

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/57983> holds various files of this Leiden University dissertation.

Author: Cai, F.

Title: Fuzzy systems and unsupervised computing: exploration of applications in biology

Issue Date: 2017/12/12

Samenvatting (Dutch Summary)

In dit proefschrift wordt de toepassing van het fuzzy systeem onderzocht in het bijzonder in het domain van bioinformatica. In de afgelopen tientallen jaren is de belangstelling voor bioinformatica bij zowel biologen als computerwetenschappers fors toegenomen. Dit geldt voor biologen omdat zij in hun onderzoek grote hoeveelheden data produceren die op een robuuste en accurate manier moeten worden geanalyseerd. Dit geldt voor computerwetenschappers omdat zij ingaan op de uitdaging toegevoegde waarde voor deze grote data volumes te creëren door het ontwikkelen en ontwerpen van methodologieën waarmee zulke accurate en robuuste analyses kunnen worden gerealiseerd.

Data uit het onderzoek in de levenswetenschappen zijn in oorspronkelijke ruwe vorm, notoir vaag, dubbelzinnig, onnauwkeurig, en soms missen er data of zijn er delen niet bekend. Om met het begrip dat data niet perfect zijn om te kunnen gaan, is het nodig dat er verschillende nieuwe benaderingen worden bestudeerd en geïmplementeerd. Het Fuzzy Systeem – gebaseerd op zogenaamde “vage” logica – is een relatief nieuwe heuristische benadering die de geschikt is voor het simplificeren van een anders complexe beslissing, door het toestaan van meerdere hypothesen in de data analyse. Met andere woorden, de logica van het fuzzy systeem erkent het feit dat significantie de belangrijkste factor in het modelleren van data is; dit terwijl in andere benaderingen aan precisie meer waarde wordt toegekend.

De fuzzy systeem benadering wordt geacht potentie te hebben voor data analyse, vandaar dat het onderzoek beschreven in dit proefschrift de intentie heeft de nadruk te leggen op het ontwerpen van efficiënte en betrouwbare heuristische oplossingen voor analyse om daarmee verborgen informatie in biologische experimenten te ontdekken. Ons onderzoek heeft als doel een speciale analysemethodiek te bouwen gebaseerd op fuzzy systemen; in de onderzoekshoofdstukken van dit proefschrift worden daartoe drie verschillende perspectieven uitgewerkt. Uiteindelijk integreren we deze drie verschillende benaderingen van fuzzy systemen in een analysemethodiek die wordt geïllustreerd aan de hand van een voorbeeld studie.

(1) We hebben fuzzy systemen gebruikt voor de correctie van achtergrondbelichting. Een microscoopbeeld bevat ruis en andere effecten die niet bijdragen aan het daadwerkelijke signaal dat wordt gemeten. Behalve willekeurige ruis die wordt veroorzaakt door de electronica van het apparaat, wordt het beeld beïnvloed door ongelijke belichting. Om de invloed van deze ongelijke belichting te reduceren moet er een achtergrondcorrectie worden uitgevoerd.

In Hoofdstuk 2 van dit proefschrift wordt een achtergrondcorrectie methode voorgesteld die gebaseerd is op mathematische morfologie en deze wordt

samengevoegd met randvoorwaarden uit de fuzzy logica, waarmee het achtergrondverloop wordt geëlimineerd. Vergeleken met de algemeen gebruikte EMI methode en het, in commerciële software veel gebruikte, “rolling ball” algoritme, demonstreert de door ons ontwikkelde DCBC methode robuuste prestaties met beelden die karakteristiek zijn voor biomedische microscopie.

(2) We hebben fuzzy systemen bestudeerd in relatie tot kenmerkselectie (eng. feature selection) en verwijderen van overvloedige data. De procedure van kenmerkselectie speelt een belangrijke rol in het omzetten van fenotypische waarnemingen naar een matrixvorm voor statistische berekeningen. Hierbij dienen twee aspecten onder ogen gezien te worden: (i) selectie van de kleinste hoeveelheid kenmerken waarmee de dataset het best kan worden beschreven en waaruit voorspellingen kunnen worden afgeleid; (ii) reductie van de dimensionaliteit van de dataset door het verwijderen van overbodige informatie.

In Hoofdstuk 3 van dit proefschrift wordt een filter-gestuurde aanpak voor kenmerkselectie aangedragen, waarin criteria uit de fuzzy logica worden gebruikt als één van de kostenfuncties. Daarbij wordt een zogenaamd multi-objectief evolutionair algoritme gebruikt voor het produceren van optimale oplossingen langs de Pareto grens. Deze aanpak verschilt van andere strategieën, daar de door ons werk geïntroduceerde methode voorziet in een set van kandidaat kenmerk combinaties. Met deze combinaties kan de besluitvormer zijn voordeel doen door de voor het specifieke geval meest waardevolle combinatie te kiezen.

(3) We hebben het gebruik van fuzzy systemen verkend voor het clusteren van informatie. Onder de patroonextractie methoden, i.e. samenvatting, associatie en predictie, is het clusteren van informatie van groot belang; dat geldt voor onderzoek alsook voor de dagelijkse praktijk. Dit is met name het geval voor een dataset die weinig of geen labels heeft.

In Hoofdstuk 4 van dit proefschrift hebben we een nieuwe clustering methode verwezenlijkt waarbij de zogenaamde fuzzy rough c-means benadering en het deeltjes zwerm optimalizatie (PSO) algoritme worden gecombineerd. Deze combinatie integreert tot een zinnig globaal resultaat. Het concept van fuzzy logica kan omgaan met onzekerheid, vaagheid en partities die in de dataset overlappen, terwijl het PSO algoritme helpt met het zoeken naar vrijwel optimale oplossingen.

(4) We hebben fuzzy systemen onderzocht als onderdeel van een speciale analysemethodiek, dit is een samenvoeging van de benaderingen die in de vorige hoofdstukken behandeld zijn. De toepasbaarheid wordt gedemonstreerd aan de hand van een casus.

In Hoofdstuk 5 van dit proefschrift beschrijven we een case studie die gericht is op een overkoepelende data analyse van een 1-dimensionale gel-electroforese experiment. De voorgestelde analysemethodiek gebaseerd op fuzzy systemen heeft laten zien in staat te zijn tot nauwkeurige extractie van kenmerken uit beelden van eiwitgels die karakteristiek voor kankercellen. De beelden van de eiwitgels zijn verre van ideaal, vandaar dat een aantal correcties moeten worden toegepast zodat er juiste metingen uit deze beelden kunnen worden verkregen. Deze metingen kunnen worden gebruikt om de significante verschillen tussen kankercel-lijnen, en de daarbij horende groepen en subgroepen te onderscheiden en te karakteriseren. Dit draagt bij aan het verdere begrip met betrekking tot de ontwikkeling van kanker. Daarenboven kan de analysemethodiek worden gebruikt voor het verkrijgen van begrip in eiwit/DNA patronen en hun karakteristieken zoals die in gelbeelden worden gezien.

Concluderend, het onderzoek beschreven in dit proefschrift heeft als doel het onderzoeken van fuzzy systemen in combinatie met patroonherkenning in het onderzoeksveld van de bioinformatica. Uit het bestuderen van bestaande methodologieën is een ensemble van fuzzy logica en unsupervised algoritmen ontworpen en geïntegreerd in een analysemethodiek waarmee vragen met een oriëntatie in de biologie en in patroonherkenning kunnen worden begrepen en aangepakt.