J. Indian Inst. Sci., 2002, 82, 63–70
© Indian Institute of Science.

BOOK REVIEWS

Understanding the digital economy edited by Erik Brynjolfsson and Brian Kahn, The MIT Press, 5, Cambridge Centre, Cambridge, MA 02142, USA, 2002, pp. 401, $22.95.

The book under review is a contemporary research collection of scholarly papers in the emerging area of digital economy. Evolved out of the conference organized in May 1999 by the US Department of Commerce, the book addresses the following four major issues

- Macro economic perspective
- Market structure
- Employment & workforce
- Organizational change

through 14 articles written by well-known experts in this area (Erik Brynjolfsson and Warda Orlikowski of MIT Sloan, Hal Varian of UC Berkeley, Paul David of Oxford University, Kathleen Carley of Carnegie-Mellon, Andrew Whinston of the University of Texas at Austin) and others. Coming at a time after the dot.com bust and 'tech meltdown', the research questions raised in this volume are timely, pertinent and deep. With the hype gone, one hopes that a set of scholarly professionals would address the right questions, conduct deep research and help the humanity at large to steer the evolving digital economy in the right direction.

The Introduction sets the stage by defining *information economy* (the trend towards the dominance of information and knowledge-based assets relative to tangible assets associated with agriculture and manufacturing) and *digital economy* (transformation of all sectors of the economy through digitisation–the recent and still largely unrealised transformation) and addresses the following sets of questions

- How should we identify and measure digital economy?
- What are the implications for employment and productivity?
- What are the economic characteristics of digital products?
- What are the key determinants of prices, market structure and competition in digital economy?
- How reliable are current methods for projecting the size and composition of labour markets?
- What barriers impede the diffusion of e-commerce across the society?
- To what extent and under what conditions will a digital economy lead to new organizational cultures?

The authors have brought out a wealth of interesting questions—the answers will be found over time only. The famous 'productivity paradox'—you can see the computer age everywhere but in the productivity statistics—brought out originally by the Nobel Laureate Solow is related to serious measurement problems. The authors then move on to the question of data needs with respect to IT infrastructure, e-commerce quantification, firm and industry structure, worker characteristics and price behaviour. They also indicate useful suggestions to Census Bureau and other statistical organizations. It has a telling message for India, though IT is much talked about, the macro and micro economic measurements relating to IT industry in India leave much to be desired. Several other challenges of measurements of digital economy are brought out well, for example, if you receive a newspaper online, is it a 'good' or a 'service' ?

Paul David in his paper powerfully brings home the fact that major technology like IT requires the development and coordination of a vast array of complementary elements—new plant and equipment, new workforce skills, new organizational forms, new forms of legal property and new mindsets—to bring in significant productivity increases that typically require decades rather than years.

The author has drawn a very interesting and deep parallel to the electrification of industry through the years 1899–1914 to the impact of IT across all sectors of the economy through the years 1979–1997.

Another interesting point that once again takes us back to the measurement question is the fact that the availability of services round the clock (the 24 x 7 jargon in IT industry) leads to significant benefit to the consumers (a bank branch is open all the time); though current economic measures do not capture this large societal benefit.

Yet another point forcefully brought out by the author is the 7.0% per annum rate of decline in electric power over the period 1899–1948; a fact that is significant and force many IT professionals to think hard about the 'uniqueness' of IT industry often quantified through Moore's Law and the resultant falling cost of computing power.

Yet another interesting research question is the nature of information market and the impact of Internet with its ability to deliver, not merely support the process of delivery through a set of transactions—digital goods. In fact, it may be the most interesting development to watch.

The authors lament that intuition, trial and error, and venture capital can sometime substitute for general understanding, but not always; much could be learnt from well-designed research. It is a telling story of the state of research in this area.

Hal Varian in his characteristic style puts forth his view "technology changes, economic laws do not" underlining the fact that many of the fundamental economic principles will be as relevant to e-commerce as it is commerce.

Wanda Orlikowski in her piece puts it more bluntly "to treat digital economy as if it were an external, independent, inevitable phenomenon "out there" is inappropriate and misleading—it is a product of our own making. We have the opportunity, the challenge, and perhaps more importantly, the responsibility to shape the phenomenon. There is a well-known saying that the best way to predict the future is to invent it. As such collectively we are in the process of inventing the digital economy".

The set of papers in the book are excellent pieces that address in their own ways this emerging phenomenon. Though a set of coincidences, India has emerged as a 'thought leader' in several areas of IT. It is now the opportunity for economists and other social scientists in India to get actively involved in research into this study of digital economy and actively contribute to this emerging area. Undoubtedly, a careful study of the different articles in this book will contribute to such an understanding.

