

Rekonstruktion av ofullständigt specifierade objekt

Slutrapport

Peter Jonsson

Projektet och dess resultat

Vi har studerat ett generellt problem med tillämpningar inom i första hand bioinformatik men också inom en rad andra områden som databassäkerhet, schemaläggning och nätverksplanering. Konkreta bioinformatiska tillämpningar som studerats inkluderar olika varianter av proteinsekvensiering, studiet av 3D-struktur hos proteiner samt automatiserad utvärdering av data från masspektrometriska experiment. Inom biokemisk industri betraktas dessa problem som mycket relevanta men också som mycket svårlösta. Ett konkret exempel är följande: Givet olika klasser av receptorer, dvs molekyler inom eller på ytan av en cell vid vilka en substans (tex ett hormon eller läkemedelssubstans) selektivt binder till, vilket orsakar en aktivitetsförändring i cellen, avgöra vilka molekylära aspekter, tex sekvensstruktur, som bidrar till olika bindingsgrad, dvs aktivitetsförändring. Bioinformatik kan i denna kontext bidra med att generera testbara hypoteser, vilka beskriver vilka egenskaper, deras distribution, och relationer mellan olika egenskaper som skulle kunna bidra till en uppmätt aktivitet.

En abstrakt beskrivning av vårt problem ser ut på följande sätt: Vi ställs inför ett objekt O som vi har ofullständig information om, och vi vill veta om den informationen räcker för att skapa en fullständig beskrivning av hela O eller delar av O . Problem av denna typ kan elegant omformuleras som *the frozen variable problem* (FVP) där informationen om O modelleras som relationer mellan variabler vilka beskriver olika aspekter av objektet O . Studiet av FVP och liknande problem har revolutionerats de senaste åren genom upptäckten av nya algebraiska metoder. Dessa metoder ger inte endast möjlighet att studera teoretiska egenskaper hos FVP utan ger i många fall upphov till användbara algoritmer för problemen.

Den ursprungliga konkreta frågeställningen var om en kombination av FVP och genontologier är användbar för bioinformatiska tillämpningar. En genontologi är en databas som taxonomiskt beskriver olika geners funktion med avseende på till exempel molekylär funktion, deltagande i biologiska processer och förekomst i olika celltyper. Forskning av bland andra Cliff Joslyn visar hur avancerad information kan extraheras genom användandet av enkla måttbaserade metoder och dessa metoder verkade initialt att vara möjliga att integrera med våra tekniker. Detta ledde till utveckling av metoder som visade sig vara generellt användbara inom främst databasområdet. Samtidigt drogs slutsatsen att effektivare metoder för FVP måste hittas för att kunna hantera större probleminstanser. Arbetet koncentrerades sålunda på två olika problem:

1. att konstruera snabbare algoritmer för FVP och besläktade problem; och
2. att studera approximationsmetoder för FVP och besläktade problem.

Algoritmer för problem såsom MAX CSP and MAX SOL konstruerades och gränserna för deras tillämpbarhet analyserades. *Algoritmisk tillämpbarhet* var ett nyckelord i detta sammanhang. Metodiken var att först konstruera generella algoritmer och därefter analysera vilka problem de är möjliga att applicera på. Ett konkret exempel är algoritmer för supermodulära problem: Man vet att en mycket stor del av de exakt lösbara problemen kan lösas med denna metod. Man vet också att det finns problem som inte kan lösas med denna metod men man vet mycket lite om gränslinjen. Flera frågor av denna typ klargjordes i projektet.

I senare delen av projektet breddades det mot ytterligare en bioinformatisk tillämpning: s.k. fylogenetisk rekonstruktion. Detta problem är ett av de mest studerade inom bioinformatik och det innefattar flera viktiga delproblem som trippel- och kvartettproblem, förbjudna delträdsproblemet med flera. Både algoritmiska resultat och hårdhetsresultat har visats för denna typ av problem.

Examina inom projektet

Fyra doktorer har producerats under projektets gång:

- Fredrik Kuivinen, *Algorithms and Hardness Results for Some Valued CSPs* (2009)
- Gustav Nordh, *Complexity Dichotomies for CSP-related Problems* (2007)
- Magnus Wahlström, *Algorithms, Measures, and Upper Bounds for Satisfiability and Related Problems* (2007)
- Wilhelm Dahllöf, *Exact Algorithms for Exact Satisfiability Problems* (2006)

Examensarbeten inom projektet

Fyra examensarbeten har producerats under projektets gång:

- Robert Engström "Approximability distances between circular complete graphs" (2008)
- Patrik Larsson "Analyzing and adapting graph algorithms for large persistent graphs" (2008)
- Fredrik Kuivinen "Tight approximability results for the maximum solution equation problem over abelian groups" (2005)
- Mikael Klasson "Supermodularity and the complexity of Max CSP" (2004)

Finansiering

Peter Jonsson, Magnus Wahlström och Mikael Klasson har delvis finansierats av projektet.

Samarbete

Projektet har i huvudsak samarbetat med två industriella partners:

- AstraZeneca R & D, Mölndal (M. Bjärelund, M. L. Andersson) där vi i första hand studerat applikationer inom bioinformatik. Detta samarbete har i första hand handlat om att utvärdera FVP-baserade analysmetoder t.ex. rörande aspekter av molekylär binding.
- Neo Technology (E. Eifréim). rörande tillämpningar inom databasteknik. I detta samarbete arbetade vi med grafmetoder för representation och strukturering av databaser. Detta mynnade ut i P. Larssons examensarbete som låg till grund för delar av databassystemet Neo.

