Fuel 219 (2018) 426–445

Contents lists available at ScienceDirect

Fuel

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

Review article

Methods of coke quality prediction: A review

Lauren North[a,*], Karen Blackmore[b], Keith Nesbitt[b], Merrick R. Mahoney[a]

[a] Centre for Ironmaking Materials Research, School of Engineering, The Newcastle Institute for Energy and Resources, The University of Newcastle, Callaghan, NSW 2308, Australia

[b] School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Keywords:
Coking coal
Cokemaking
Coke quality
Prediction
Regression
Neural network

ABSTRACT

The prediction of coke quality from global coal basins is critical to coke producers and steel makers for both the selection and effective utilisation of coals. This review analysed the methods described within published models for the prediction of coke quality. Of particular focus were methods that sought to predict coke strength after reaction (CSR) and the related coke reactivity index (CRI). Using the cross industry standard process for data mining (CRISP-DM) as an analysis framework, the models were compared in terms of their data treatment and use of analytical techniques. On reviewing these papers, our results indicate that it is difficult to apply models beyond the conditions under which they were derived, and that many models do not report enough detail to allow complete replication.

1. Introduction

Despite the rise in collection of operating and experimental data, a global prediction model of coke quality from the underlying coal properties has been relatively unsuccessful [1,2]. Historically, traditional regression based approaches have dominated coke quality prediction methods. More recently, non-linear data mining approaches have been applied with varying degrees of success. In this context, data mining is the search for meaningful patterns and information in large volumes of data. In essence, the techniques used, typically termed machine learning, represent the region between traditional statistical analysis and artificial intelligence, although there is no clearly defined boundary between the techniques [3]. At the broadest view, data mining is either predictive (model creation) or descriptive (pattern identification) [4]. Beyond the commonly used predictive regression analysis, data mining is suitable for performing a number of tasks, defined by Witten, Frank [3] as:

- Classification – Assignment of previously unseen instance to a category
- Association – Identification of relationships between features
- Clustering – Grouping similar behaviour
- Numeric Prediction – Similar to classification, but the prediction of continuous values

In terms of applications, data mining has been extensively utilized in the medicine, advertising, and manufacturing domains [5,6]. Whilst not an exhaustive list of applications, the adoption within the metallurgical domain has been limited in comparison [7]. The limited adoption could be attributed to the data mining and metallurgical domains requiring a certain level of knowledge in order to make meaningful and informative contributions that can be usefully executed [8], as well as a lack of publicly available data sources, and a tendency to have poor data collection and management practices [7].

In order to assess the benefits and limitations of existing approaches to coke quality prediction, and to establish a best practice baseline, a review of the literature was conducted. The focus of this review is on data analysis approaches that are undertaken as a key step in the knowledge discovery process, and literature is evaluated in terms of its alignment to a formal framework for describing this process.

1.1. A brief introduction to the knowledge discovery process

Data mining or statistical analysis forms one step of the knowledge discovery process; the overarching method from which information is extracted from available data. The Cross-Industry Standard Process for Data Mining (CRISP-DM) [9] is reported to be the most commonly applied formal method of data analytics [10]. The six step process focuses on obtaining meaningful results from the data, through the intersection of domain knowledge and thorough analytical practice. A summary of the CRISP-DM method is discussed in the following paragraph, and is represented in Fig. 1, whilst a more thorough discussion of the method can be found in [11].

The first step in the process is *business understanding*, which focuses

* Corresponding author.
E-mail address: Lauren.North@uon.edu.au (L. North).

https://doi.org/10.1016/j.fuel.2018.01.090
Received 3 November 2017; Received in revised form 22 January 2018; Accepted 23 January 2018