Director
Indian Institute of Information Technology
Bangalore 560 066
(ss@iiitb.ac.in)

S. Sadagopan

Principles of data mining by David J. Hand, Heikki Mannila and Padhraic Smyth, The MIT Press, Cambridge Centre, Cambridge, MA 02142, 2001, pp. 546, $50.

The field of data mining is attracting the attention of engineers, scientists, and business people due to its application potential in a variety of important areas including internet science and engineering, medical sciences, bioinformatics, forecasting, large-scale pattern recognition, and knowledge discovery in databases. It can have a positive impact on every large-scale decision-making endeavor. There are not many good books on data mining; it is because the area is still in a state of infancy. Also, there is no standard to assess whether some activity can be classified as data mining activity or not. In this context, a comprehensive book on data mining is highly welcome.

This book offers a wealth of information on a variety of topics coming under the broad umbrella of data mining. It deals with both the descriptive and predictive data-mining paradigms. It is an authoritative source for researchers and practitioners in the area of statistical pattern recognition. Even though the authors recommend it for using as a textbook for teaching senior undergraduate and junior graduate-level courses, it will be necessary for the instructor to provide details missing in the book; work out additional examples and implementation details. A major difficulty in using the book as a teaching resource is that it deals with most of the material at a conceptual level. However, the reader can access the missing details through the valuable pointers, provided at the end of each chapter in the book, to the literature on data mining.

Material in the book is very well structured into 14 chapters dealing with various aspects of data mining, an appendix on random variables, and an excellent collection of references. Chapters 1–4 deal with fundamental aspects of data and data analysis to provide an excellent platform for describing and analyzing statistical data-mining components, tasks and algorithms in the rest of the book. The first four chapters are written in an authoritative and a reasonably thorough manner. However, there are occasions when clarity is missing. For example, Section 2.3 deals with distance measures and the notion of metric as dissimilarity measure is introduced here; however, it is not clear why the dissimilarity between two objects should be a metric. There are several dissimilarity functions that do not satisfy metric properties and yet are successfully used in practice; the notion of mutual neighborhood value (MNV) discussed by Jain and Dubes (reference Jain and Dubes (1988) in the book) is one such function.

Chapters 5–8 describe various building blocks that can be used to synthesize and analyze data-mining algorithms. Several important issues including 'Curse of dimensionality', 'Over-fitting', and 'Bias-variance trade-off' are very well explained in these chapters. They are dealt with such a high degree of clarity that the reader will get an excellent understanding of these important notions. However, notions like 'Score function' can be confusing; the 'classification accuracy of a classifier', 'angle between two vectors' in the vector space model, and 'sum of squared errors' are all score functions. Chapters 9–14 deal with specific data-mining tasks like clustering, classification, regression, and association rule and pattern discovery. Chapter 14 titled 'Retrieval by content' is an important application but not a basic data-mining task as categorized by the authors in the preface.

On the whole, the book provides very good references to recent literature on data mining. Also, it deals with several important building blocks and data-mining tasks. It clearly brings out the differences between machine learning and database communities in terms of the nature of their contributions to data mining. A major problem with the presentation is that there are too many forward references; for example, the term 'over fitting' is used on pages 19 and 150 and is explained on page 222. Similarly, 'cross-validation' is used on page 148 (Chapter 5) and is detailed in Chapter 7. Terms like 'NP-Hard' and 'regularization' are used without pointers to the related literature. Also, there are some syntactic mistakes; for example, on page 162, in the second paragraph, in the second sentence 'Once the use vector representation' should be read as 'Once the use of vector representation'.

There are some semantic problems also. For example, the expression for Mahalanobis distance given on page 276 corresponds to squared Mahalanobis distance. On page 287, it is stated that "kernel models are really practical for relatively low-dimensional problems"; however, support vector machines are based on kernel models and they can deal with high-dimensional data sets. On page 301, in the second paragraph, it is stated that "for partition-based methods, however, it is necessary to decide at the start how many clusters we want"; there are partition-based clustering techniques like ISODATA (reference Hall and Ball (1965) in the book) that do not require the value of number of clusters.

This book appears to be an outcome of an ambitious plan of dealing with a very wide variety of topics; the authors have successfully achieved it. In the process, some important contributions in the literature are either ignored or are not dealt with in detail. For example, Thomas Dietterich, who wrote the Series Foreword to the book, has identified four important topics in machine learning in his paper "Machine learning research: Four current directions" (*AI Mag.*, 1997, **18**(4), 97–136). Of these four, the current book did not discuss reinforcement learning; ensembles of classifiers is discussed in a sketchy manner. Also, one of the important topics in machine learning, namely, support vector machines, is not discussed in detail. From a pragmatic viewpoint, it could have been beneficial to the reader to have seen linear time algorithms for clustering in the book; there are several linear time clustering algorithms in the literature and some of them should have been discussed instead of describing quadratic time algorithms. One of the important foundational topics for descriptive data mining is mathematical logic. There are several knowledge-based pattern recognition methods based on logic and they do not find a place in the book. Expectation maximization algorithms are discussed at length, but they do not offer a pragmatic appeal to the data-mining community.

