بیست و هشتمین کنفرانس اپتیک و
فوتونیک ایران و چهاردهمین کنفرانس
مهندسی و فناوری فوتونیک ایران،
دانشگاه شهید چمران اهواز،
خوزستان، ایران.
۱۴-۱۲ بهمن ۱۴۰۰

مقایسه کاهش ابعاد حاصل از تکنیک‌های PCA و LDA در طبقه‌بندی داده‌های بیناب‌نمایی فروشکست القایی لیزری به روش SVM

یاشار احمدی عالی‌نسب، مرضیه همتی فارسانی، سید محمدرضا دربانی

پژوهشکده علوم و فناوری اپتیک و لیزر دانشگاه صنعتی مالک اشتر، اصفهان

چکیده – کاهش ابعاد تکنیکی است که به منظور افزایش دقت مدل طبقه‌بندی مورد استفاده قرار می‌گیرد. در این پژوهش تاثیر کاهش ابعاد به وسیله‌ی روش‌های تحلیل مولفه اصلی(PCA) و تجزیه و تحلیل تفکیک خطی(LDA) در طبقه‌بندی داده‌های بیناب‌نمایی فروشکست القایی لیزری(LIBS) نمونه‌های آلیاژی آهن-کروم-نیکل با استفاده از روش ماشین بردار پشتیبان(SVM) به همراه دو تابع کرنل خطی و شعاعی(RBF) مورد بررسی قرار گرفت. در روش PCA نتایج طبقه‌بندی SVM با کرنل خطی و شعاعی به ترتیب ۱۰۰٪ و ۹۸٪ حاصل و در روش LDA این نتایج به ترتیب برابر ۱۰۰٪ و ۶۸٪ بدست آمد. نتایج پژوهش کارآمدی روش PCA را در مقایسه با روش LDA در این حوزه به خوبی نشان داد.

کلید واژه- ماشین بردار پشتیبان، بیناب‌نمایی فروشکست القایی لیزری، تحلیل مولفه اصلی، تجزیه و تحلیل تفکیک خطی، توابع کرنل

Comparison of dimensional reduction of PCA and LDA techniques in laser induced breakdown spectroscopy data classification with SVM method

Yashar Ahmadi Alinasab, Marziyeh Hemati Farsani, S. Mohammad Reza Darbani

Optics-Laser Science and Technology Research Center, Malek Ashtar University of Technology, Isfahan

Abstract- Dimensional reduction is a technique used to increase the accuracy of the classification model. In this study, the effect of dimensional reduction by Principle Component Analysis(PCA) and Linear Discriminant Analysis(LDA) on the classification of laser-induced breakdown spectroscopy(LIBS) data of iron-chromium-nickel alloy samples was investigated using the Support Vector Machine(SVM) method with two linear and Radial Basis kernel(RBF) functions. For PCA method, classification results for SVM method with linear and radial kernels were %100 and %98, respectively, and for LDA method, these results were %100 and %68, respectively. The results showed that the PCA method was more effective than the LDA method in this field.

Keywords: Support Vector Machine(SVM), Laser Induced Breakdown Spectroscopy(LIBS), Principle component Analysis(PCA), Linear Discriminant Analysis(LDA), Kernel Function

[Downloaded from opsi.ir on 2024-11-24]

۱- مقدمه

آلیاژها ترکیبات مهمی هستند که در صنایع مختلف مورد استفاده قرار می‌گیرند. به‌منظور افزایش کاربری این آلیاژها، در اختیار داشتن یک الگو برای شناسایی دسته‌ای از این آلیاژها امری حیاتی است. روش بیناب‌نمایی فروشکست القایی لیزری(LIBS) یک فناوری تحلیل عنصری با مزایای منحصربه‌فردی است که از آن جمله می‌توان به درجا و غیرتماسی بودن، تجزیه و تحلیل همزمان چندین عنصر و تشخیص از راه دور بدون آماده‌سازی نمونه اشاره کرد که سبب گسترش دامنه عملکردی آن در اغلب صنایع شده است. تلفیق داده‌های حاصل از این تکنیک با روش‌های آماری مرسوم به‌منظور افزایش توانایی عملکردی این روش در سالیان اخیر بسیار مورد استفاده قرار گرفته است[۱]. پیچیدگی این مجموعه داده‌ها باعث به‌کارگیری روش‌های آماری متنوع از جمله الگوریتم‌های یادگیری ماشین[۱] در مسائل تجزیه و تحلیل طیفی شده است. در این پژوهش در نظر است، تاثیر کاهش ابعاد داده‌ها به‌وسیله‌ی روش تحلیل مولفه اصلی(PCA) و تجزیه و تحلیل تفکیک خطی(LDA) بر طبقه‌بندی داده‌های بیناب نمونه‌های آلیاژ فلزی آهن-کروم-نیکل با استفاده از روش ماشین بردار پشتیبان (SVM) بررسی شود.

