

Verslag krediet aan navorsers

1 Wetenschappelijke resultaten

1.1 Aanwending van de middelen

Het krediet aan navorsers besloeg de laatste twee jaar (2004-2006) van mijn eerste mandaat als postdoctoraal onderzoeker FWO en het eerste jaar (2006-2007) van mijn verlenging als postdoctoraal onderzoeker FWO. Het project sloot nauw aan bij het onderzoeksthema van deze mandaten als postdoctoraal onderzoeker.

Meer bepaald leverde het krediet aan navorsers werkingsmiddelen voor mijn onderzoek rond het efficiënter en bruikbaar maken van ILP algoritmen (vooral voor bio-informatica-toepassingen). Het krediet werd voornamelijk aangewend voor

- aankoop van literatuur
- aankoop hardware en software nodig voor het uitvoeren van het onderzoek
- kosten verbonden aan het bijwonen van conferenties waar onderzoeksresultaten voortvloeiend uit het project gepresenteerd werden.

Hierna volgt een overzicht van de belangrijkste onderzoeksresultaten.

1.2 Nieuwe data mining methoden

1.2.1 Efficiënte algorithmes voor graph mining

De huidige technieken in het gebied van Inductief Logisch programmeren en Graph Mining veronderstellen meestal data voorgesteld met (algemene) grafen, waardoor de computationele kost zowel in theorie (worst case) als in de praktijk exponentieel stijgt met de grootte van de grafen. In de praktijk echter, kan men vaststellen dat veel data kan voorgesteld worden met gemakkelijker klassen grafen. bv. veel moleculen kunnen met planaire grafen voorgesteld worden, en elk atoom heeft hoogstens 6 bindingen. In samenwerking met Tamas Horvath (Bonn) werd een algoritme ontwikkeld om efficiënt patronen te ontdekken in verzamelingen outerplanaire grafen [6]. Gelijkaardige technieken kunnen ook toegepast worden voor het efficiënt berekenen van afstanden tussen outerplanaire grafen [12].

Specifiek voor het genereren van (niet-isomorfe) patronen (dus niet voor het berekenen van de frequentie in een databank) werd een algoritme ontwikkeld dat onder ruime voorwaarden (bv. voor de klasse van alle grafen) slecht een polynomiale tijd nodig heeft per gegenereerd patroon [11].

1.2.2 Actief leren

Er werd tijdens een studieverblijf aan de computational biology groep van de University of Wales (Aberystwyth, UK) gewerkt rond actief leren. Hier willen we vooral nagaan hoe lerende algoritmen kunnen omgaan met reële wereld-experimenten (onzekerheid, real-time) en hoe ze door zelf experimenten te kiezen het leerproces kunnen verbeteren. De juiste (meest informatieve) experimenten kiezen kan zeer kostenbesparend zijn in (duur) bio-medisch onderzoek.

De toepassing op coherente lasercontrole waaraan in Aberystwyth gewerkt werd is uitdagend door de hoge dimensionaliteit van de ruimte van mogelijke experimenten. Een aantal eerste resultaten werden beschreven in [4].

1.2.3 Incrementele leeralgorithmes

Niet alle kennis blijft onveranderd geldig. Omstandigheden kunnen veranderen en soms is het nodig geleerde patronen aan te passen. Recentelijk is er in het domein van reinforcement learning veel aandacht voor “transfer learning”. Hiermee bedoelt men dat men eerst een taak leert oplossen en daarna probeert de opgedane kennis te hergebruiken voor het oplossen van een andere taak. Niet alle kennis is nog geldig, maar een aantal elementen blijven wel toepasselijk.

Een vergelijkbaar probleem doet zich voor bij online leren, waar men niet alle informatie meteen krijgt, maar de informatie slechts geleidelijk aan binnenkomt. In onze intensieve zorgen toepassing bijvoorbeeld heeft men bij opname van een patient weinig informatie, maar moet men toch al beslissingen nemen. Naarmate de patient langer op de afdeling blijft zal de kennis over de patient toenemen. Mogelijks moet men daarom in de loop van het verblijf bepaalde hypothesen bijstellen. Aan de andere kant bestaat de hoop dat een nieuwe patient op een redelijk gelijkaardige manier zal reageren als de vorige, zodat men vroeger opgedane kennis (gedeeltelijk) kan hergebruiken.