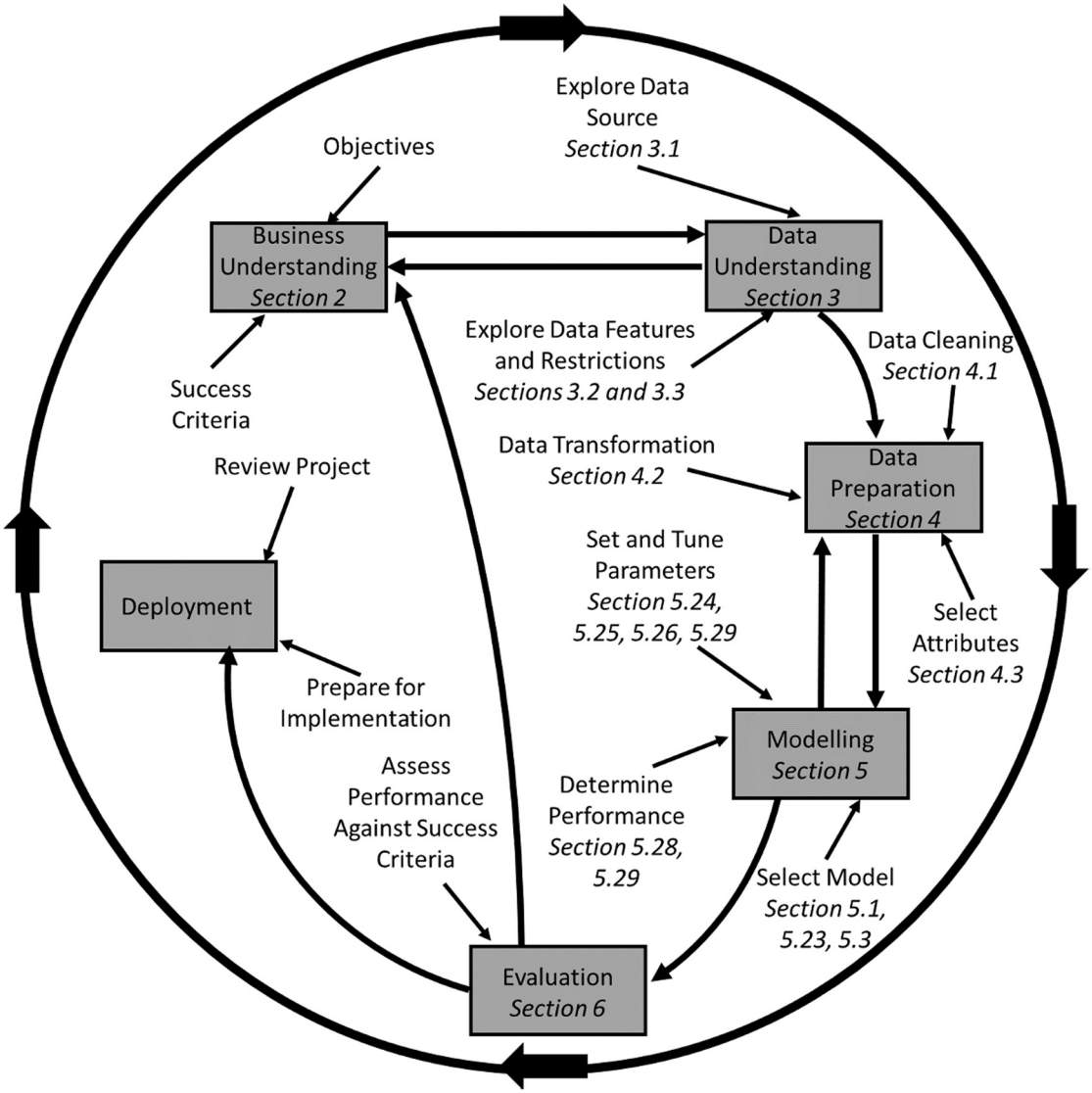

Fig. 1. Flow diagram of the CRISP-DM process. The sections where each stage of the process is discussed with respect to coal and coke quality prediction are shown in italics.

on the project objectives and defining the data mining problem to be addressed, and ensures alignment of the analysis requirements. Strongly linked to the business understanding is the second step of *data understanding*, involving data collection and exploration activities to further develop the problem and potential hypotheses. A final data set is constructed in the *data preparation* phase, with this pre-processing step incorporating attribute selection and creation, and data cleaning. The *modelling* stage applies one or more data mining techniques to the final data set. Following the data mining phase, a 'sanity check' on the outcome or results is performed in the *evaluation* step. It is confirmed that the model does satisfy the business purpose identified in the first stage of the process before the model is implemented in the *deployment* phase. This is where the knowledge is presented in an accessible manner for the end-user. Whilst presented as a linear process here, in practice, the method is iterative as understanding of the problem domain increases at each step of the process. The effective use of such a technique allows for the development of a robust model that addresses business needs and follows a logical process that facilitates validation.

Whilst none of the investigators reported upon in this study explicitly use the CRISP-DM method (although Tiwari et al. [12] present an unreferenced diagram with identical steps to the CRISP-DM method), the concepts from the method are used in this study to aid in the assessment of models presented.

1.2. Method of study

The models discussed in this paper were collected between 30th March 2016 and 30th March 2017, with no bounds placed on the publication date of the model. Due to the nature of publications in the area, which includes conference papers that are not peer reviewed, not all publications are searchable using conventional databases. Therefore, any relevant papers, not found in the search, but referenced by these papers, were added to the review. Not all referenced articles were able to be collected due to restricted availability. Where professional translations were not available, papers not in English were translated using Google Translate [13].

As previously stated, this review focusses on the statistical and data mining based methods for prediction of coal or coke properties presented within the collected papers. Predominantly, this work focuses on the coke quality measures of CSR and CRI (defined in the following section), due to their prevalence in the literature. With regard to regression analysis, papers not modelling a coke quality measure were eliminated. However, in order to expand the available methods for comparison of data mining techniques, consideration was also given to the prediction of coal properties from commonly measured parameters where a data mining technique was used.

In the following sections, the analysis of coke quality prediction