

Inhalt

Editorial	2
Iterationsschleife	5
Efficient Label Propagation	6
“Intel MIC Programming” & “HPC for Water-Related Hazards” Workshops @ LRZ	13
SPPEXA Doctoral Retreat	17
TUM IAS Workshop “Hybrid Simulation Methods in Fluid Dynamics”	18
Compact course “Scalable kernel methods in machine learning”	19
Wire Follows Form Follows Function	21
PhD Forum auf der ISC 2016	22
Big Data Meets Computation @ IPAM	24
BGCE Student Paper Preis	25

Quartl* - Impressum

Herausgeber:

Prof. Dr. A. Bode, Prof. Dr. H.-J. Bungartz, Prof. Dr. U. Rüde

Redaktion:

S. Herrmann, Dr. S. Zimmer

Bildquelle: <https://openclipart.org/detail/176598/cartoon-wolf>

Technische Universität München, Fakultät für Informatik

Boltzmannstr. 3, 85748 Garching b. München

Tel./Fax: ++49-89-289 18636/18607

e-mail: herrmasa@in.tum.de, **www:** <http://www5.in.tum.de/quartl>

Redaktionsschluss für die nächste Ausgabe: **01.12.2016**

* Quartl: früheres bayerisches Flüssigkeitsmaß,

→ das Quartl: 1/4 Kanne = 0.27 l

(Brockhaus Enzyklopädie 1972)

Das Quartl erhalten Sie online unter <http://www5.in.tum.de/quartl/>



Das Quartl ist das offizielle Mitteilungsblatt des Kompetenznetzwerks für *Technisch-Wissenschaftliches Hoch- und Höchstleistungsrechnen in Bayern* (KONWIHR) und der *Bavarian Graduate School of Computational Engineering* (BGCE)

Editorial

Sagt Ihnen der Name Pater Pio etwas? Falls nein, dann lesen Sie weiter; falls ja, auch weiterlesen! Mich hat die Erhellung auf der diesjährigen Ferienakademie ereilt, und so gebe ich meine neusten diesbezüglichen Erkenntnisse gerne weiter. Ein weiteres Mal zeigt sich somit, was man beim interdisziplinären Austausch mit Kollegen auf Wanderungen im Sarntal so alles erfahren kann ...

Francesco Forgione wurde geboren im Jahre 1887 in Pietrelcina in Kampanien (Italien). Bereits im Alter von sechs Jahren hatte er ein sehr reges und daher ungewöhnliches „Gebetsleben“ entwickelt, wie es so schön heißt. Schon in seiner Kindheit ereigneten sich wundersame Dinge. So wird beispielsweise berichtet, dass sich bei einer Wallfahrt vor seinen Augen die Heilung eines behinderten Kindes vollzog. Nach der Schule trat er mit 15 dem Kapuzinerorden bei und nahm seinen Ordensnamen Bruder Pio an. Die folgenden Jahre waren geprägt von einer starken Frömmigkeit, der Erkrankung an Tuberkulose sowie von mehreren mysteriösen weiteren Krankheiten mit Symptomen wie plötzlichem Erglühen. Die konsultierten Ärzte waren ratlos (was, zugegebenermaßen, nicht notwendigerweise etwas mit den Krankheiten zu tun haben muss, sondern auch an den beschränkten Fähigkeiten der Ärzte gelegen haben kann). Göttliche Visionen und dämonische Qualen nahmen immer mehr zu. Seine Antwort war das Gebet. Ende 1915 kam der italienische Staat auf die wenig geistreiche Idee, den kranken Pio zum Militärdienst einzuziehen – um ihn ein gutes Jahr später aus Gesundheitsgründen wieder aus selbiger zu entlassen.

Noch im Jahr 1918 bekam Pio, zwischenzeitlich vom Frater zum Pater avanciert, dann die „Wundmale Christi“, eine Art Stigmatisierung. Bald war von einem Wunder die Rede, und so wuchs die Schar derer, die Pater Pio um Rat fragen wollten. Neben den Wundmalen passierten angeblich noch andere wunderbare Dinge. So wird von einem ungewöhnlichen Duft berichtet, den er ausstrahlte (mei, das kann jetzt wieder irdische Gründe haben ...). Die Aufmerksamkeit des Quartls erweckte unser italienischer Freund allerdings durch eine andere Fähigkeit, die sich so mancher von uns auch schon mehrfach gewünscht hat: die Fähigkeit zur Bilokation. Gleich mehrfach

Einreichungsschluss für den BGCE Student Paper Prize ist der 11. Dezember 2016. Beiträge werden bitte in pdf-Format an rippm@in.tum.de

geschickt. Wir freuen uns neben den traditionell zahlreichen internationalen Einreichungen natürlich auch über viele Beiträge aus dem Quartl-Land!

Wer es genauer wissen will: Die Preisträger der vergangenen fünf BGCE-Preise sind:

- 2015: David Emerson (Tufts University, Massachusetts)
- 2013: Tobin Isaac (ICES, UT Austin)
- 2011: Andrea Manzoni (EPF Lausanne)
- 2009: Gisela Widmer (ETH Zürich) and Chad Lieberman (MIT)
- 2007: Alfonso Bueno Orovio (Universidad de Castilla-La Mancha, Spain)

Michael Rippl

wird sie ihm zugeschrieben. Ein besonders prägnantes Beispiel: zeitgleich auf einem Schlachtfeld, den Suizidversuch eines Generals nach verlorener Schlacht vereitelnd, und in seiner Klosterzelle. Fragen Sie mich jetzt nicht nach der damaligen Synchronisation der Zeitmessung – eine derart popelige Frage wäre der Erhabenheit des Phänomens überhaupt nicht angemessen.