In summary, this book is of good quality and is strongly recommended as an excellent reference material for people interested in statistical pattern recognition, machine learning and data mining.

Professor
Department of Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.
Tel: +91-80-3092779
email: mnm@csa.iisc.ernet.in

M. Narasimha Murty

Computational methods for representations of groups and algebras (Progress in Mathematics, Vol. 173) edited by P. Dräxler, G. O. Michler, C. M. Ringel, Birkhaüser Verlag AG, Klösterberg 23, CH-4010 Basel, Switzerland, 1999, pp. 376, sfr.108.

This is a collection of survey articles presented at a conference held at the University of Essen during April 1–4, 1997. This conference was the culmination of a joint research project of algebraists from the Universities of Antwerp, Bielefeld, Essen, Leeds, Paris VI and Trondheim on 'Invariants and representations of Algebras' from 1991 to 1997. The objective of these surveys is to provide "an informative account of the tremendous possibilities of computational methods in the representation theory of groups and algebras".

It has three introductory articles on classification problems in representation theory of finite-dimensional algebras by Peter Dräxler and Rainer Nörenberg, Non-commutative Gröbner bases and projective resolutions by Edward L. Green and Construction of finite matrix groups by Robert A. Wilson.

This is followed by 16 keynote articles of a more specialized nature on diverse topics: Presentations for Lyon's simple group (one article by Holger W. Gollan and George Havas and another by George Havas and Charles C. Sims), An existence proof of Janko's simple group J_4 (by G. D. Cooperman, W. Lempken, G. O. Michler and M. Weller), 2-Modular decomposition numbers of the Conway group Co_2 (by Jürgen Müller and Jens Rosenboom), Computational problems in the theory of Kazhdan–Lusztig polynomials and coxeter groups (by Fokko du Cloux), Calculations of group cohomologies (by Jon F. Carlson), Computational approach to decide the equivalence of certain classes of algebras to tubular algebras (by M. Barot and J.A. de la Peña), Non-commutative Gröbner bases and Anick's resolution (by Svetlana Cojocaru, Alexander Podoplelov and Victor Ufnarovski), A new algorithm to compute the normalization $\bar{R}$ (i.e., the integral closure of a ring R in its total ring of quotients) of an affine ring R (by Wolfram Decker, Theo de Jong, Gert-Martin Greuel and Gerhard Pfister), A computer algebra approach to sheaves over weighted projective lines (by Piotr Dowbor and Thomas Hübner), Relative trace ideals and Cohen Macanlay quotients (by Peter Fleischmann), Reduction of weakly definite integral quadratic forms (Hans-Joachim von Höhne), Decision problems in finitely presented groups (by Derek F. Holt), Some algorithms in invariant theory of finite groups (by Gregor Kemper and Allan Steel), Coxeter transformations associated with finite dimensional algebras (by Helmut Lenzing) and Bimodules and matrix problems (by Serge Ovsienko).

All the articles here are carefully written and refereed. The emphasis of each of them is on computational aspect that can be implemented on a machine and so should be of interest to both algebraists and computer scientists interested in these aspects. These, together with introductory lectures by H. Gollan, are hoped to present the state of the art in the subject. The content of the lectures of H. Gollan is available in the lecture notes Linear algebra over small finite fields on parallel machines, by P. Fleischman *et al.*, *Vorlesungen Fachbereich Math.*, Univ. of Essen, 23 (1995).

Indian Statistical Institute
Bangalore 560 059.

N. S. Narasimha Sastry

Abstract root subgroups and simple groups of Lie-type by Franz Georg Timmesfeld, Birkhauser Verlag AG, Klösterberg 23, CH-4010 Basel, Switzerland, 2001, pp.389, sFr.168.

In 1955, Chevalley classified simple algebraic groups over an arbitrary algebraically closed ground field by proving that they were obtained by base change from certain Chevalley group (scheme)s over **Z**. He proved further that the group or rational points of these group schemes over a finite field were simple groups (modulo their centres). Thus he produced remarkable and new examples of finite simple groups (Chevally groups corresponding to exceptional Lie groups, and the Ree–Suzuki groups). These are called simple of Lie type. They are generalization of groups of the form $SL_n(\mathbf{F}_q)$ (mod centre), where $\mathbf{F}_q$ is the finite field with q elements.

We mention here that according to the classification of finite simple groups, apart from alternating groups and the groups considered by Chevally (including the Ree–Suzuki groups), there are only finitely many finite simple groups; the latter are called sporadic groups.

