

Сравнение различных типов многомерных регрессий при анализе содержания углерода в низколегированных сталях по спектрам ЛИЭС низкого разрешения

В. В. Кирис, К. Ю. Кацалап, М. В. Бельков

*Институт физики им. Б. И. Степанова НАН Беларуси, Минск, Беларусь,
e-mail: v.kiris@ifanbel.bas-net.by*

На основе набора эмиссионных спектров лазерно-искровой эрозионной плазмы образцов из низколегированных сталей исследована эффективность различных регрессионных моделей (регрессия на латентные структуры, деревья регрессии, регрессия опорных векторов, регрессия гауссовского процесса, ансамбли деревьев регрессии, узкая однослойная нейронная сеть) для анализа содержания углерода. Возбуждение лазерной плазмы проводилось в коллинеарном двухимпульсном режиме на длине волны 1064 нм. Исходные спектры зарегистрированы в диапазоне 172,5–507,5 нм с разрешением ~0.4 нм.

Ключевые слова: лазерно-искровая эмиссионная спектроскопия; сталь; регрессия на латентные структуры; деревья регрессии; регрессия опорных векторов; регрессия гауссовского процесса; ансамбли деревьев регрессии; узкая однослойная нейронная сеть.

Comparison of different types of multivariate regressions in the analysis of carbon content in low-alloy steels using low-resolution LIBS spectra

V. V. Kiris, K. Y. Catsalap, M. V. Belkov

*B. I. Stepanov Institute of Physics of The National Academy of Sciences of Belarus, Minsk, Belarus,
e-mail: v.kiris@ifanbel.bas-net.by*

Based on a set of laser spark emission spectroscopy of low-alloy steels, the effectiveness of various regression models (partial least square regression, regression trees, support vector machine regression, gaussian process regression, ensembles of trees, narrow single-layer neural network) for the analysis of carbon content was studied. The laser plasma was excited in a collinear double-pulse mode at a wavelength of 1064 nm. The initial spectra were recorded in the range 172.5–507.5 nm with a resolution of ~0.4 nm.

Keywords: laser induced breakdown spectroscopy, steel, partial least square regression, regression trees, support vector machine regression, gaussian process regression, ensembles of trees, narrow single-layer neural network.

Введение

Несмотря на обилие новых технологий в производстве материалов, одним из наиболее используемых и распространенных в настоящее время остается сталь. Сталь представляет собой сплав железа, содержащий до 2 % углерода. Добавление углерода в железо существенно изменяет его свойства, что позволяет использовать сталь в различных типах деталей и конструкций. Ежегодно во всем мире производится почти 2 миллиарда тонн стали более 3500 различных марок. Сталь является наиболее часто используемым конструкционным материалом в мире. Как продемонстрировано в работе [1], при характерных для стали концентрациях углерода его

линии в спектрах ЛИЭС (лазерно-искровая эмиссионная спектроскопия) в видимой и ближней ультрафиолетовой областях явно не проявляются, и использование классического метода построения калибровочных графиков невозможно. Тем не менее применение многопараметрической регрессии [2] может обеспечить получение калибровочной зависимости даже в таких условиях.

1. Описание установки и методики получения и обработки спектров

Эрозия и возбуждение вещества анализируемых образцов осуществлялись излучением двухимпульсного Nd:YAG лазера с диодной накачкой (производства ИФ НАН Беларуси) при следующих основных параметрах: длина волны излучения 1064 нм, энергия и длительность импульса генерации (на полувысоте) 50 мДж и 12 нс, соответственно, задержка между импульсами в пакете 10 мкс, частота следования двухимпульсных пакетов – 10 Гц. Излучение лазера проходило через трехкратный оптический телескоп и фокусировалось линзой ($F = 100$ мм) на поверхность мишени. Спектры регистрировались в диапазоне 172,5–507,5 нм посредством малогабаритного спектрометра Firefly4000 производства компании CNPlaser (Китай). Процедура накопления спектральных данных производилась для 5-ти различных точек образца, в каждой из которых регистрировалось пятьдесят кадров, первые десять из которых отбрасывались (обжиг образца), а остальные усреднялись. Регрессионные модели строились на основе набора из 65 стандартных образцов низколегированных сталей производства ИСО (Россия) и IMZ (Польша), концентрация углерода в которых варьировалась от 0,0034 % до 1,32 %. Для построения калибровочной модели регрессии на латентные структуры (РЛС) использовалось программное обеспечение Unscrambler X от CAMO Software AS, для остальных моделей пакет Matlab R 2022b от MathWork. Для проверки моделей использовался метод восьмикратной кросс-валидации.