In [1] beschrijven we een transfer learning aanpak waarbij we abstracties gebruiken om te veralgemenen. In [2] beschrijven we een incrementeel leersysteem gebaseerd op relationele kernel regressie. In [9] beschrijven we een benadering om incrementeel beslissingsbomen te leren en, als fouten worden ontdekt, te reviseren. Het artikel beschrijft een aantal experimenten in het domein van reinforcement learning.

1.2.4 Probabilistische relationele modellen en leertechnieken

Er werd ook gewerkt aan het ontwikkelen van methoden voor Bayesiaanse netwerken op relationele data. Er werden zowel voorstellen gedaan voor het leren van lokale modellen als voor het leren van de structuur van dergelijke Bayesiaanse netwerken.

[3] is een vergelijkende studie van verschillende methoden voor het leren van relationele probabiliteitsbomen. Deze bomen laten toe om functies voor

te stellen van relationele data naar probabiliteiten. Om dergelijke bomen optimaal te leren bleek dat men beter andere heuristische gebruikt dan voor bv. regressiebomen. Een meer uitgebreid tijdschriftartikel is ingediend. De relationele probabiliteitsbomen werden ondertussen in meerdere applicaties (o.a. HIV) toegepast.

[8] beschrijft een veralgemening van ordering search voor relationele data. Deze methode laat toe om de structuur van Logische Bayesiaanse Netwerken (LBN) en gelijkaardige modellen te leren in de vorm van ordeningsfuncties die aangeven in welke volgorde de knopen van het netwerk moeten beschouwd worden. Als lokale modellen gebruiken we relationele probabiliteitsbomen.

1.3 Applicaties

1.3.1 Intensieve geneeskunde

Een samenwerking met de eenheid intensieve zorgen van het UZ-Leuven werd opgestart. Resultaten op manueel ingevoerde data met dagelijkse patient-gegevens werden o.a. gepubliceerd in [7]. Sinds midden 2006 is het online monitoring systeem van de afdeling voldoende operationeel en uitgetest om betrouwbare hoge-resolutie data te leveren. Deze data bevat o.a. gegevens over de waarden van meerdere parameters van de patienten die elke minuut worden gemeten. In eerste instantie is er gewerkt aan het voorspellen van de evolutie van deze parameters met Gaussiaanse processen [5]. Voor korte-termijnvoorspellingen blijkt deze aanpak succesvol, maar voor langere-termijnvoorspellingen neemt de nauwkeurigheid sterk af. Meer recent wordt daarom gewerkt aan twee ideeën die dit probleem mogelijks kunnen oplossen. De eerste is het beter parameteriseren van de dynamische eigenschappen van de data (in samenwerking met de Biores groep van de faculteit Bioingenieurswetenschappen). De tweede is een benadering met Dynamische Bayesiaanse netwerken waarbij we het probleem op een hoger niveau aanpakken.

1.3.2 HIV onderzoek

De DTAI onderzoeksgroep werkt reeds geruime tijd samen met het Rega-instituut van de K.U.Leuven rond HIV-onderzoek. In dit project werd specifiek gewerkt aan een algoritme [10] dat gegeven een aantal sequenties van virussen van verschillende patienten, enerzijds een phylogenetische boom bouwt die zo goed mogelijk alle polymorphismen verklaart en anderzijds een functie leert die voor een gegeven sequentie zegt hoe waarschijnlijk elke mogelijke volgende mutatie is.

2 Samenwerkingsverbanden

Tijdens het project verbleef ik 6 weken in Aberystwyth, UK voor samenwerking met de computational biology groep rond coherente lasercontrole en 5 weken in Bonn, Duitsland voor samenwerking met Tamas Horvath rond efficiënte algoritmen voor graph mining.

