Complete Sets of Hamiltonian Circuits for Classification of Documents

Bernd Steinbach[1] and Christian Posthoff[2]

[1] Freiberg University of Mining and Technology, Institute of Computer Science, D-09596 Freiberg, Germany
[2] The University of The West Indies, St. Augustine Campus, Trinidad & Tobago

Abstract. The calculation of Hamiltonian Circuits is an NP-complete task. This paper uses slightly modified complete sets of Hamiltonian circuits for the classification of documents. The solution method is based on a SAT-instance with a huge number of clauses which is flattening the knowledge about the problem. We suggest an even more compact model of Boolean equations that preserves the knowledge by summarizing restrictions and requirements. The presented *implicit two-phase SAT-solver* finds efficiently the solution using operations of the XBOOLE library. This solver can be included easily as signal processing unit into the device where the classification of the documents is required.

1 Introduction

A Hamiltonian path in a graph is a sequence of edges that uses each node precisely once. A Hamiltonian circuit [1], also called Hamiltonian cycle, is a cycle in the graph which visits each node exactly once and returns to the starting node. It is well-known that the problem of determining whether such paths or cycles exist is an NP-complete problem [3].

We suggest to modify this problem slightly and apply it to the classification of documents by means of an efficient signal processing unit. Basically our approach can be applied to each document that is transmitted by a sequence of bits. As an example we motivate our approach in the context of FAX transmissions. In spite of several new possibilities the FAX transmission is popular for the fast exchange of textual and graphic date because only simple equipment is required. A disadvantage of the FAX transmission from Alice to Bob is that Eve (an eavesdropper) can see the transmitted information. An easy way to reduce the value of the information seen be Eve is that both *good* and *bad* information is transmitted by FAX. It is necessary that the receiver Bob can select the *good* FAX, and a signal processing unit throws away the *bad* information immediately.

A couple of bits of the transmitted document is used for this detection. These bits may form an $n \times n$ adjacency matrix of a graph. Reflexive edges that begin and end on the same node are ignored. Instead of a single Hamiltonian circuit a set of Hamiltonian circuits will now be is allowed. A complete set of Hamiltonian circuits covers all nodes of the graph. Any received FAX without a complete set of Hamiltonian circuits will be rejected. The number of different complete sets

of Hamiltonian circuits is an additional information for the receiver Bob, but hidden for Eve. Due to the missing number of nodes n and the position of the classification bits in the transmitted FAX the eavesdropper Eve is not able to distinguish between *good* and *bad* FAX transmissions.

2 Boolean Model

Boolean variables are introduced for each edge and each direction as follows:

$$x_{ik} = \begin{cases} 1 & \text{if the edge is used from node } i \text{ to node } k \\ 0 & \text{otherwise} \end{cases}$$

Hence, the number of variables needed to express all conditions of complete sets of Hamiltonian circuits is equal to the number of values 1 in the $n \times n$ adjacency matrix of a graph where values 1 on the main diagonal are replaced be values 0.

The use of the edge from node i to node j, i.e. $(x_{ij} = 1)$ implies $(x_{ji} = 0)$, as next $(x_{id_l} = 0)$, and finally $(x_{s_m j} = 0)$. The solution of this system of equations is a partial solution set that can be stored in a single ternary vector [2].

3 Two Phase SAT Solver

In this extended abstract we can give only a very restricted sketch of our very efficient algorithm. In the first phase partial solution vectors for the restrictions of each given edge are generated based on the Boolean model described above. These stored partial solution sets are used in the second phase such that the requirement *there must be for each node exactly one outgoing edge* is fulfilled.

4 Experimental Results

A small adjacency matrix of 6×6 bits with 12 values 1, for instance, is sufficient to sign a FAX as *good* or *bad* and deliver additional classification information. The power of such a signal processing unit for classification becomes visible when during the time of receiving a FAX this unit could find all existing 185,868 complete sets of Hamiltonian circuits in an adjacency matrix of 36×36 with 154 values 1.

References

1. Ore, O.: A note on Hamiltonian Circuits. Amer. Math. Monthly. **67** (1960), 55.
2. Posthoff, Ch., Steinbach, B.: Logic Functions and Equations - Binary Models for Computer Science. Springer, Dordrecht, The Netherlands, 2004.
3. Wegener, I.: Complexity Theory - Exploring the Limits of Efficient Algorithms. Springer, Dordrecht, The Netherlands, 2005.

This article was processed using the LaTeX macro package with LLNCS style