videnskabelige artikler — en søster til den almindelige og populære Google web-søgemaskine.
- BrainMap databasen. <http://www.brainmap.org/>  
Denne database kræver at man downloader et java-program.
- BrainInfo (NeuroNames). <http://braininfo.rprc.washington.edu/>
- Brede databasen. <http://hendrix.imm.dtu.dk/services/jerne/brede/>
- Der er mange andre databaser, for eksempel fMRIDC, en database med rå hjerneskan fra magnetisk resonans skannere, og CoCoMac, som indeholder information om anatomiske forbindelser mellem hjerneområder.  
<http://www.fmridc.org/> og <http://cocomac.org/>

*Tak til Daniela Balslev og Claus Svarer for kommentarer.*