2. Результаты исследования

Регрессионные модели построенные на основе полных спектров в диапазоне 172,5–507,5 нм содержащих 3629 переменных, обладали очень низкой точностью (например среднеквадратичная ошибка по проверочной выборке (СКОПВ) для оптимизированного дерева регрессии (ОДР) составила $\sim 0,30$ %; регрессии опорных векторов (РОВ) $\sim 0,18$ %; регрессии на латентные структуры $\sim 0,14$ %; для регрессии гауссовского процесса (РГП) построение модели не завершилось; для ансамбля деревьев регрессии $\sim 0,30$ %; для узкой однослойной нейронной сети (УОНС) содержащей десять нейронов – 0,42 %). С учетом того, что концентрация углерода в наборе, на котором происходило обучение, лежит в диапазоне 0,0034 % – 1,32 % только результаты РЛС и РОВ могут представлять какую-то “ценность”. Лучший результат на полных спектрах продемонстрировала РЛС, однако получен этот результат при 20-ти латентных переменных.

Спектры элементов в ЛИЭС содержат набор атомных и ионных линий, перекрывающих небольшие участки спектра. Остальная часть спектрального диапазона не содержит информации о содержании конкретного элемента, но может давать значительный вклад в регрессионную модель (рис. 1). Кроме того, из рис. 1 видно,

что может наблюдаться и обратная ситуация, когда спектральные участки, содержащие сильную линию углерода, не дают значительного вклада в регрессионную модель. Для уменьшения количества входных переменных был выбран участок спектра 420 – 430 нм (127 переменных), содержащий ионные линии углерода С II 426,73 нм С II 426,7 нм, дающие значительный вклад в модель РЛС, хотя они и не проявляются в спектре явно (рис. 2). Кроме того, следует отметить, что наиболее интенсивная линия углерода в спектре С I 247,86 нм не дает значительного вклада в модель РЛС (рис. 1). Вероятно, причиной этого является то, что линия перекрывается линиями железа Fe I 247,98 нм, Fe I 247,48 нм, Fe II 248,02 нм, которые на порядки более интенсивные, чем линия С I 247,86 нм.

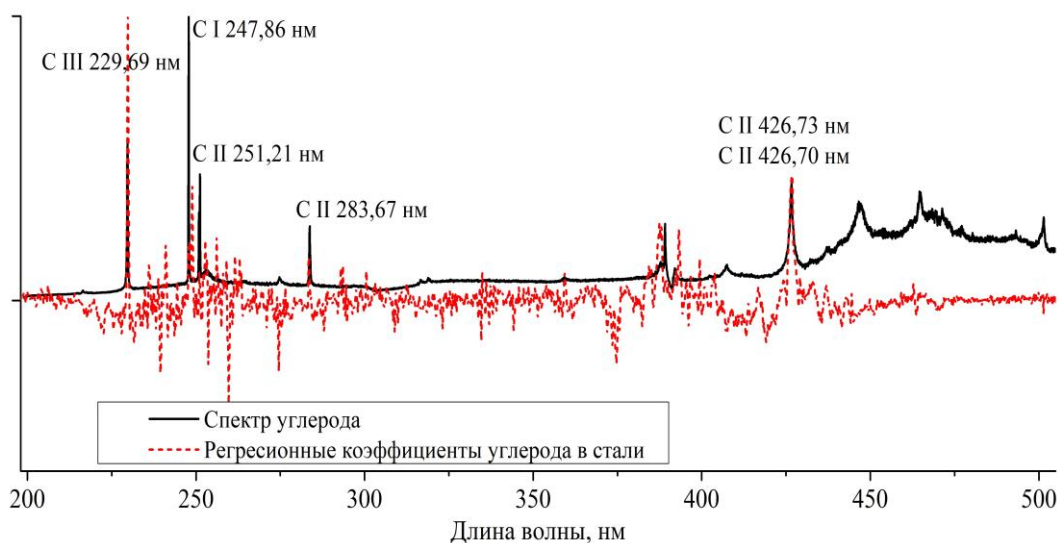


Рис. 1. ЛИЭС спектр графита и коэффициенты регрессионной модели (РЛС) для углерода в сталях