Auch bei Pio gilt der Spruch „Wer hat, dem wird gegeben“ – immer größer wurde die Liste seiner Wunder. Angeblich konnte er bei der Beichte seinem Gegenüber „ins Herz schauen“ und so ermitteln, ob die gezeigte Reue echt oder vorgegeben war. Dass er auch zu Prophezeiungen in der Lage war, wird da fast zur Selbstverständlichkeit. So wird ihm nachgesagt, dem jungen Priester Karol Wojtyła 1947 sowohl die Wahl zum Oberhaupt der katholischen Kirche als auch das Attentat von 1981 vorausgesagt zu haben. Allerdings nahmen auch seine Leiden und mit diesen auch vermeintlich abstruse Verhaltensmuster zu.

In der Folge zeigte die katholische Kirche dann das gesamte Spektrum ihrer Aktionsmöglichkeiten. Die Glaubenskongregation erklärte Pater Pio zunächst für einen Besessenen. Es kam zu einem Verfahren, an dessen Ende Mess- und Beichtverbot in der Öffentlichkeit standen. 1933 wurde selbiges wieder aufgehoben. Sein Leben jedoch blieb geprägt von Frömmigkeit, Leiden und wundersamen Vorkommnissen. 1968 schließlich verstarb er – kurz nachdem sich seine Wundmale zum fünfzigsten Male geöffnet hatten.

Bereits 1973 wurde der Prozess der Seligsprechung eingeleitet, und 1999 vollzog Papst Johannes Paul II. diese dann. Am 16. Juni 2002 erfolgte vor angeblich fast 1 Million Gläubiger schließlich die Heiligsprechung. Noch nie in der neueren Kirchengeschichte war bis dato eine Person so kurz nach ihrem Tod heiliggesprochen worden.

Als Evangelischer stehe ich diesem Rummel um Selige und Heilige ja eher etwas distanziert gegenüber. Und in der Tat gab und gibt es massive Zweifel an so ziemlich allen Wundern von Pater Pio. Zum Beispiel hält sich der Vorwurf der Selbstzerstümmelung recht hartnäckig, und auch mit der aufkommenden faschistischen Bewegung wird Pater Pio in Verbindung gebracht. Aber falls dieser Kerl tatsächlich einen Durchbruch in Sachen Bilokation erreicht haben sollte, dann wäre er mein Mann, kein Zweifel. Als Informatiker interessiert mich dann natürlich auch sofort die Frage der Verallgemeinerung

in Richtung Multilokation. Induktion, Rekursion, Tensorprodukt ... irgendwas müsste dann doch gehen. Noch schöner wäre, wenn sich das irgendwie mit Heisenbergs Unschärfe koppeln ließe: Ich kann gleichzeitig an zwei Orten sein, aber alle anderen wissen nicht wann ... oder so ähnlich. Da kommt die Phantasie auf Touren. Also behalten wir unseren Pater lieber auf dem Radar – wer weiß, vielleicht wird's ja doch noch was mit meiner Bilokation. Damit wäre ich schon ganz zufrieden.

Doch genug der einleitenden Worte – die Quartl-Redaktion wünscht allen einen goldenen Herbst. Und vor allem wünschen wir natürlich anhaltenden Spaß mit dieser neuen Ausgabe Ihres Quartls!

Hans-Joachim Bungartz

BGCE Student Paper Preis

Auf der SIAM CS&E Konferenz 2017, die vom 27. Februar bis 03. März in Atlanta, Georgia, stattfinden wird, wird der 6. BGCE Student Paper Preis¹ für hervorragende studentische Arbeiten im Bereich des Computational Science and Engineering vergeben werden.

Stifter des Preises ist die Bavarian Graduate School of Computational Engineering (BGCE), die als Elitestudiengang im Elitenetzwerk Bayern ein gemeinsames Aufsatzmodul für die besten und motiviertesten Studierenden der internationalen Masterprogramme Computational Engineering (CE) der Universität Erlangen-Nürnberg (FAU) sowie Computational Mechanics (CO-ME) und Computational Science and Engineering (CSE) an der TU München (TUM) darstellt. Der Gewinner bzw. die Gewinnerin des Preises wird im kommenden Sommer eine Woche lang kostenlos Bayern, die FAU und die TUM besuchen und intensiven Kontakt zu Vertretern der BGCE pflegen. Das Hauptanliegen des Preises ist es, exzellente Studierende im Bereich CS&E zu fördern und internationalen Austausch in einer frühen Karrierephase zu unterstützen.

Für den BGCE Student Paper Preis können sich MasterstudentInnen und DoktorandInnen bewerben, indem sie ein kurzes Paper von maximal vier Seiten einreichen. Die Preis-Finalisten werden ihre Arbeit in einem speziellen "CS&E Student Prize Minisymposium" auf der SIAM CS&E 2017 vorstellen. Die Paper und Vorträge werden dann von einer internationalen Jury bewertet. Ausgeschlossen von der Teilnahme sind lediglich Studierende der FAU und der TUM.

