Fuel 206 (2017) 171–179

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Full Length Article

Artificial neural network based predictions of cetane number for furanic biofuel additives

Travis Kessler [a], Eric R. Sacia [b], Alexis T. Bell [b], J. Hunter Mack [c,*]

[a] Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA, United States
[b] Department of Chemistry, University of California at Berkeley, Berkeley, CA, United States
[c] Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA, United States

HIGHLIGHTS

- An artificial neural network is used to accurately predict the cetane number of molecules.
- A methodology for extending predictions to underrepresented classes is demonstrated.
- Model predictions are compared to literature values where applicable.
- Two of the furanic candidates posses CN's in a suitable range for use in traditional diesel engines.

ARTICLE INFO

Article history:
Received 4 April 2017
Received in revised form 30 May 2017
Accepted 2 June 2017

Keywords:
Neural networks
Machine learning
Biofuels
Cetane number
Prediction
Quantitative structure property relationship

ABSTRACT

The next generation of alternative fuels is being investigated through advanced chemical and biological production techniques for the purpose of finding suitable replacements for diesel and gasoline while lowering production costs and increasing process yields. Chemical conversion of biomass to fuels provides a plethora of pathways with a variety of fuel molecules, both novel and traditional, which may be targeted. In the search for new fuels, an initial, intuition-driven evaluation of fuel compounds with desired properties is required. Due to the high cost and significant production time needed to synthesize these materials at a scale sufficient for exhaustive testing, a predictive model would allow chemists to preemptively screen fuel properties of potentially desirable fuel candidates. Recent work has shown that predictive models, in this case artificial neural networks (ANN's) analyzing quantitative structure property relationships (QSPR's), can predict the cetane number (CN) of a proposed fuel molecule with relatively small error. A fuel's CN is a measure of its ignition quality, typically defined using prescribed ASTM standards and a cetane testing engine. Alternatively, the analogous derived cetane number (DCN), obtained using an Ignition Quality Tester (IQT), is a direct measurement alternative to the CN that uses an empirical inverse relationship to the ignition delay found in the constant volume combustion chamber apparatus. DCN data points acquired using an IQT were utilized for model validation and expansion of the experimental database used in this study. The present work improves on an existing model by optimizing the model architecture along with the key learning variables of the ANN and by making the model more generalizable to a wider variety of fuel candidate types, specifically the class of furans and furan derivatives, by including specific molecules for the model to incorporate. The new molecules considered include tetrahydrofuran, 2-methylfuran, 2-methyltetrahydrofuran, 5,5′-(furan-2-ylmethylene)bis(2-methyl furan), 5,5′-((tetrahydrofuran-2-yl)methylene)bis(2-methyltetrahydrofuran), tris(5-methylfuran-2-yl) methane, and tris(5-methyltetrahydrofuran-2-yl)methane. Model architecture adjustments improved the overall root-mean-square error (RMSE) of the base database predictions by 5.54%. Additionally, through the targeted database expansion, it is shown that the predicted cetane number of the furan-based molecules improves on average by 49.21% (3.74 CN units) and significantly for a few of the individual molecules. This indicates that a selected subset of representative molecules can be used to extend the model's predictive accuracy to new molecular classes. The approach, bolstered by the improvements

Abbreviations: ANN, artificial neural network; ASTM, American Society for Testing and Materials; CFR, Cooperative Fuels Research; CN, cetane number; DCN, derived cetane number; HMF, 5-(hydroxymethyl)furfural; IQT, Ignition Quality Tester; mdRMSE, mean-delta-root-mean-square-error; QSPR, quantitative structure property relationship; RMSE, root-mean-square error; SMILES, Simple Molecular-Input Line-Entry System.

* Corresponding author.
E-mail address: hunter_mack@uml.edu (J.H. Mack).

http://dx.doi.org/10.1016/j.fuel.2017.06.015

presented in this paper, enables chemists to focus on promising molecules by eliminating less favorable candidates in relation to their ignition quality.

1. Introduction

Research into next-generation alternative fuels has gained significant interest due to concern over global warming, decreasing reserves of conventional fossil fuels, and drawbacks associated with first-generation biofuels like corn ethanol. Biofuels are typically derived from renewable sources such as sugars, starch, and vegetable oil; however, the oxygenated functional groups in biofuel molecules add an additional layer of complexity over traditional hydrocarbons. Though these fuels offer many benefits, especially when derived from cellulosic biomass, the next-generation of biofuels have proven challenging to produce at scale cost-effectively.