Ny forskningsgrupp

Anslagen från CENIIT är uppenbarligen för små för att på egen hand kunna användas för att starta en grupp. Däremot är anslagsperiodens stora längd (6 år) mycket viktig för att utjämna skillnader i finansiering över tid. Denna utjämningsseffekt har hjälpt till vid uppbyggnaden av en forskningsgrupp som för närvarande består av tre disputerade forskare och fyra doktorander.

Publikationer

Bokkapitel

[BC2] Peter Jonsson and Gustav Nordh. Introduction to the MAXIMUM SOLUTION problem. In Nadia Creignou, Phokion Kolaitis, and Heribert Vollmer

(eds.), *Complexity of Constraints*, LNCS 5250, Springer-Verlag, pp. 255–282, 2008.

Tidskriftsartiklar

- [J34] Tomás Feder, Pavol Hell, Peter Jonsson, Andrei Krokhin, and Gustav Nordh. Retractions to pseudoforests.. *SIAM Journal on Discrete Mathematics*. To appear.
- [J33] Peter Jonsson and Gustav Nordh. Approximability of clausal constraints. *Theory of Computing Systems*. To appear.
- [J32] Peter Jonsson, Andrei Krokhin, and Fredrik Kuivinen. Hard constraint satisfaction problems have hard gaps at location 1. *Theoretical Computer Science* 410(38–40):3856–3874, 2009.
- [J31] Vladimir Deineko, Peter Jonsson, Mikael Klasson, and Andrei Krokhin. The approximability of MAX CSP with fixed-value constraints. *Journal of the ACM* 55(4), 2008.
- [J30] Peter Jonsson and Andrei Krokhin. Computational complexity of auditing finite attributes in statistical databases. *Journal of Computer and System Sciences* 74:898-909, 2008.
- [J29] Peter Jonsson, Fredrik Kuivinen, and Gustav Nordh. MAX ONES generalised to larger domains. *SIAM Journal on Computing* 38(1):329-365, 2008.
- [J28] Peter Jonsson and Andrei Krokhin. Maximum H -colourable subdigraphs and constraint optimization with arbitrary weights. *Journal of Computer and System Sciences* 73(5):691-702, 2007.
- [J27] Peter Jonsson, Mikael Klasson, and Andrei Krokhin. The approximability of three-valued MAX CSP. *SIAM Journal on Computing* 35(6):1329-1349, 2006.
- [J25] Vilhelm Dahllöf, Peter Jonsson, and Magnus Wahlström. Counting models in 2SAT and 3SAT formulae. *Theoretical Computer Science* 332(1–3):265-291, 2005.
- [J24] Peter Jonsson and Andrei Krokhin. Recognizing frozen variables in constraint satisfaction problems. *Theoretical Computer Science* 329(1–3):93-113, 2004.
- [J23] Víctor Dalmau and Peter Jonsson. The complexity of counting homomorphisms seen from the other side. *Theoretical Computer Science* 329(1–3):315-323, 2004.

Konferensbidrag

- [C39] Robert Engström, Tommy Färnqvist, Peter Jonsson, and Johan Thapper. Properties of an approximability-related parameter on circular complete graphs. In *Proceedings of the 5th Latin-american Algorithms, Graphs, and Optimization Symposium (LAGOS-2009)*, 2009. To appear.
- [C38] Peter Jonsson and Johan Thapper. Approximability of the maximum solution problem for certain families of algebras. In *Proceedings of the 4th International Computer Science Symposium in Russia (CSR-2009)*, pp. 215–226, 2009.
- [C37] Tommy Färnqvist, Peter Jonsson, and Johan Thapper. Approximability distance in the space of H -colourability problems. In *Proceedings of the 4th International Computer Science Symposium in Russia (CSR-2009)*, pp. 92–104, 2009.
- [C36] Manuel Bodirsky, Peter Jonsson, and Timo von Oertzen. Semilinear program feasibility. In *Proceedings of the 36th International Colloquium on Automata, Languages, and Programming (ICALP-2009)*, pp. 79–90, 2009.
- [C35] Tommy Färnqvist and Peter Jonsson. Bounded tree-width and CSP-related problems. In *Proceedings of the 18th International Symposium on Algorithms and Computation (ISAAC-2007)*, pp. 632–643, 2007.
- [C34] Peter Jonsson, Andrei Krokhin, and Fredrik Kuivinen. Ruling out polynomial-time approximation schemes for hard constraint satisfaction problems. In *Proceedings of the 2nd International Computer Science Symposium in Russia (CSR-2007)*, pp. 182–193, 2007.
- [C33] Peter Jonsson, Gustav Nordh, and Johan Thapper. The maximum solution problem on graphs. In *Proceedings of the 32nd International Symposium on Mathematical Foundations of Computer Science (MFCS-2007)*, pp. 228–239, 2007.
- [C32] Peter Jonsson, Fredrik Kuivinen, and Gustav Nordh. Approximability of integer programming with generalised constraints. In *Proceedings of the 12th International Conference on Principles and Practice of Constraint Programming (CP-2006)*, pp. 256–270, 2006.
- [C31] Peter Jonsson and Gustav Nordh. Generalised integer programming based on logically defined relations. In *Proceedings of the 31st International Symposium on Mathematical Foundations of Computer Science (MFCS-2006)*, pp. 549–560, 2006.
- [C30] Vladimir Deineko, Peter Jonsson, Mikael Klasson, and Andrei Krokhin. Supermodularity on chains and complexity of maximum constraint satisfaction problems. In *Proceedings of the European Conference on Combinatorics, Graph Theory and Applications (EuroCOMB-2005)*, pp. 51–56, 2005.