¹ http://www.bgce.de/news/bgce_prize.html

Big Data Meets Computation @ IPAM

Dear friends of high performance computing and Big Data, we are glad to invite you to the upcoming workshop “Big Data Meets Computation” that will take place between January 30th and February 3rd, 2017 in the venues of the Institute for Pure & Applied Mathematics (IPAM) at UCLA.

The workshop aims to advance the fusion of HPC and Big Data, bringing together leading innovators and pioneers from applied mathematics, computer science, and various areas of applications.

While the fusion of HPC and Big Data is a young field, the potential of considering computing and Big Data technologies as parts of a wholesome solution is immense and promises the development of new and efficient methods for representation, transformation, and transfer of large amounts of data. These methods can be applied to a large number of scientific and industrial applications.

Please visit www5.in.tum.de/DMC2017 for further details and registration information. Registered participants will have an opportunity to present and discuss their research in a poster session.

The organizing committee (Rick Archibald (ORNL), Hans Bungartz (TUM), Frank Jenko (UCLA), and Stan Osher (IPAM)) will be glad to welcome you at the workshop.

Hans-Joachim Bungartz

Iterationsschleife

N=20

19. September 2016

Glaubt man den Gebrüdern Grimm so ereignete sich die Geschichte so, dass ein Wolf sieben Geißlein fraß als deren Mutter nicht zu Hause war. Die Geißlein konnten schließlich von der Mutter gerettet werden und der Wolf kam zu Tode. So weit, so gut.

Betrachten wir jedoch die Sache näher. Eine Geiß ließ ihre Kinder alleine zu Hause, weil sie für Essen sorgen musste. Das ist zunächst nicht verworlich doch stellen sich einige Fragen.

Zunächst ist zu fragen, wo der Vater der Kinder abgeblieben ist. Möglicherweise war er schon bei der Arbeit oder vielleicht in dieser Woche gänzlich auf Dienstreise. Insofern wäre seine Abwesenheit verzeihlich, jedoch müsste ein Teil der Schuld dann vielleicht ihn treffen. Doch da die Brüder Grimm keinerlei Hinweise geben, so muss davon ausgegangen werden, dass es sich um einen alleinerziehende Geiß gehandelt hat.

Die Tatsache der sieben Kinder wiederum wirft weitere Fragen auf. Sieben Kinder bedeuten einen ausgesprochen hohen Aufwand. Warum die alleinstehende Geiß in einer solchen Situation nicht von der Caritas oder einer anderen vielleicht sogar staatlichen Stelle bei der Betreuung der Geißlein unterstützt wurde ist doch überraschend. Für eine Betreuung zu Hause reicht das Kindergeld natürlich nicht aus, doch sollten die einschlägigen gesetzlichen Vorschriften über die Bereitstellung ausreichender Kinderbetreuungskapazitäten bis zum 6ten Lebensjahr gerade in so einer Situation greifen (wir können mit Sicherheit davon ausgehen, dass die Geißlein weniger als 6 Jahre alt waren). Hier stellt sich die Frage, ob ein verpflichtender Besuch eines Kindergartens die Geißlein nicht vor einem schweren Trauma bewahrt hätte. Andererseits hätte eventuell ein Betreuungsgeld der Mutter die Beschäftigung einer adäquaten Betreuungskraft ermöglicht. Dann stellt sich die Frage nach der Angemessenheit der von der Geißleinmutter getroffenen Abwehrmaßnahmen gegen allfällige Gefahren. Letztlich überläßt sie ihre Kinder mit einer doppelten IF-Abfrage einem ausgesprochen unzuverlässigen Algorithmus. Einzig Stimme und Potentfarbe werden überprüft, während eine visuelle Kontrolle (etwa eine einfache facial recognition) die Kinder vor weiterem Schaden hätte bewahren können. Hier zeigt sich möglicherweise ein Mangel in der IT-Ausbildung in Deutschland. Die Geißleinmutter war sich der Komplexität der Situation nicht bewusst und konnte daher nicht angemessen algorithmisch reagieren. Zuletzt stellt sich die Frage der Mittäterschaft für den Bäcker, der ganz offenbar die Absicht einer Straftat erkennend dennoch Beihilfe geleistet hat. Die Tatsache, dass nur der Wolf zur Verantwortung gezogen wird hinterläßt beim Leser doch einen unguuten Nachgeschmack. In Summe wirft die Grimmsche Geschichte mehr Fragen auf als sie beantwortet. Insbesondere bei kritischer Betrachtung der sozialen Umstände erscheint das Verhalten der handelnden Personen/Tiere unangemessen.

Efficient Label Propagation

Starting October 2016, Stephan Guennemann will be Professor for Data Mining and Analytics at TUM. His research group focuses on the development of robust data mining and machine learning principles with the goal to support the analysis and understanding of the massive amounts of data collected in science and industry. In particular, the group studies principles for analyzing networks/graphs and temporal data. For further information visit <http://www.kdd.in.tum.de>

1 Introduction

The increasing potential of storage technologies combined with enhanced information systems and ubiquitous computing infrastructure has opened the possibility to conveniently and affordably gather massive amounts of complex data – ranging from online services, over cyber-physical systems, to the health sciences. Extracting meaningful information from this ever-growing data is one of the crucial challenges to realize the next generation of high impact applications.

