Bachelorarbeiten am Lehrstuhl 2

Prof. Dr. Amin Coja-Oghlan
amin.coja-oghlan@tu-dortmund.de

20. Juni 2023

Effiziente Algorithmen bilden das Rückgrat aller IT-Systeme und gute Algorithmiker sind dementsprechend heiß begehrt. Am LS2 befassen wir uns mit der Entwicklung und Analyse von Algorithmen, mit Anwendungen in den verschiedensten Bereichen. Wenn Sie sich für eine Bachelorarbeit bei uns interessieren, können Sie wahlweise einen eher angewandten oder einen theoretischen Schwerpunkt setzen. Für ersteres sollten Sie Programmierkenntnisse und Experimentierfreude, für letzteres Begeisterung an einer analytischen Arbeitsweise mitbringen. Einige mögliche Themen in den Bereichen

- *Kryptographie,*
- *Maschinelles Lernen,*
- *SAT-Problem,*
- *Kombinatorik*

sind im folgenden aufgelistet, wir sind aber immer auch für Vorschläge Ihrerseits offen. ***Kontaktieren Sie mich gern per email.***

Kryptographie

In der Kryptographie geht es darum, vertrauliche Informationen vor unbefugtem Zugriff zu schützen. Beispiele sind die verschlüss elte Kommunikation über öffentliche Netzwerke (z.B. “https”) sowie die Verschlüsselung von Daten auf Speichermedien. In der Kryptographie kommen vielfältige Algorithmen zur Ver- und Entschlüsselung zum Einsatz. Eine weitere Aufgabe besteht im Nachweis der Sicherheit von Verschlüsselungsverfahren, oder umgekehrt in der Untersuchung möglicher Angriffsstrategien.

Thema: Enigma

Die Enigma ist eine mechanische Verschlüsselungsmaschine aus der 1. Hälfte des 20. Jahrhunderts. Das Brechen dieses Verschlüsselungssystems, an dem Alan Turing beteiligt war, gilt als einer der Anfangspunkte der Informatik. In diesem Projekt soll ein Angriff auf die Enigma mit modernen Methoden versucht werden.

Referenzen:

- Wikipedia-Artikel [link]
- Katz, Lindell: Introduction to modern cryptography.

Thema: Primzahltests

Public Key Verschlüsselungsverfahren wie RSA benötigen schnelle Algorithmen zur Erzeugung zufälliger großer Primzahlen. Eine Strategie dafür ist, (im wesentlich) eine große zufällige Zahl (mit vielleicht 1000 Ziffern) zu erzeugen und anschließend zu prüfen, ob es sich um eine Primzahl handelt. Dazu benötigt man natürlich einen effizienten Primzahltest. In diesem Projekt geht es darum, einen solchen effizienten Primzahltest experimentell und/oder theoretisch zu untersuchen.

Referenzen:

- Wikipedia-Artikel [link]
- M. Dietzfelbinger: Primality testing in polynomial time.

Thema: Faktorisierungsalgorithmen

Bekannte Verschlüsselungsverfahren wie beispielsweise das häufig verwandte RSA-Verfahren beruhen auf der Annahme, daß es nur mit großem Aufwand möglich ist, eine große gegebene Zahl in ihre Primteiler zu zerlegen. Die Zahlen, von denen dabei die Rede ist, bestehen typischerweise aus einigen hundert oder einigen tausend Ziffern und haben keine "kleinen" Primteiler. Effiziente Faktorisierungsalgorithm könnten also diese Verschlüsselungen brechen. In dieser Arbeit geht es um den Vergleich verschiedener Faktorisierungsalgorithmen. Welche Verfahren existieren, auf welchen mathematischen Techniken beruhen sie, wie effizient lassen sie sich implementieren und wie groß sind die Zahlen, die sich mit diesen Verfahren faktorisieren lassen? In der Konsequenz: wie sicher ist das RSA-Verfahren?