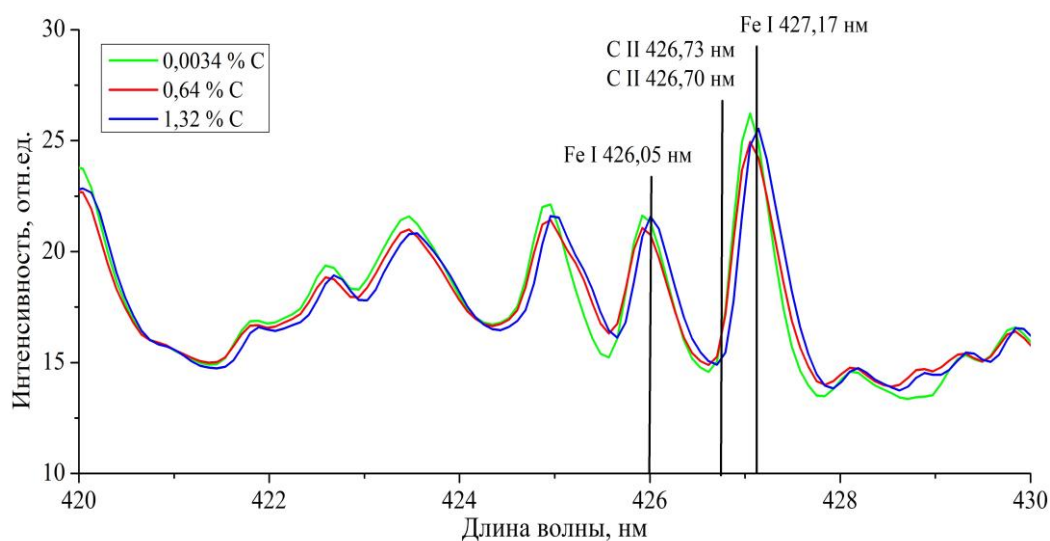


Рис. 2. Спектры образцов стали с различным содержанием углерода

Локализация спектрального диапазона 420-430 нм обеспечила улучшение (СКОПВ) для большинства видов регрессии. Так для РОВ СКОПВ составила 0,11 %; для РЛС – 0,098 %; для РГП – 0,10 %; УОНС – 0,20 %; для деревьев регрессии и ансамблей деревьев регрессии 0,28 % и 0,27 %, соответственно.

При уменьшении входного спектрального диапазона до 77 переменных улучшение СКОПВ происходит только для УОНС – 0,16 %, для остальных методов происходит рост СКОПВ, для РОВ 0,12 %, для РЛС – 0,10 %; для РГП – 0,11 %; для деревьев регрессии и ансамблей деревьев регрессии 0,30 % и 0,28 % соответственно.

Дальнейшее уменьшение входного спектрального диапазона до 47 переменных приводит к увеличению СКОПВ в том числе для УОНС – 0,20 %.

Заключение

При построении многомерных регрессионных моделей не все интенсивные спектральные линии углерода дают соразмерный вклад в модель. Вероятно, вклад определяется не только интенсивностью самой линии, но и интенсивностями линий других компонент стали, налагающихся на данную линию углерода.

Из рассмотренных в работе типов регрессий наилучшую точность в определении содержания углерода на использованном наборе ЛИЭС спектров продемонстрировала РЛС, также данный тип регрессии наименее восприимчив к размерам входных данных и наиболее прост в интерпретации. Наихудшие результаты продемонстрировали деревья регрессии и ансамбли деревьев регрессии. Разброс получаемых результатов для данных типов регрессий сравним с величиной максимальной концентрации углерода в использованном наборе образцов и слабо зависит от размера входных данных. РГП оказалась наиболее требовательна к вычислительным ресурсам и построение модели с использованием полного спектра не удалось завершить, однако после уменьшения размеров входных данных метод показал результаты близкие к РЛС. УОНС наиболее чувствительна к размерам входных данных. На полном спектре ее результаты были наихудшими, но в наибольшей степени улучшились при уменьшении размера анализируемого диапазона.

Благодарности

Данная работа частично выполнена в рамках Государственной программы научных исследований Республики Беларусь «Фотоника и электроника для инноваций», задание 1.5.

Библиографические ссылки

1. Определение углерода в низколегированных сталях с помощью лазерной атомно-эмиссионной спектроскопии в воздухе / В. С. Бураков // Журнал прикладной спектроскопии. 2018. Т. 85, № 5. С. 760–766.
2. Многопараметрическая калибровка концентраций C, Mn, Si, Cr, Ni и Cu в низколегированных сталях методами лазерно-искровой эмиссионной спектроскопии по необработанным спектрам с низким разрешением / М. В. Бельков [и др.] // Журнал прикладной спектроскопии. 2021. Т. 88, № 5. С. 737–742.