While standard data analysis techniques are known for decades, their inherent assumptions often do not meet the characteristics of today's data, thus, rendering them rather useless: (a) Many standard techniques assume that the observations we make are independent. However, in reality, data shows complex dependency structures. Sensors, for example, interact with each other in complex cyber-physical systems. Similarly, people exchange information in social networks. (b) Standard techniques assume the data to be relatively clean. However, data gathered in experiments and real life applications is often corrupted, prone to outliers, or contains missing information. Errors in the measurement devices or incomplete user inputs are only a few examples.

Mit dem Thema *Handling Hard and Soft Faults with the Sparse Grid Combination Technique* konnte ich die Jury überzeugen und sicherte mir somit den *PhD Forum Award*. Als Preis gab es ein iPad Air 2, welches von den Organisatoren gesponsert war. Das ist definitiv ein guter Anreiz für eine Teilnahme nächstes Jahr, denn das PhD Forum hat es auch in das Programm der ISC 2017 geschafft.



Verschiedene SPPEXA Projekte waren auf der Veranstaltung vertreten. Organisiert wurde das Forum von Prof. Lorena Barba, Dr. Georg Hager, Prof. Gerhard Wellein und Dr. Rolf Rabenseifner. Eine Veranstaltung, die man im Auge behalten sollte!

Alfredo Parra



PhD Forum auf der ISC 2016

Dieses Jahr fand wieder die *ISC High Performance* Konferenz in Frankfurt statt. Mit über 3.000 Teilnehmern aus 53 verschiedenen Ländern und 146 Firmen und Forschungseinrichtungen aus dem Bereich Hochleistungsrechnen hat sich die ISC als eine der wichtigsten Informatik-Konferenzen weltweit etabliert. Wer sich also für die aktuellsten Trends im HPC interessiert, darf die ISC nicht verpassen. Zum 5-tägigen Programm zählen Tutorials, Workshops, Ausstellungen, Poster Sessions, die Student Cluster Competition und verschiedene Social Events.

Neu im Programm war dieses Jahr das PhD Forum, eine ganz besondere Veranstaltung, welche gezielt für Doktoranden aus dem HPC Bereich gedacht war. Das Ziel der Veranstaltung war es, den Ideenaustausch in einem dynamischen Format zu fördern. Das Forum bestand aus zwei Teilen: Im ersten Teil durfte jeder der insgesamt 14 Doktoranden seine Forschungsergebnisse als *Lightning Talk* präsentieren: 5 Folien in genau 4 Minuten. Die Folien wurden alle 48 Sekunden automatisch gewechselt - der Vortrag musste also perfekt getimt sein, um peinliche Übergänge, wie etwa abrupte Unterbrechungen oder zu lange Stille, zu vermeiden. Dabei haben sich ein paar Vortragende entschieden, sicher zu gehen und ein vorgefertigtes Skript vorzulesen, was allerdings keinen allzu guten Eindruck gemacht hat. Der zweite Teil war eine klassische einstündige Poster Session, bei der man tiefer auf die verschiedenen Themen eingehen konnte.

Da das Forum gleichzeitig ein Wettbewerb war, wurde die Qualität der Vorträge und des Auftretens der Teilnehmer bei der Poster Session von einer Jury von HPC-Experten evaluiert und diskutiert. Dies motivierte die Teilnehmer, gute Beiträge zu präsentieren - sicher kein schlechter Ansatz.

Based on these observations, our research targets at the design and analysis of mining and learning principles taking *dependency and impurity* of data into account. Specifically, we focus on principles for temporal data and complex networks, where we study tasks such as clustering, anomaly and event detection, or label propagation and prediction. In the following, we give a brief introduction to the task of label propagation and we show one of our recent advances in this field.

2 Label Propagation

How can we tell which accounts are fake or real in a social network? How can we spot fraudulent sellers or products in an online recommendation site like Amazon? How can we label news articles (politics/sport/other) and events? Often we can answer such questions accurately by knowing the labels of a few instances only and exploiting the underlying "network effects" of the data.

Technically, network effects capture the inherent dependency structure of the data. A typical example is homophily ("birds of a feather flock together"): if we know, e.g., the political leanings of most of Alice's friends on Facebook, then we have a good estimate of her leaning as well. Assuming homophily, the node labeled with '?' in Fig. 1 (left), is most likely blue/liberal and not orange/conservative since it is connected to many blue nodes. Occasionally, the reverse is true as well, also called heterophily ("opposites attract"). In an online dating site, we may observe that talkative people prefer to date silent ones, and vice versa. Assuming heterophily, the node a in Fig. 1 (middle) is most likely orange since it is connected to two blue nodes. Node b , in contrast, is likely blue since it connects to a and another orange node.

