Application Areas of Combinatorics, Especially Permutations and Combinations

1. Introduction

Combinatorics, or combinatorial theory, is a major mathematics branch that has extensive applications in many fields such as engineering (e.g., pattern such as image analyses, communication networks), computer science (e.g., languages, graphs, intelligent computing), natural and social sciences, biomedicine (e.g., molecular biology), operations research (e.g., transportation, scheduling), and business (Liu, 1968; Roberts, 198; *Combinatorial*, 1992-2005; Mycielski, et al., 1997). The most common scenario is that many real world problems are mathematically intractable. In these cases, combinatorics techniques are needed to count, enumerate, or represent possible solutions in the process of solving application problems. Generation of combinatorial sequences, such as permutations and combinations, has been studied extensively because of the fundamental nature and the importance in practical applications. There has been interest in the generation of these sequences in a parallel or distributed computing environment (Akl, 1989; Kapralski, 1993).

2. Application Areas of Combinatorics

o Communication networks, cryptography and network security
Permutations are frequently used in communication networks and parallel and distributed systems (Massini, 2003; Yang and Wang, 2004). Routing different permutations on a network for performance evaluation is a common problem in these fields. Many communication networks require secure transfer of information, which drives development in cryptography and network security (Kaufman, et al., 2003; Stallings, 2003). This area has recently become particularly significant because of the increased use of internet information transfers. Associated problems include protecting the privacy of transactions and other confidential data transfers and preserving the network security from attacks by viruses and hackers. Encryption process involves manipulations of sequences of codes such as digits, characters, and words. Hence, they are closely related to combinatorics, possibly with intelligent encryption process. For example, one common type of encryption process is interchanging--i.e., permuting parts of a sequence (Nandi, et al., 1994**)**. Permutations of fast Fourier transforms are employed in speech encryption (Borujeni, 2000).

o Computer architecture
Design of computer chips involves consideration of possible permutations of input to output pins. Field-programmable interconnection chips provide user programmable interconnection for a desired permutation (Bhatia and Haralambides, 2000). Arrangement of logic gates is a basic element for computer architecture design (Tanenbaum, 1999).

o Computational molecular biology
This field involves many types of combinatorial and sequencing problems such as atoms, molecules, DNAs, genes, and proteins (*Combinatorial*, 1992-2005; Doerge and Churchill, 1996; Chiang and Eisen, 2001; Siepel, 2003). One-dimensional sequencing problems are essentially permutation problems under certain constraints.

o Languages
Both natural and computer languages are closely related to combinatorics (*Combinatorial*, 1992-2005; Kotz, 1995). This is because the components of these languages, such as sentences, paragraphs, programs, and blocks, are arrangements of smaller elements, such as words, characters, and atoms. For example, a string searching algorithm may rely on combinatorics of words and characters. Direct applications of this can include word processing and databases. Another important application area is performance analysis of these string searching algorithms. The study of computability--what we can compute and how it is accomplished--draws heavily on combinatorics.

o Pattern analysis
In a broad sense, all the above-mentioned areas can be viewed as special cases of pattern analysis. Molecular biology, for example, studies patterns of atoms, molecules, and DNAs whereas languages treat patterns of sentences, words, and strings. Patterns can have many other forms; for example, visual images, acoustic signals, and other physical quantities such as electrical, pressure, temperature, etc., that appear in engineering problems. Patterns can also be abstract without any associated physical meaning. These patterns may be represented in various ways such as digital, analog, and other units. Some of these types of patterns can be associated with combinatorics. There has been extensive research on combinatorial pattern matching (*Combinatorial*, 1992-2005). Computer music can be a specialized application domain of combinatorics of acoustic signals.

o Scientific discovery
For certain types of knowledge discovery problems, generation of combinatorial sequences may become necessary in the process of yielding candidate solutions. For example, in scientific discovery, we may want to have a sequence of plausible chemical/biological reactions and their formations (Valdes-Perez, 1999). In each step of the sequence, we may generate combinatorial sequences of chemical/biological radicals, bases, and molecular compounds as candidate solutions and may select the most likely ones under certain rules and constraints. In another example, certain areas of mathematics, such as graph theory and number theory, may generate combinatorial sequences as candidate solutions.

o Databases and data mining
Queries in databases are multiple join operations that are permutations of the constituent join operations. Determining an optimal permutation that gives minimum cost is a common and important problem (Kumar Verma, and Trimbak Tamhankar, 1997). Data mining or knowledge discovery in databases is a relatively new field that aims at distilling useful information, often from large databases. In this area, techniques employing symbolic AI can manipulate combinatorial sequences of atoms or information elements. Non-symbolic knowledge discovery techniques, such as genetic algorithms and genetic programming, most commonly deal with solutions in the form of sequences of bits, digits, characters, and sometimes Lisp program elements. Neural networks, another domain of non-symbolic AI, sometimes deal with combinatorial patterns. Knowledge discovery techniques under uncertainty, such as Bayesian networks, Dempster-Shafer theory, fuzzy logic, and rough set theory, may have combinatorial solutions (Munakata, 1998).

