

Inference of biological networks by data mining

Michael Hörnquist, ITN

Viktigaste vetenskapliga resultat

Projektet har på kort sikt syftat till att utveckla algoritmer för att utvinna information om biologiska nätverk ur storskaliga molekylärbiologiska experiment. På längre sikt, bortanför projektperioden, är målet att utveckla en systemförståelse för biologiska nätverk.

Viktigast vetenskapliga resultat är att vi påvisat att det verkligen är möjligt att alls genomföra en nätverksbaserad informationsutvinning med biologisk relevans ur storskaliga molekylärbiologiska experiment. Detta har vi till dels visat genom en serie publikationer, se publikationslistan, men framför allt genom att delta i den internationella tävlingen DREAM – Dialogue on Reverse Engineering and Assessment Methods – och vinna den i utvalda grenar både 2007 och 2008 (2009 saknade vi möjlighet att ställa upp). År 2007 handlade det om att återskapa ett in-silico nätverk där data från tidsserier och från knock-out experiment presenterades (<http://wiki.c2b2.columbia.edu/dream09/index.php/D2c4>), medan år 2008 var mer biologiskt orienterat och syftade till att prediktera gennivåer från faktiska mätningar av jäst (<http://wiki.c2b2.columbia.edu/dream/index.php/D3c3>).

Vägen till framgång på den internationella scenen har kantats av många mindre resultat, vilka dock uppmärksammats positivt av forskarsamhället. Vi lyfter några av dessa här. Vi har påvisat möjligheten att göra en nätverksinferens där antalet noder är ca hundra gånger större än antalet mätpunkter (artikel 2, med referens till publikationslistan nedan); vi har visat hur befintliga klustringstekniker har mycket att ge när man söker efter grannskap i nätverk (artikel 4); hur ett mycket enkelt mått kan prediktera om ett booleskt system är ordnat eller kaotiskt (artikel 6); hur man kan kvantifiera begreppen robusthet och flexibilitet till att vara avpassade för lineariserade system med stor osäkerhet i individuella kanter (artikel 8); etc.

Examina, examensarbeten och befordringar

- Michael Hörnquist, docent, 2004
- Fredrik Karlsson, civilingenjörsexamen ED, 2005, "Dynamics in Boolean networks"
- Mika Gustafsson, fil. lic., 2006, "Large-scale topology, stability and biology of gene networks"
- Mika Gustafsson, fil. dr, 2010, "Gene networks from high-throughput data – reverse engineering and analysis" (planerad disputation 26 mars)

Personer finansierade av projektet

Större delen av Mika Gustafssons doktorandtjänst och viss forskningstid (forskarhandledningstid) för Michael Hörnquist har finansierats av CENIIT via detta projekt.

Industrikontakter

I början av projektet hade vi kontakter med AstraZeneca i Mölndal. Vi besökte dem vid några tillfällen och gav seminarier om våra forskningsresultat, varvid de även gavs möjlighet att återkoppla kring vilka frågeställningar de såg som mest relevanta ur sitt perspektiv.

Företaget bidrog även med data vilka vi kunde använda för att testa de algoritmer vi utvecklade. Efter en omorganisation i Mölndal kom vi dock uppenbart mellan två stolar, och trots upprepade påstötningar från vår sida upphörde kontakten.

Därefter har vi försökt finna nya industripartners, bland annat företaget Innetics, men där våra intressen varit alltför åtskilda för att leda till något fruktbart.

Kontakter med andra CENIIT-projekt

Närmaste CENIIT-projekten parallella med det egna torde ha varit de som drevs och drivs av Patrick Lambrix respektive Lena Strömbäck. Här påbörjade vi ett samarbete kring att utforska och sammanställa relevanta fria databaser för att utröna deras överlapp och hur man genom att integrera dem skulle kunna få en mer heltäckande bild av den organism som studeras. Av olika skäl blev dock projektet fördröjt, och då databaser är färskvara kom arbetet att bli inaktuellt innan det hann publiceras.

Vissa beröringspunkter har även funnits med Peter Jonssons projekt, men där har samverkan inskränkt sig till att undertecknad medverkat i betygskommittén för en doktorand.

Formande av forskargrupp och externa kontakter

Under åren har forskargruppen inom beräkningssystembiologi på ITN bestått av doktoranden Mika Gustafsson, undertecknad Michael Hörnquist, samt lektorn Anna Lombardi i den mån hon haft finansiering för forskning. För att ha en god forskningsmiljö har vi haft tät kontakt med gruppen kring Jesper Tegnér, först på IFM och senare på Karolinska Institutet. Trots goda utvärderingar från Vetenskapsrådet har ansökningarna dit inte gett några forskningsmedel, vilket omöjliggjort expansion av forskargruppen med nuvarande tillämpade grundforskningsinriktning. Undertecknads övriga uppdrag har försvårat att ändra inriktning till mer tillämpad forskning för att på så sätt kunna attrahera medel från andra finansierare såsom Cancerfonden, Vinnova etc. Försök att attrahera examensarbetare för att göra inbrytningar och förstudier inom andra områden har också varit mindre lyckosamma, då relevanta utbildningar på avancerad nivå ligger i Linköping och det fysiska avståndet har gjort att presumtiva studenter till sist valt något som legat dem närmare geografiskt.