Thus, knowing the labels of a few nodes in a network, plus knowing which network effects hold, we can usually give good predictions about the remaining nodes' labels by propagating the label information through the network. Formally, the different kinds of network effects can be represented by a so-called compatibility matrix - denoted with H in the figures. It models the relative affinities between the labels of two adjacent nodes. In Fig. 1 (left),

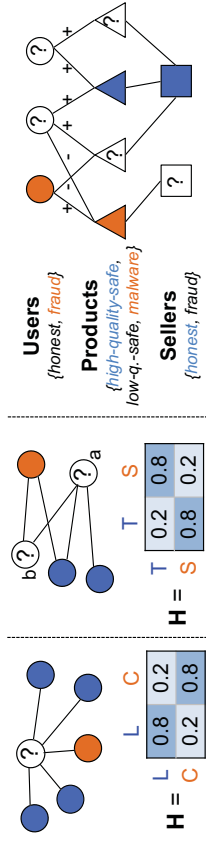


Figure 1: Inferring the labels of nodes marked with '?' based on network effects. Left: Homophily; Middle: Heterophily; Right: Heterogenous graph with different types of nodes. Each edge might represent different network effects.

for example, we assume that it is more likely to observe adjacent nodes having a similar label, $H(i, i) = 0.8$, than having different ones, $H(i, j) = 0.2$.

Naturally, this problem generalizes to more complex scenarios as illustrated in Fig. 1 (right): Here, we illustrate a network having different node types, each with a specific set of labels, and different types of edges, each with a potentially different compatibility matrix H_* . Given, e.g., a *positive* user-product rating, it is likely that either the user is honest and the product of high quality, or the user is a fraudster and the product is malware. However, (given the positive rating) it is unlikely that the user is honest and the product is malware. An edge representing a *negative* rating, in contrast, would describe different affinities. Such scenarios are typical for e-commerce and rating websites as we will see later.

In all of the above scenarios, the underlying problem we are trying to solve is the same: Given a network with a few labeled nodes as well as the compatibility matrices H_{uv} between pairs of connected nodes (u, v) , try to predict the labels of the unlabeled nodes. This learning scenario, where we reason from observed training cases directly to test cases, is also called transductive inference.

Wire Follows Form Follows Function

Auf der diesjährigen ISC High Performance haben wir, angestoßen durch das ExaHyPE Projekt, einen Workshop organisiert: Form Follows Function lud Sprecher ein um auszuleuchten, wie Hardware-Entwicklungen unsere algorithmische Arbeit prägen, ob es eine umgekehrte Beeinflussung gibt und ob das vielzitierte Co-Design auch außerhalb dedizierter Co-Design-Projekte stattfindet.

Es war sicher ein Erfolg, dass der Workshop bei einer Akzeptanzquote von unter 20% angenommen worden ist—obwohl offensichtlich einer der fünf Gutachter ein Co-Designer war und sich sichtbar von der im Workshop-Konzept zum Ausdruck gebrachten Skepsis gegenüber Co-Design angegriffen gefühlt hat. Es war auch sicher ein Erfolg, dass während des Workshops eine sehr lebendige Diskussion und Austausch stattgefunden haben. Es war aber mit Sicherheit der größte Erfolg, dass die Zusammenfassung des Workshops auf arXiv von HPCWire und anderen News-Seiten diskutiert wurde. Interessierte finden das Material unter

- <http://exahype.eu>
- <http://arxiv.org/pdf/1607.02835v1.pdf>
- <https://www.hpcwire.com/2016/07/12/isc-workshop-tackles-co-developments-thorny-challenges/>

Tobias Weinzierl

The course was open to all interested students but in particular students of BGCE, TopMath, CSE, Mathematics, and Informatics.

Prof. Biroš also gave a presentation at the informatics department about his recent publication in SC16 conference about “A Fast Solver for Constrained Diffeomorphic Image Registration”.

Arash Bakthiari

2.1 Belief Propagation and its Limitations

One of the most frequently used principles for the above kind of transductive inference is Belief Propagation (BP [1]). Here, the task is considered from a probabilistic point of view: the given network structure is treated as a (pairwise) Markov random field and the compatibility matrices are used as the corresponding clique potentials. Accordingly, label propagation becomes a probabilistic inference task: inferring the most likely label of a node – and Belief Propagation can be used to perform approximate inference in such a probabilistic model.

Intuitively, BP (also known as sum-product message passing) sequentially propagates messages (say, m_{uv}) from one node u to another v which represents u 's belief about v 's labels. Here, m_{uv} is a vector representing the beliefs $m_{uv}(i)$ regarding each label i . Starting with the prior beliefs for a certain subset of the nodes in a graph (e.g., e_u ; prior knowledge about u 's labels), this process is carried out until a steady state is reached – assuming convergence. Formally, the message update for all connected pairs of nodes u, v and labels i is given by [2]:

$$m_{vu}(i) \leftarrow \frac{1}{Z_{vu}} \sum_{j \in L(v)} H_{uv}(i, j) e_v(j) \prod_{w \in N(v) \setminus u} m_{vw}(j). \quad (1)$$

Here, H_{uv} denotes the compatibility matrix between the corresponding two nodes, $N(v)$ the set of nodes adjacent to v in the network, and $L(v)$ the set of potential labels of node v . Finally, the normalization factor Z_{vu} ensures that the messages sum up to a constant (typically 1) at any iteration.

