Chapter 18
An Efficient Algorithm for Automating Classification of Chemical Reactions into Classes in Ugi's Reaction Scheme

Sanjay Ram
Bengal Engineering and Science University, Shibpur, India

Somnath Pal
Bengal Engineering and Science University, Shibpur, India

ABSTRACT

There are two approaches for classification of chemical reactions: Model-Driven and Data-Driven. In this paper, the authors develop an efficient algorithm based on a model-driven approach developed by Ugi and co-workers for classification of chemical reactions. The authors' algorithm takes reaction matrix of a chemical reaction as input and generates its appropriate class as output. Reaction matrices being symmetric, matrix implementation of Ugi's scheme using upper/lower tri-angular matrix is of $O(n^2)$ in terms of space complexity. Time complexity of similar matrix implementation is $O(n^4)$, both in worst case as well as in average case. The proposed algorithm uses two fixed size look-up tables in a novel way and requires constant space complexity. Time complexity both in worst and average cases of the algorithm is linear.

INTRODUCTION

One of the major tasks of chemoinformatics is to predict the outcome of a chemical reaction. For this purpose the first step is to classify existing chemical reactions, on the basis of which unknown reactions can be predicted. In the mid 1970s, Ugi and his co-workers developed a technique to represent chemical reactions by means of Reaction matrices (R-matrices) (Dugundji & Ugi, 1973; Ivanciuc, 2010). Later researchers had studied (Bart & Garagnani, 1977) several reaction schemes and finally, Ugi's work, which classifies chemical reactions into 30 different classes, came

DOI: 10.4018/978-1-4666-4010-8.ch018

to be known as Ugi's scheme (Bart & Garagnani, 1977; Gasteiger & Engel, 2003). In this paper we have proposed an efficient algorithm to automate classification of chemical reactions based on Ugi's scheme. In the following section we first describe the background to elaborate the representation of molecules and chemical reactions using graph theoretic techniques - Bond-Electron matrix (BE-matrix) and Reaction matrix (R-matrix) respectively. Next Ugi's scheme is discussed under the Classification section. In the fourth section, Methods and Algorithm, we have presented our algorithm based on a novel mapping of Ugi's scheme into look-up tables for classification of chemical reactions. In the next section we have theoretically obtained space and time complexity results and discussed their implications compared to upper/lower triangular matrix method of implementation of Ugi's scheme. The last section summarizes our conclusions.

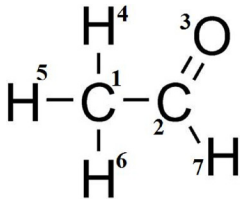
BACKGROUND

Among several representations of chemical structures graph theoretic representation (Ivanciuc, 2010; Kier & Hall, 1999; Trinajstic, 1992) is most suitable for obtaining information using model-driven approach. In the following subsections we have discussed such representation and detailed bond-electron matrix and reaction matrix used in this paper under graph theoretic representation.

Graph Theoretic Representation of Molecules and Reactions

Chemical structures are usually stored in a computer as molecular graphs (Beck et al., 1969). A graph is an abstract structure that contains nodes connected by edges. One example of graph that corresponds to chemical structure of ethanal is shown in Figure 1. In such molecular graphs the nodes correspond to the atoms and the edges to the bonds. Such graphs can also be represented as

Figure 1. Ethanal can be represented in graph theory as a labelled graph



matrices. Their major advantage is that the calculation of paths and cycles can be performed easily by well-known matrix operations. The matrix of a structure with n atoms consists of an array of n × n entries. A molecule with its different atoms and bond types can be represented in matrix form in different ways depending on what kind of entries are chosen for the atoms and bonds. Thus, a variety of matrix representations for chemical compounds are in use (Gasteiger & Engel, 2003): adjacency (Ivanciuc, 2000; Markovic et al., 2001), distance (Ivanciuc & Ivanciuc, 1999; Mihalic et al., 1992), incidence, bond, and bond-electron matrices. Among these representations bond-electron matrix representation provides all information given by bond matrix and additionally number of free electrons of an atom. In this work we have used BE-matrix, the representation of which is explained in the following.

Bond-Electron Matrix

The bond-electron matrix (BE-matrix) was introduced in the Dugundji-Ugi model (Dugundji & Ugi, 1973). The BE-matrix of a molecule with n atoms is an n × n symmetric matrix, with 0th row/column representing each atom of the molecule. It indicates bond order in the off-diagonal elements and the number of free valence electrons of the corresponding atom in the diagonal elements (e.g., O3 = 4 in Table 1). The graph in Figure 2 represents the same ethanal as in Figure 1 but additionally incorporates free electrons of one oxygen atom and corresponding bond electron matrix is shown in the Table 1.