I skrivande stund pågår samverkan kring en förstudie för åderförkalkning ihop med KI och RIKEN, och en förstudie kring livmodercancer ihop med HiS. Beroende på utfallet av dessa studier hoppas vi kunna inkomma med mer tillämpningsinriktade ansökningar för att på den vägen fortsätta försöka etablera en forskargrupp av större omfattning än den hittillsvarande.

Publikationer inom projektet

Artiklar i peer-review tidskrifter

1. K. Astrup Eriksen, M. Hörnquist and K. Sneppen, *Functional and Integrative Genomics* 4, 241-245 (2004) "Visualization of large-scale correlations in gene expressions".
2. M. Gustafsson, M. Hörnquist and A. Lombardi, *IEEE/ACM --- Transactions on Computational Biology and Bioinformatics* 2 (2), 254-261 (2005) "Constructing and analyzing a large-scale gene-to-gene regulatory network - Lasso-constrained inference and biological validation".
3. V. Thorsson, M. Hörnquist, A.F. Siegel and L. Hood, *Statistical Applications in Genomics and Molecular Biology* 4 (1), article 28 (24 pages) (2005) "Reverse Engineering Galactose Regulation in Yeast through Model Selection".
4. M. Gustafsson, M. Hörnquist and A. Lombardi, *Physica A* 367C, 559--576 (2006), "Comparison and validation of community structures in complex networks".
5. A. Lombardi and M. Hörnquist, *Phys. Rev. E* 75, 056110 (2007), "Controllability analysis of networks".
6. F. Karlsson and M. Hörnquist, *Physica A* 384, 747-757 (2007), "Order or chaos in Boolean gene networks depends on the mean fraction of canalizing functions".
7. M. Gustafsson, M. Hörnquist, J. Lundström, J. Björkegren and J. Tegnér, *Ann. N.Y. Acad. Sci.* 1158, 265-275 (2009), "Reverse Engineering of Gene Networks with LASSO and Non-Linear Basis Functions".
8. M. Gustafsson, M. Hörnquist, J. Björkegren and J. Tegnér, *IET-Systems Biology* 3, iss. 4, 219-238 (2009) "Genome-Wide System Analysis Reveals Stable yet Flexible Network Dynamics".
9. H. Suzuki, ... , M. Hörnquist, ... and Y. Hayashizaki (totally 158 names), *Nature Genetics* 41, iss. 5, 553-562 (2009), "The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line".
10. M. Gustafsson and M. Hörnquist, accepted for *PLoS ONE*, to appear 2010 "Gene Expression Prediction by Soft Integration and the Elastic Net"

Bokkapitel

11. M. Gustafsson and M. Hörnquist: *Handbook of Research on Computational Methods in Gene Regulatory Networks*, eds. S. Das, D. Caragea, S. Welch and W. Hsu., pp. 476-496, Hershey, New York (2009), "Integrating various data sources for improved quality in reverse engineering of gene regulatory networks"

Konferensbidrag

12. M. Hörnquist and V. Thorsson, "Reverse engineering as a subset selection problem" (International Conference on Systems Biology, ICSB 2003, St Louis, USA, pp. 141-142)

13. M. Gustafsson, M. Hörnquist and A. Lombardi, "Large scale reverse engineering by the Lasso" (International Conference on Systems Biology, ICSB 2003, St Louis, USA, pp. 135-136)
14. M. Gustafsson, M. Hörnquist and A. Lombardi, "Combinations of Gene Clusterings Enhance Biological Validity" (Reglermöte 2004, Göteborg, 6 pages)
15. M. Gustafsson, M. Hörnquist and A. Lombardi, "Community structure in gene-to-gene networks" (Bioinformatics 2004, Linköping, ed. by B. Persson et al., p. 41)
16. M. Gustafsson, M. Hörnquist, A. Lombardi and J. Tegnér, "Gene network dynamics in yeast - communities, motifs and hubs" (6th Swedish Bioinformatics Workshop, Göteborg, 2005, p. 21)
17. M. Gustafsson, M. Hörnquist, J. Björkegren and J. Tegnér, "Out-degrees, modules and motifs in a gene-to-gene regulatory network" (DIMACS Workshop on Bimolecular Networks, New Jersey, 2005, on-line)
18. M. Gustafsson, M. Hörnquist, J. Björkegren and J. Tegnér, "Genome-Wide System Identification and Analysis Reveals Stable yet Flexible Network Dynamics in Yeast", (International Conference on Systems Biology, ICSB 2006, Yokohama, Japan, 2 pages on CD)
19. M. Gustafsson, M. Hörnquist, J. Björkegren and J. Tegnér, "Soft Integration of Data for Reverse Engineering", (International Conference on Systems Biology, ICSB 2008, Göteborg, p. 127)
20. M. Gustafsson and M. Hörnquist, "Gene expression prediction by the elastic net" (DREAM 2008, Boston, p. 48)
21. M. Gustafsson and M. Hörnquist, "In-silico network predictions by ODE and lasso" (DREAM 2008, Boston, p. 138)

Avhandlingar

22. M. Gustafsson, "Large-scale topology, stability and biology of gene networks", Linköping Studies in Science and Technology, Thesis. No. 1256.
23. M. Gustafsson, "Gene networks from high-throughput data – reverse engineering and analysis", Linköping Studies in Science and Technology, Dissertations. No. 1301.

Norrköping, 27 januari 2010,
Michael Hörnquist