After convergence of the iterative update process, the final beliefs about a node's label (e.g., b_u ; the inferred information about the label of u) are recovered from the messages that a node receives:

$$b_u(i) \leftarrow \frac{1}{Z_u} e_u(i) \prod_{v \in N(u)} m_{vu}(i) \quad (2)$$

Again, Z_u is a normalization factor ensuring that the beliefs sum up to 1.

While BP has been successfully applied in many applications such as error-correcting codes, stereo matching in computer vision, and fraud detection, it has one strong limitation: its convergence for graphs with loops can not be guaranteed in general [1]. These convergence problems lead to a questionable use in real world applications (see [3] for a detailed discussion from a practitioner’s point of view). Furthermore, as we will see later, BP’s absolute runtime is relatively high for very large graphs.

3 Proposed Method

Due to the drawbacks of BP, we developed a novel label propagation principle (we presented a special case in [4]). Unlike traditional BP, our method provides a closed-form solution for the beliefs and also comes with strong convergence guarantees for its iterative algorithm. Our method scales linearly with the graph size and is up to 600 times faster than BP – still it almost perfectly approximates the result of BP.

Core ideas: Residuals, Linearization, and Message Absorption. Our method is based on three intuitive ideas. (1) Instead of considering the ‘original’ messages/beliefs/compatibility values (m, b, H, e) as defined before, we focus on the corresponding residuals. That is, given appropriate center values, we only consider the values’ deviations. For the compatibility matrix in Fig. 1 (left) and the center value 0.5, for example, we obtain the residual matrix

$$\hat{H} = \begin{bmatrix} 0.3 & -0.3 \\ -0.3 & 0.3 \end{bmatrix}. \quad \text{Residuals are denoted with a hat. (2) Rewriting Eq. (1)}$$

and (2) in terms of residuals, and transforming them into the logarithmic domain, opens the door for the second step: a linearization of the update system. Since usually the residuals are small w.r.t. the center, this linearization leads to good approximation results. (3) Since now the (residual) update of \hat{m}_{v_u} corresponds to a linear system, we can derive a closed-form equation for the (residual) message values when they reach their steady state. That is, the computation of the steady state message values \hat{m}_{v_u} does *not* depend on other messages anymore. By plugging in this closed-form into the residual update equation of b_u – the residual and linear version of Eq. (2) –, we obtain

Compact course “Scalable kernel methods in machine learning”

In 2014, Prof. George Biros (Institute for Computational Engineering Sciences, University of Texas at Austin) joined the focus group “HPC& the Institute for Advanced Study (IAS), as a Hans Fischer Fellow to focus on the Multi-Core Challenge. He has visited TUM and collaborated with his doctoral candidate, Arash Bakhtiari, each summer since then.

During his visit this year, Prof. Biros and PhD candidate Arash Bakhtiari worked on developing an arbitrary-order accurate time marching scheme based on the spectral differer correction method. The method will be integrated with a Semi-Lagrangian advection-diffusion solver developed by Arash, which is already arbitrary-order accurate in space, and a related publication was accepted in the Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC16) in Salt Lake City. The team also worked on extending the advection-diffusion solver to an incompressible Navier-Stokes solver, where the nonlinear convective term will be solved with the Semi-Lagrangian method and the Stokes term with a volume integral equation formulation.

Prof. Biros, together with Dr. Philipp Neumann from the SCCS chair, also organized the workshop “Hybrid Simulation Methods in Fluid Dynamics: Models, Software and Application”. The workshop took place from June 20. to 21. at the TUM IAS and it was designed to encourage collaboration and exchange between researchers from physics, computer science and engineering. In the workshop, the state-of-the-art methods for hybrid flow simulation models, algorithms and software were discussed. In addition, Prof. Biros organized the compact course “Scalable kernel methods in machine learning” for TUM master students. In the course, machine learning related topics like supervised, unsupervised, and approximation algorithms, in particular nearest-neighbors, regression, density estimation, and scattered data approximation, were discussed. In addition, the kernel methods were introduced and their applicability and power in well-understood machine learning algorithms were explained.

TUM-IAS Workshop “Hybrid Simulation Methods in Fluid Dynamics”

Am 20./21. Juni 2016 fand am TUM Institute for Advanced Study der Workshop *Hybrid Simulation Methods in Fluid Dynamics: Models, Software and Applications* statt.

Das Ziel des von George Biros, Jens Harting, Philipp Neumann und Jason Reese organisierten Workshops war der interdisziplinäre Austausch von Experten aus der Physik, dem Ingenieurwesen und der Informatik in Bezug auf die Entwicklung von hybriden—beispielsweise Molekulardynamik-Kontinuum—Simulationen in der Fluidmechanik. Welche hybriden Ansätze sind derzeit verfügbar? Wie können akkuratere hybride Simulationsmodelle in effiziente Algorithmen und Software übertragen werden? Welche Herausforderungen birgen aktuelle Anwendungsbereiche?

Dies sind nur einige der Fragestellungen zur Modellierung hybrider Systeme, zu potentiellen Anwendungen und zu HPC-Aspekten, die im Rahmen des Workshops näher beleuchtet und erörtert wurden. Neben Plenar-Vorträgen und vielseitigen Diskussionen in den Kaffee- und Mittagspausen fand zudem reger Austausch zu dieser Thematik in einer Poster-Session statt.