Referenzen:

- Wikipedia-Artikel [link]
- T. Kleinjung: Factorization of a 768-bit RSA modulus. [link]
- What is Shor's factoring algorithm? [link]

Maschinelles Lernen

Aufgrund spektakulärer Erfolge wie AlphaGo ist in den letzten Jahren das maschinelle Lernen wieder ins Zentrum der Informatikforschung gerückt. Algorithmen spielen dabei natürlich eine fundamentale Rolle, z.B. zum Trainieren eines neuronalen Netzwerkes. Ziel der folgenden Themen ist eine wissenschaftliche Auseinandersetzung mit maschinellem Lernen: was geht, was geht (noch) nicht? Was können verschiende Paradigmen leisten?

Thema: Netzwerkanalyse

Ein grundlegendes Problem im maschinellen Lernen ist die Identifikation von "Struktur" in Daten. Dabei spielen Netzwerkdaten eine besondere Rolle. In diesem Projekt geht es um die Analyse eines großen, frei zugänglichen Netzwerkes, nämlich des Kollaborationsgraphen von Informatikforschern. Anhand dieses Datensatzes sollen Techniken wie z.B. Clusteringalgorithmen experimentell untersucht werden. Ist es beispielsweise möglich, mit einem solchen rein netzwerkbasieren Ansatz die fachliche Gliederung der Informatik zu rekonstruieren?

Referenzen:

- E. Abbe: Community detection and stochastic block models. [link]
- C. Moore: The computer science and physics of community detection. [link]
- dblp computer science bibliography. [link]

Thema: Das Planted Clique-Problem

Eine fundamentale algorithmische Herausforderung im maschinellen Lernen besteht im Feststellen von "Auffälligkeiten" in einem Datensatz. Ein bekanntes "Laborproblem", an dem sich diese Fragestellung besonders gut untersuchen läßt, ist das *planted clique problem*. Dabei wird eine relativ große Clique (vollständig verbundener Untergraph) in einen ansonsten rein zufälligen Graphen eingefügt. Die Problemstellung ist, den Graphen mit der eingebauten Clique vom Nullmodell, d.h. einem rein zufälligen Graphen, zu unterscheiden. Das Projekt eigent sich sowohl für eine eher praktische als auch für eine theorieorientierte Arbeit.

Referenzen:

- Wikipedia-Artikel [link]
- N. Alon, M. Krivelevich, B. Sudakov: Finding a large hidden clique in a random graph. [link]
- Q. Berthet, P. Rigollet: Complexity theoretic lower bounds for sparse principal component detection. [link]

Thema: Compressed sensing

Compressed sensing ist ein grundlegendes Problem im maschinellen Lernen. Es geht dabei grob gesagt darum, die wichtigsten Elemente eines Datensatzes möglichst effizient zu erkennen. In den letzten Jahren sind verschiedene neue Algorithmen für dieses Problem vorgeschlagen worden. Teils beruhen diese neuen Algorithmen auf Ideen aus der Informationstheorie und der statistischen Physik. Thema dieser Arbeit ist die experimentelle und/oder theoretische Analyse dieser Algorithmen.

Referenzen:

- Wikipedia-Artikel [link]
- D. Donoho: Compressed sensing. [link]
- F. Krzakala, M. Mezard, F. Sausset, Y. Sun, L. Zdeborova: Statistical physics-based reconstruction in compressed sensing. [link]

Das SAT-Problem

Das Erfüllbarkeitsproblem ('SAT') ist eines der wichtigsten algorithmischen Probleme überhaupt, weil es als Unterproblem in einer Vielzahl von Anwendungen begegnet (z.B. Datenbankabfragen, Korrektheit von Programmen). Dabei geht es darum, für eine gegebene aussagenlogische Formel eine Belegung der Variablen zu finden, die die Formel erfüllt. Dieses Problem ist NP-vollständig, d.h. es gibt sehr wahrscheinlich keinen effizienten Algorithmus, der dieses Problem immer löst. Durch diese Erkenntnis verschwindet das SAT-Problem aber natürlich nicht von der Bildfläche. Stattdessen stellt sich die Aufgabe, dem SAT-Problem mit möglichst guten Heuristiken beizukommen. Darum geht es bei den folgenden Themenvorschlägen.