Providing predictive insight into key properties, such as the cetane number (CN), can accelerate the development of new alternative fuels. Ultimately, the screening and selection process would address multiple properties; this paper presents a method for increasing the accuracy and extensibility of cetane number predictions, an integral part of the screening process. By shortening the feedback loop inherent to research, scientists can quickly identify the most promising compounds and focus on increasing yield and decreasing costs.

1.1. Cetane number

One of the most important parameters for evaluating a fuel for use in a diesel engine is the cetane number, a measure of the fuel's ignition quality. It is a correlation based on ignition delay from the start of injection and includes both physical (vaporization) and chemical components. There are two widely used methods in determining CN, either using a Cooperative Fuel Research (CFR) engine or an Ignition Quality Tester (IQT). Experimental determination of CN using the single-cylinder CFR is specified through the American Society for Testing and Materials (ASTM) Standard D613 [1]. Other approaches for determining CN include ASTM Standard D7170 [2] and ASTM Standard D6890 [3], both of which use a constant volume combustion chamber. The latter standard utilizes an Ignition Quality Tester (IQT); the method determines the ignition delay in a constant volume combustion chamber by measuring the time between the start of fuel injection and the onset of combustion. The various methods provide accurate CN measurements, although the CN obtained on the CFR is preferred since it reflects combustion behavior in an actual engine. Furthermore, the correlation between DCN and CN is based on an empirical relation and has limited accuracy when used across a range of fuels [4]. The potential impact of this limitation in relation to novel fuels has not been fully characterized. However, the IQT offers a distinct advantage in terms of increased speed and lower volumetric requirements, typically about 100 mL.

Even with the advantages provided by the IQT, the sheer number of potential fuel molecules makes testing prohibitive in terms of both cost and time. This reinforces the need for a rapid and robust screening method for predicting CN, and potentially other properties, in order to aid in alternative fuel development.

1.2. Predicting the cetane number

Predicting cetane numbers and other fuel properties from molecular structure has an extensive history. Prior models based on quantitative structure property relationships (QSPR) have been developed to predict the CN of different compounds, which included an early, but limited, application of backpropagating neural networks for predicting the CN of isoparaffins and diesel fuels [5]. Though the study was limited to branched paraffins, the model showed a superior predictive power compared to conventional equations [6]. A subsequent study used quantitative structure property relationship (QSPR) software to generate 100 molecular descriptors for a set of 275 compounds, including 147 hydrocarbons and 128 oxygenates [7]; a genetic algorithm, or a search heuristic mimicking natural selection in regards to optimization problems, was used to identify which descriptors might influence CN. Although the model did not accurately predict CN (RMSE = 9.1 CN units), the work served as a basis for future models focused on predicting CN using QSPR inputs.

Other types of models have been used to predict CN. One approach utilized an inverse function method to predict the CN of pure hydrocarbons [8]. Though the model is accurate for the range of compounds considered, it is unable to predict the CN of compounds outside the test range. A recent model considered chemical families likely found in diesel fuels using the genetic function approximation (GFA), an iterative approach to generate relationships between molecular descriptors and CN [9]. Though the approach could not satisfactorily predict CN when including all 147 molecules in the data set, it utilized an approach of dividing the set into four different groups based on their chemical families to improve the model's predictive power. The method provides a sufficient local predictive tool for compounds within the same chemical family, but is unable to extend predictions to a larger and disparate data set, including fuels that are combinations of these families.

Another recent model extended the applicability to include alcohols and esters using "consensus" modeling, which averaged results from the outputs of various linear and nonlinear models (including neural networks) [10]. The approach considered 279 compounds from 7 chemical families and predicted CN with a RMSE of 6.3.

In light of the advances and drawbacks inherent to previous models, this paper adopts a backpropagation neural network approach since it appears to be more robust across multiple molecular classes/families due to their nonlinear architecture, which allows for a representation of very complex relationships between input and output vectors [11]. The goal of this paper is twofold: (1) improve upon the state-of-the art models for predicting CN for a diverse data set, and (2) extend the model to consider a new molecular class (furanic compounds). The model's accuracy in regards to the furanic compounds can be compared for two cases, without new experimental data and with new experimental data. As a model's predictive power is only as good as the input data, it is expected that the inclusion of some new furanic compounds will increase the accuracy of the model without affecting the overall RMSE. Ultimately, the objective is to minimize the RMSE as much as possible.

1.3. Furanic biofuel additives

Many strategies exist for converting the sugar units produced by biomass via photosynthesis into fuels. One particularly attractive method is to generate furan derivatives through acid-