Het onderzoek uit dit project droeg in belangrijke mate bij tot de resultaten behaald in de samenwerkingsprojecten met de afdeling intensieve geneeskunde (K.U.Leuven) en het Rega-instituut (K.U.Leuven) ivm. HIV-onderzoek.

3 Eventuele initiatieven tot ruimere bekendmaking of valorisatie

De samenwerkingsverbanden met andere onderzoeksgroepen rond intensieve geneeskunde en virologie kan gezien worden als een eerste stap naar valorisatie van sommige van de ontwikkelde technieken. In het bijzonder zijn een aantal prospectieve studies voorbereid in de afdeling intensieve geneeskunde die hopelijk de waarde van onze methoden kunnen aantonen (o.a. bij planning van niet-urgente ingrepen door predictie van ontslag van behandelde patienten).

4 Publicaties

Zie bijgevoegde publicatielijst.

References

- [1] K. Driessens, J. Ramon, and T. Croonenborghs. Transfer learning for reinforcement learning through goal and policy parametrization. In *ICML Workshop on Structural Knowledge Transfer for Machine Learning (Online Proceedings)*, 2006.
- [2] K. Driessens, J. Ramon, and T. Gärtner. Graph kernels and Gaussian processes for relational reinforcement learning. *Machine Learning*, 64(1-3):91-119, 2006.
- [3] Daan Fierens, Jan Ramon, Hendrik Blockeel, and Maurice Bruynooghe. A comparison of approaches for learning probability trees. In *Proceedings of 16th European Conference on Machine Learning*, volume 3720 of *Lecture Notes in Artificial Intelligence*, pages 556-563, 2005.
- [4] N. Form, R. Burbidge, J. Ramon, and J. Whitaker. Parameterisation of an acousto-optic programmable dispersive filter for closed-loop learning experiments. *Journal of Modern Optics*, 99999(1):1-13, January 2007.

- [5] F. Guiza, J. Ramon, and H. Blockeel. Gaussian processes for prediction in intensive care. In Neil D. Lawrence, Anton Schwaighofer, and Joaquin Quinonero, editors, *Proceedings of the Gaussian Processes in Practice Workshop*, Bletchley Park, U.K., June 2006.
- [6] Tamás Horváth, Jan Ramon, and Stefan Wrobel. Frequent subgraph mining in outerplanar graphs. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 197–206, Philadelphia, PA, August 2006.
- [7] J. Ramon, D. Fierens, F. Guiza, G. Meyfroidt, H. Blockeel, M. Bruynooghe, and G. Van Den Berghe. Mining data from intensive care patients. *Advanced Engineering Informatics*, 21(3):243–256, 2007.
- [8] Jan Ramon, Tom Croonenborghs, Daan Fierens, Hendrik Blockeel, and Maurice Bruynooghe. Generalized ordering-search for learning directed probabilistic logical models. *Machine Learning*, 2007. to appear.
- [9] Jan Ramon, Kurt Driessens, and Tom Croonenborghs. Transfer learning in reinforcement learning problems through partial policy recycling. In *Proceedings of the 18th European Conference on Machine Learning*, Lecture Notes in Artificial Intelligence. Springer-Verlag, 2007.
- [10] Jan Ramon, Snezhana Dubrovskaya, and Hendrik Blockeel. Learning resistance mutation pathways of HIV. In *Proceedings of The Sixteenth Annual Machine Learning Conference of Belgium and the Netherlands, Amsterdam, The Netherlands*, 2007. URL: <http://www.cs.kuleuven.ac.be/cgi-bin/dtai/publ.info.pl?id=42687>.
- [11] Jan Ramon and Siegfried Nijssen. General graph refinement with polynomial delay. In *Proceedings of the Workshop on Machine Learning and Graphs (MLG'07)*, 2007.
- [12] L. Schietgat, J. Ramon, and M. Bruynooghe. A polynomial-time metric for outerplanar graphs. In *Benelearn 2007: Proceedings of the Annual Machine Learning Conference of Belgium and the Netherlands*, pages 97–104, 2007.