In 1964, Tits proved a beautiful generalization of the foregoing result of Chevalley. Let K be any field and let G be a simple connected simple algebraic group defined and isotropic over the field K. Here, 'isotropic' means that there exists a proper parabolic subgroup P defined over the field K. Assume P is minimal with this property, and let U^+ denote its unipotent radical. Let U^- be an unipotent subgroup which is conjugate to U^+ by an element of $G(K)$ and is opposed to U^+, i.e. $U^- \cap U^+ = \{1\}$. For example, if $G = SL_n$ then U^+ (resp. U^-) may be taken to be the group of unipotent upper (resp. lower) triangular matrices.

The theorem of Tits says that the quotient of $G(K)$ by its centre is an abstract simple group. This generalizes the Theorem of Chevalley.

In proving this theorem, Tits made, as did Chevalley, a detailed study of root subgroups, corresponding to the root system of G with respect to a maximal split torus T and commutator relations between them (in the case of Chevalley groups, these are called the Chevalley commutator relations). Again, in the case of SL_n, the root groups correspond to the groups $\{1 + tE_{ij} : t \in K, 1 \le i \ne j \le n\}$. One may thus say that the finite group of Lie type is described in terms of (root subgroups which lie in) a rank one group, namely, $SL(2)$ or $SU(2,1)$ (the group $SU(2,1)$ may be defined for an arbitrary field with a quadratic Galois extension).

While analyzing the structure of arbitrary finite simple groups, it was found that (by several authors like Fischer, Aschbacher, Timmesfeld himself) an important class of finite simple groups (not just the Lie-type ones) were generated by certain subgroups which corresponded to such root subgroups. Thus it seems worthwhile to axiomatise the notion of abstract root groups. For this, one must first introduce the notion of an abstract rank one group, in terms of which the whole group may be described by generators and relations.

In the book under review, a theory of abstract root subgroups is developed for an arbitrary simple group of Lie type or classical type (not only finite simple ones). Much of the book is an exposition of the research of the author himself, and includes theorems and proofs which are not easily available elsewhere.

The first chapter deals with rank one groups, the building blocks of the groups under consideration. In the case of split Lie type groups over finite fields, these are just $SL(2)$. Otherwise (for other Lie-type groups over finite fields), the rank one groups are $SU(2,1)$.

Here is the definition of an abstract rank one group. *A* group *X* generated by two different nilpotent subgroups *A* and *B* satisfying

for each $a \in A - (1)$ *there exists an element* $b \in B-(1)$

such that $b^{-1} Ab = a^{-1} Ba$ *and vice versa*

is called a rank one group.

The conjugates of *A* and *B* will be called unipotent groups, and conjugates of $H = N_X(A) \cap N_X(B)$ will be called diagonal groups.

The second chapter defines a set of abstract root subgroups of a group *G*. This is a set Σ of abelian nonidentity subgroups of *G* with the following properties.

(i) *G* is generated by elements of Σ. Moreover, for each $S \in \Sigma$ its conjugate $g^{-1}Sg$ lies in Σ for all $g \in G$.

(ii) For each pair A, $B \in \Sigma$ one of the following conditions hold:

(α) *A* and *B* commute.

(β) The group *X* generated by *A* and *B* is a rank one subgroup with unipotent subgroups *A* and *B*.

(γ) The centre $Z(< A, B >)$ of the group generated by *A* and *B* contains the commutator subgroup $[A, B]$ which equals group generated by the commutator sets $[a, B] = [A, b]$ for all $a \in A-(1)$ and $b \in B-(1)$.

In Chapter Two, it is also shown that simple groups of Lie type are generated by a conjugacy class Σ which is a set of abstract root subgroups. The results are largely due to the author.

Chapter Three is the heart of the book. In it, the author gives the classification of simple groups generated by abstract root subgroups. Certain geometries corresponding to such groups are studied and shown that the flag complex of these geometries is a spherical building.

Finally, in Chapter Five, the author gives a number of applications.

In conclusion, this is a beautiful book, written by one of the main contributors to the subject. The reviewer found that the material is presented in a way that is enjoyable and very easy to read. There are a number of exercises at the end of each chapter, giving many interesting applications. Mathematicians interested in finite groups or buildings, or infinite (simple) subgroups of Lie groups will find it very useful.

School of Mathematics
Tata Institute of Fundamental Research
Mumbai 400 005, India
email: venky@math.tifr.res.in

T. N. Venkataramana

Published by Prof. P. Venkataram, Executive Editor, *Journal of the Indian Institute of Science*, Bangalore 560 012; Printed at Tholasi Prints India Pvt. Ltd., Bangalore 560 020.