Alkanetraz

I. Overview

Alkanetraz is a program which purpose is to calculate the boiling point of a linear alkane drawn on the screen by the user. The boiling point is a function parametrized over a set of 187 linear alkanes. The user may also add more alkanes to the data set to improve the parametrization.

Requires the python libraries:

- Numpy
- Scipy
- Tkinter

II. Features

a) Draw an alkane

The graphical interface takes the form of a panel of button. By clicking successively on one button, different character, either “C”, “-“ or “|” will appear. “C” represents a carbon, “-“ a bond between two carbons on the left and right-hand side of the button, and “|” represents a bond between two carbons up and down from the button. By clicking the buttons in the panel, the user may thus draw the carbon tree corresponding to the alkane. Of note, for cases where bulky substituents are involved, several “-“ or “|“ character may be drawn to describe one unique bond. Hence, the following succession of characters:

“C” “-“ “-“ “C”

Represents the carbon tree of ethane, and is equivalent to “C” “-“ “C”.

b) Calculate a boiling point

Once the alkane has been drawn on the graphical interface, pressing the button “Boil. Point” on the lower end of the panel will calculate its Boiling point, which appears below the button. The calculation relies on a set of variable that are extracted from the carbon tree (see part III below), and on a set of parameters which are written upon first use of this feature in a file called “parameters.txt”. This set of parameters is the default and have been previously

optimized for a default training set of alkanes (see below). These parameters may be modified either by hand via modifying the file or automatically after changing the training set file (see below).

c) Add an alkane to the data set

Once the alkane is drawn, if the experimental boiling point is known, one can press the button “add to the base” in the lower right corner of the screen. This triggers a procedure which extracts the variables from the carbon tree together with the boiling point provided by the user in the entry field above the button. This allows the user to expand the size of the data set in order to later refine the function parametrization. On the first use of this module, the default training set is written in a file (“dataset.txt”) in the same folder as the program. This file contains 187 lines each containing the variables and experimental boiling points for 187 alkanes and corresponds to the default parameters. It then adds the line corresponding to the drawn alkane.

d) Reparametrize the boiling point function

At any moment,the user may press the button “get new set of parameters”. In this case, the program minimizes the standard deviation of the calculated vs experimental boiling point, calculated on the current training set using a Nelder-Mead algorithm (careful: this step may take a few seconds). The function uses the current parameters as a starting point for the optimization. In total, it requires information contained in the files “dataset.txt” and “parameters.txt”. If any is missing, it is written as the default. After the parameters have been optimized, the file “parameters.txt” is replaced by the new optimal set. This new set will be used for future calculation of the boiling point.

IMPORTANT: To restore the default parameters, delete the file dataset.txt and parameters.txt. Any of those will be rewritten as default whenever the program needs it.

e) Print the exp. Vs calculation data points

By clicking the button calc/exp agreement, the program writes a file called “training-set.txt” which contains the couples of calculated and experimental boiling points for all data points present in the file “dataset.txt”, using the current set of parameters and data set.

III. Mathematical treatment

The program relies on chemical graph theory. In details, the temperature is given in the following function:

$$T_{calc} = A(0) + A(1) \times {}^{1}P^{A(9)} + A(2) \times {}^{2}P^{A(10)} + A(3) \times {}^{3}P^{A(11)} + A(4) \times {}^{4}P^{A(12)} + A(5) \times {}^{5}P^{A(13)} + A(6) \times {}^{6}P^{A(14)} + A(7) \times {}^{th}M^{A(15)} + A(8) \times Z^{A(16)}$$

Where ${}^{n}P$ corresponds to the Wiener path numbers of order n, ${}^{th}M$ is the methyl index and Z is the Hosoya index for the drawn alkane. The A(i) are optimized against a training set of alkanes which experimental temperature (T_{exp}), ${}^{n}P$, ${}^{th}M$ *and* Z are known.

Of note, in the file "dataset.txt", each line is organized as follows

${}^{1}P$ ${}^{2}P$ ${}^{3}P$ ${}^{4}P$ ${}^{5}P$ ${}^{6}P$ ${}^{th}M$ Z T_{exp}

While in the file "parameters.txt", the first line corresponds to the parameters A(0) to A(8), and the second line corresponds to the parameters A(9) to A(16) (from left to write).