Die Organisatoren bedanken sich beim IAS und SPPEXA für die finanzielle Unterstützung, bei der IAS-Crew für die Vorbereitungen und die Unterstützung in der Durchführung sowie bei allen eingeladenen Sprechern für die interessantesten Vorträge und die somit gelungene Gestaltung des Workshops.

Philipp Neumann



equations for the beliefs which solely depend on \hat{b} , \hat{H} , and \hat{e} . Thus, the messages got absorbed and do not need to be computed anymore.

The overall result. Exploiting the previous three steps leads to a linear system for computing the belief values \hat{b}_v . For simplicity, we present here the result for a frequently used special case only: A homogeneous graph where all compatibility matrices are identical. This scenario corresponds to, e.g., a social network of users as shown in Fig. 1 left or middle. In this case, the beliefs can be computed as shown in Fig. 2 where the values were arranged in matrix form for convenience.

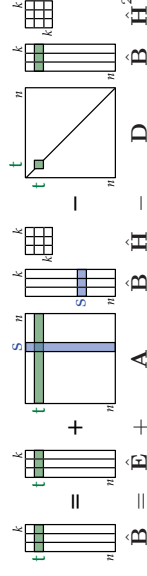


Figure 2: Residual belief update. Notice our matrix conventions: $\hat{B}(s, i)$ is the belief that node s has label i , and $\hat{E}(s, i)$ the prior belief (0 for unlabeled nodes). $\hat{H}(j, i)$ indicates the affinity of a node’s label j to its neighbor’s label i . $A(s, t) = A(t, s) \neq 0$ if the edge (s, t) exists in the network, and D is the network’s degree matrix.

This matrix equation shows the full potential of our method: (a) Since the computation of the beliefs is now based on simple matrix operations, it is straightforward to implement and interpret. Intuitively, the belief of a node consists of three terms: The prior beliefs about the node’s labels (\hat{E}), the beliefs of the *adjacent* nodes *modulated* by the compatibility matrix ($\hat{A}\hat{B}\hat{H}$), and a term which prevents two (or more) nodes mutually reinforcing each others’ beliefs ($D\hat{B}\hat{H}^2$). (b) Given this linear system, we can easily provide necessary and sufficient convergence guarantees. More precise, the above belief computation converges if and only if $\rho(\hat{H}\hat{A} - \hat{H}^2 \otimes D) < 1$ where ρ denotes the spectral radius and \otimes the Kronecker product. This property cannot be achieved by BP. (c) Since the network A is usually sparse the computations are highly efficient by exploiting well-established sparse-matrix computation engines.

4 Results and Future Work

As one of the many applications of our method, we present here its strength for detecting e-commerce fraud. This study was done in collaboration with Flipkart, an e-commerce company that provides a platform for sellers to market their products to customers. The dataset we used consists of about 1M users, their 3.3M ratings to 512K products, and connections between 4K sellers and their products. This matches the setting of Fig. 1 (right). In addition, we also have 1.7M ratings from users about sellers. Our goal is to classify users, products and sellers as fraudulent or not, given 50 manually labeled fraudsters as seed labels. We obtained the following results: (i) Our method identifies fraudulent sellers with near-perfect precision: 92.3%. (ii) The labels obtained by our method and BP show almost perfect correlation. That is, our approximations perform well empirically. (iii) Our method is around 600 times faster than BP and runs on this graph in a few seconds. In summary, our proposed method enables us to perform label propagation on very large data with high accuracy and efficiency.

As some ongoing work, we are currently exploring scenarios where the network contains additional feature information to further improve the prediction accuracy. Moreover, we are investigating principles to enrich the prediction by uncertainty scores.

Stephan Guennemann

References

- [1] J. Pearl. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo, 1988.
- [2] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. Exploring artificial intelligence in the new millennium, 8: 236-239, 2003.
- [3] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3): 93-106, 2008.
- [4] W. Gatterbauer, S. Guennemann, D. Koutra, and C. Faloutsos. Linearized and single-pass belief propagation. *VLDB*, 8(5): 581-592, 2015.

SPPEXA Doctoral Retreat: Vive la France!

Von 5.-9. September 2016 fand das diesjährige SPPEXA Doctoral Retreat in Obernai, Frankreich, statt. Nach einer kurzen Kennenlernrunde am ersten Tag wurde den 16 Teilnehmern aus 8 SPPEXA-Projekten das Programmieren von Beschleuniger-Hardware näher gebracht. Hierzu fanden zwei 1.5-Tage-Tutorials zu GPU- und zu Xeon-Phi-Programmierung statt, in denen die Teilnehmer die Gelegenheit bekamen, verschiedene entsprechende Programmiersätze kennenzulernen, direkt auszuprobieren und anzuwenden.

In den Pausen wurde das Klischee des schlemmenden Franzosen in Form von umfangreichen Mittags- und Abendmenüs bedient. Am Mittwoch, 7. September, ging es für einen Nachmittag ins nahe gelegene Straßburg zum gemütlichen Bummel durch die Altstadt.



Die SPPEXA-Koordination bedankt sich an dieser Stelle sowohl bei den Dozenten Dominik Göddeke und Karl Furlinger sowie ihren Helfern Malte Schirwon und Tobias Fuchs für die gelungenen Tutorials als auch bei den Teilnehmern für die aktive Teilnahme, und beim LRZ für den Zugang zum SuperMIC. À bientôt!