Thema: Message Passing-Algorithmen

Eine neue Familie von Algorithmen für das SAT-Problem beruht auf "message passing". Die Idee ist, daß die Klauseln und Variablen der Formel einander Nachrichten senden, die immer wieder aufgrund der anderen Nachrichten, die die Variable/Klausel empfängt, aktualisiert werden. Wenn die Nachrichten konvergieren (d.h. sich nicht mehr wesentlich verändern), kann möglicherweise eine erfüllende Belegung abgelesen werden. Interessanterweise scheinen Message Passing-Algorithmen gerade mit solchen Instanzen gut zurechtzukommen, an denen andere Algorithmen scheitern. Ziel dieses Projektes ist es, Message Passing-Algorithmen auf verschiedenen Typen von SAT-Instanzen zu erproben und/oder zu analysieren.

Referenzen:

- L. Croc, A. Sabharwal, B. Selman: Survey Propagation revisited. [link]
- J. Pearl: Causality. [link]

Kombinatorik

Kombinatorische Strukturen wie Graphen, Bäume, Polytope oder Codes bilden die mathematische Grundlage der Algorithmik. Die Entwicklung von Algorithmen beruht häufig auf kombinatorischen Erkenntnisse, und führt umgekehrt auf neue kombinatorische Fragestellungen. Die folgenden Themenvorschläge behandeln einige davon.

Thema: der Vierfarbensatz

Eine der faszinierendsten Aussagen der Graphentheorie ist der *Vierfarbensatz*: die chromatische Zahl eines ebenen Graphen ist höchstens vier. Dabei ist die chromatische Zahl eines Graphen definiert als die kleinste Zahl von Farben, die genügt, um die Knoten des Graphen so zu färben, daß keine Kante zwei Knoten mit derselben Farbe verbindet. Ferner ist ein ebener Graph ein Graph, der so in die Ebene gezeichnet ist, daß sich keine zwei Kanten überkreuzen. In diesem Projekt soll der Beweis des Vierfarbensatzes, der auf Computerunterstützung beruht, neu implementiert werden.

Referenzen:

- Wikipedia-Artikel [link]
- Seite von Robin Thomas [link]

Thema: Rejection sampling

In diesem Projekt geht es um Algorithmen, die Objekte mit bestimmten erwünschten Eigenschafen zufällig erzeugen. Wir nehmen an, daß der Algorithmus einfache Münzwürfe durchführen kann, und das Ziel ist, daraus Algorithmen zur Erzeugung komplexerer Objekte zu entwickeln. In den letzten Jahren wurden neue Zugänge zu diesem Problem entwickelt, unter denen die Idee des "rejection sampling" hervorsticht. In diesem Projekt soll es darum gehen, dieses Verfahren auf neue kombinatorische Probleme zu übertragen und anzuwenden.

Referenzen:

- M. Jerrum: Fundamentals of partial rejection sampling. [link]
- B. Barak: Complexity of counting. [link]

Thema: Travelling Salesman Problem

Das TSP ist eines der bekanntesten Optimierungsprobleme: es soll eine möglichst kurze Rundreise durch eine Menge von 'Städten' gegeben werden. Obwohl das Problem NP-schwer ist, können bemerkenswert grösse Instanzen mit heuristischen Methoden schnell gelöst werden. In diesem Projekt geht es darum, solche heuristischen Methoden zu implementieren und zu evaluieren. Die Algorithmen beruhen dabei teils auf ganzzahliger Programmierung, branch-and-bound oder dynamischer Programmierung.

Referenzen:

- Wikipedia-Artikel [link]
- slides by Martin Grötschel [link]

Thema: Rubik's cube

Der "Zauberwürfel" ist ein bekanntes kombinatorisches Spielzeug, das in diesem Projekt algorithmisch angegangen werden soll. Die Zielsetzung dabei ist ein möglichst effizienter Algorithmus, der das Problem löst. Algorithmisch ergeben sich direkte Verbindungen zur Graphentheorie, aber auch zur Algebra, insbesondere Gruppentheorie. Als weiterer Aspekt kann die Komplexität des Zauberwürfels behandelt werden.

Referenzen:

- Wikipedia-Artikel [link]
- "God's number" [link]