o Operations research
Many optimization problems in operations research (OR) involve combinatorics. The job scheduling problem is essentially a sequencing problem to determine the order of jobs to be processed in an effort to minimize the total time, cost, etc. Here, jobs can be in a computer system, network, or processing plant. Many problems involving graphs or networks also deal with the order of vertices and edges. The traveling salesperson problem is to determine the order of cities to be visited to minimize the total distance (Matsumoto and Yashiki, 1999). The shortest path problem of a graph is to determine a sequence of edges, the total length of which is minimum. Oftentimes, these problems are computationally difficult--e.g., NP-complete or NP-hard--and, therefore, require extensive research.

o Simulation
Permutations and combinations can be employed for simulations in many areas. Permutations representing various genotype-phenotype associations are employed in genetics simulations (Doerge and Churchill, 1996). Other areas that employ permutations and combinations for simulations include networks, cryptography, databases and OR.

o Homeland security
This is a very specialized problem domain that has become a major national challenge after 9/11. To confront this challenge, many intelligent computing techniques have been applied, including intelligent pattern analyses of human faces, X-ray images, chemical components, data from a distributed network of wireless sensors, etc. Natural language processing and data mining techniques have been applied to sift through and monitor the tremendous

accumulation of electronic communication data. Since combinatorics are extensively applied to these intelligent computing techniques, there is a wide spectrum of potentials for the national security issue. Some specific examples may include string searching algorithms and their performance analysis in communication data, pre- and post-analysis of combinatorial sequences of information elements, and combinatorial pattern matching.

References

Akl, S.G. 1989. *The Design and Analysis of Parallel Algorithms*, Englewood Cliffs, New Jersey, Prentice-Hall, Chapter 6.

Bhatia D. and J. Haralambides. 2000. Resource requirements and layouts for field programmable interconnection chips, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 8(3): 346-55.

Borujeni. S.E. 2000. Speech encryption based on fast Fourier transform permutation, *Proceedings of the 7th IEEE International Conference on Electronics, Circuits and Systems* (ICECS 2000), Dec. 17-20, 290-293, Jounieh, Lebanon, IEEE.

Chiang D.Y. and M. Eisen. 2001. Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles, *Bioinformatics*, 17(S1): 49-55.

Combinatorial Pattern Matching; Lecture Notes in Computer Science. Proceedings of Annual Symposiums. 1992-2005. Berlin, Springer-Verlag.

Doerge R.W. and G. A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character, *Genetics*, 142: 285-294.

Kapralski, A. 1993. New methods for generation of permutations, combinations, and other combinatorial objects in parallel, *Journal of Parallel and Distributed Computing* 17: 315-326.

Kaufman, C., R. Perlman and M. Speciner. 2003. *Network Security*: *Private Communication in a Public World*, 2nd Ed., Upper Saddle River, NJ, Prentice Hall.

Kotz. D. 1995. A data-parallel programming library for education (DAPPLE), *SIGCSE Bulletin*, 27(1): 76-81.

Kumar Verma, A. and M. Trimbak Tamhankar. 1997. Reliability-based optimal task-allocation in distributed-database management systems, *IEEE Transactions on Reliability*, 46(4): 452-459.

Liu, C.L. 1968. *Introduction to Combinatorial Mathematics*, Computer Science Series, New York, McGraw-Hill, Chapter 1.

Massini. A. 2003. All-to-all personalized communication on multistage interconnection networks, *Discrete Applied Mathematics*, 128(2-3): 435-46.

Matsumoto N. and S. Yashiki. 1999. Simple approach to TSP by permutation of six cities and deletion of crossover, *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Aug. 22-24, 377-380, Victoria, BC, Canada, IEEE.

Munakata. T. 1998. *Fundamentals of the New Artificial Intelligence: Beyond Traditional Paradigms*, New York, Springer-Verlag.

Mycielski, J. et al. (Eds.), 1997. *Structure in Logic and Computer Science*, Berlin, Springer-Verlag.

Nandi, S., B.K. Kar, and P. Pal Chaudhuri. 1994. Theory and applications of cellular automata in cryptography, *IEEE Transactions on Computers*, 43(12): 1346-1357.

Roberts, F.S. 1984. *Applied Combinatorics*, Englewood Cliffs, New Jersey, Prentice-Hall.

Siepel, A.C. 2003. An algorithm to enumerate sorting reversals for signed permutations, *Journal of Computational Biology*, 10(3-4): 575-597.

Stallings, W. 2003. *Cryptography and Network Security*: *Principles and Practice*, Upper Saddle River, NJ, Prentice Hall.

Tanenbaum, A.S. 1999. *Structured Computer Organization*, 4th Ed., Upper Saddle River, NJ, Prentice Hall.

Valdes-Perez, R.E. 1999. Discovery techniques for scientific apps, *Communications of the ACM*, 42(11): 37-41.

Yang Y. and J Wang. 2004. Routing permutations on optical baseline networks with node-disjoint paths, *Proceedings of Tenth International Conference on Parallel and Distributed Systems* (ICPADS 2004), July 7-9, N.- Tzeng, F. (Ed), 65-72, Newport Beach, CA, IEEE.