Philipp Neumann



Figure 3: Participants of the scientific workshop “High Performance Computing for Water-Related Hazards” at LRZ, June 29 – July 1, 2016.

Acknowledgements:

This article will appear in the autumn 2016 edition of inSiDE (Innovatives Supercomputing in Deutschland).

The Czech-Bavarian Competence Centre for Supercomputing Applications is funded by the Federal Ministry of Education and Research. The Intel MIC programming workshops was financially also supported by the PRACE 4-IP project funded by the European Commission’s Horizon 2020 research and innovation programme (2014-2020) under grant agreement 653838.

The CzeBaCCA workshop were organized by Michael Bader (TUM), Momme Allalen, Fabio Baruffa, Arndt Bode, Anton Frank, Volker Weinberg (all Leibniz Supercomputing Centre), Ondrej Jakl, Branislav Jansik, Martin Palkovic, Vit Vondrak (IT4Innovations).

References

- <https://www.lrz.de/forschung/projekte/forschung-hpc/CzeBaCCA/>
- <http://www.lrz.de/services/compute/courses/>
- <http://www.training.prace-ri.eu/>

“Intel MIC Programming” & “High Performance Computing for Water-Related Hazards” Workshops@LRZ

To foster the Czech-German collaboration in high performance computing, the Leibniz Supercomputing Centre (LRZ), the Department of Informatics of TUM and the National Supercomputing Centre of the Czech Republic, IT4Innovations, recently joined forces and established the Czech-Bavarian Competence Centre for Supercomputing Applications (CzeBaCCA). IT4Innovations at Ostrava is hosting the Salomon supercomputer, the largest Intel Xeon Phi based system currently operating in Europe. Besides their joint research program around simulation software and tools for Salomon, one of the main objectives of the new Competence Centre is to organise a series of scientific workshops and Intel MIC (Many Integrated Core) architecture specific trainings.

The first series of workshops took place in Ostrava in February 2016 and combined a two-day “Intel MIC Programming Workshop” with a one-day scientific workshop “SeisMIC – Seismic Simulation on Current and Future Supercomputers” (see inSiDE Vol. 14 No. 1 p. 76f).

The second series of workshops took place at LRZ in Garching in June 2016 and combined a three-day “Intel MIC Programming Workshop” (June 27 – 29, 2016) with a three-day scientific workshop on “High Performance Computing for Water-Related Hazards” (June 29 – July 1, 2016). To create synergy effects between the different communities, both events shared a plenum session about Intel MIC experience on June 29.

The Intel MIC Programming Workshop attracted 50 participants from the Czech Republic, Finland, Germany, Italy, Spain and UK. The first two days provided an introduction about the Intel Knights Corner architecture and various Intel Xeon Phi programming models. The talks by LRZ staff members were interleaved with many hands-on sessions. The hands-on sessions were done on the SuperMIC system at LRZ, which was exclusively reserved for the workshop. On the last day Dr.-Ing. Jan Eitzinger from the Regional Computing Centre in Erlangen (RRZE) presented advanced MIC

programming techniques such as SIMD or assembly programming. Mr. Andrey Semin from Intel presented the newest version of the Intel Xeon Phi Knights Landing architecture, which has been launched on the International Supercomputing Conference (ISC) in Frankfurt just one week before the workshop. In a plenum session 7 invited speakers from Intel, RRZE, IPP, IT4Innovations, TUM and LRZ talked about experiences and best practices using Intel Xeon Phi based systems like e.g. Salomon (IT4I), HELIOS (Japan) or SuperMIC (LRZ). A guided tour of the Weihenstephan brewery, the oldest still-operating brewery in the world, and a self-paid dinner at the brewery restaurant, has been organised as social event.

The Intel MIC Programming Workshop was followed by a three-day scientific workshop about “High Performance Computing for Water-Related Hazards” with 44 participants from the Czech Republic, Finland, Germany, Hungary, Italy, Netherlands, Norway, UK and USA. The focus of this workshop was on the simulation of environmental catastrophes caused by water. The researchers discussed various aspects of water related natural disasters: online prediction systems for floods, rain-induced flooding, simulation of storm surges and tsunamis, rise of the sea level by melting glaciers and similar events. A special focus was on demands and desired features of (future) simulation software, parallelisation for current and novel HPC platforms, as well as establishing scalable simulation workflows on supercomputing environments. A special session on “Flood Simulation in Bavaria” with invited speakers from the research areas hydrology, regional climatology, earth and environmental sciences concluded the workshop. With the recent floods in Bavaria the workshop gained unexpected relevance. The next series of workshops will take place at IT4Innovations in early 2017. Planned is an Intel MIC workshop concentrating on simulations on the Salomon machine combined with a scientific workshop on “HPC for Atmospheric Models and Air-Related Environmental Issues”.

Volker Weinberg, LRZ



Figure 1: Participants of the “Intel MIC Programming Workshop” at LRZ, June 27–29, 2016. The main speakers in the first row are holding Intel Xeon Phi coprocessors.



Figure 2: Participants enjoying the guided tour of the Weihenstephan brewery during the social event.