۲-تئوری

ماشین بردار پشتیبان یک روش یادگیری ماشین نظارت‌شده[۲] است که در مسائل طبقه‌بندی و رگرسیون به‌کار می‌رود. در سال‌های اخیر ماشین بردار پشتیبان برای تجزیه و تحلیل داده‌های بیناب‌نمایی فروشکست القایی لیزری پیشنهاد شده و به دلیل عملکرد مطلوب، توجه بسیاری از محققان در زمینه کمومتری[۳] و بیناب‌نمایی را به خود جلب نموده است[۲]. این روش بر اساس یافتن یک ابرصفحه بهینه(طبقه‌بندی) است که نمونه‌های کلاس‌های مختلف را تا آنجا که ممکن است به درستی جدا می‌کند. الگوریتم SVM فضایی خطی یا غیرخطی می‌یابد که کلاس‌های آموزشی را با بیشترین فاصله بین داده‌ها در مرزها جدا می‌کند. در مورد داده‌های غیرخطی قابل تفکیک، SVM از ترفند کرنل(معادله۱) برای تبدیل داده‌های خام در یک بعد بالاتر قابل تفکیک‌خطی یا فضای ویژگی استفاده می‌کند.

$$K(x_i.x_j) = \emptyset(x_i)\emptyset(x_j) \qquad (۱)$$

که K نوع تابع کرنل، x_i و x_j نقاط داده آموزشی و $\emptyset(x_i)$ و $\emptyset(x_j)$ داده‌های تبدیل شده‌ی x_i و x_j هستند. از توابع مختلف کرنل می‌توان برای انجام طبقه‌بندی مانند توابع خطی، گاوسی، چندجمله‌ای، سیگموئیدال[۴] و شعاعی(RBF) استفاده کرد. به‌طورکلی، در موارد رابطه غیرخطی بین کلاس‌ها، RBFها به دلیل توانایی آن‌ها در برخورد با داده‌ها، انتخاب اول منطقی هستند[۳].

روش تحلیل مولفه اصلی(PCA) یک تکنیک کاهش ابعاد بدون‌نظارت[۵] است که برای تبدیل یک مجموعه داده اصلی به گروه کوچکتری از متغیرها به نام مولفه‌ها بدون ربط دادن آن‌ها به متغیرهای خروجی استفاده می‌شود. هر مولفه درصدی از واریانس داده‌های ورودی اصلی که شامل دو نوع متغیر تبدیل شده scoreها و loadingها است را توضیح می‌دهد. در این میان loadingها همبستگی بین متغیرهای ورودی و مشارکت آن‌ها در مجموعه داده را توصیف و scoreها، الگوها و همبستگی بین نمونه‌ها را در خود مجموعه داده‌ها کمی‌سازی می‌کند[۴].

---

[۱] Machine Learning
[۲] Supervised
[۳] Chemometry
[۴] Sigmoidal
[۵] Unsupervised

[Downloaded from opsi.ir on 2024-11-24]

روش تجزیه و تحلیل تفکیک خطی(LDA) روشی بسیار معمول برای مسائل کاهش ابعاد به‌عنوان یک مرحله پیش‌پردازش برای یادگیری ماشین و طبقه‌بندی داده‌ها است. تکنیک LDA برای تبدیل ویژگی‌ها به فضای با ابعاد پایین ساخته شده که نسبت واریانس بین کلاس به واریانس درون کلاس را به حداکثر رسانده که منجر به حداکثر تفکیک‌پذیری کلاس خواهد شد[۵].

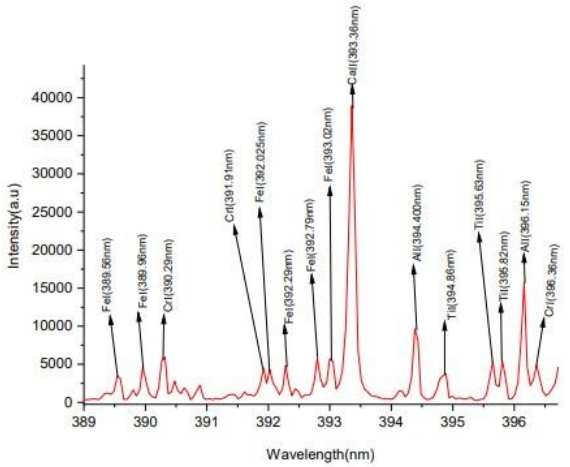
۳- چیدمان تجربی

در این پژوهش از سامانه LIBSCAN100 ساخت شرکت Applied Photonics برای ثبت بیناب ۴ نمونه آلیاژ فلزی آهن-کروم-نیکل استفاده شد. این سامانه دارای یک لیزر Q سوئیچ Nd:YAG با طول موج ۱۰۶۴نانومتر، انرژی خروجی ۱۰۰میلی‌ژول، پهنای تپ۲ $\pm$ ۷ نانوثانیه و نرخ تکرار متغیر ۱ تا ۲۰هرتز بوده و آشکارساز آن قابلیت ثبت بیناب در بازه ۱۸۱ تا ۱۰۵۷ نانومتر با دقت بین ۰/۰۱تا ۰/۰۵ نانومتر در بازه‌های مختلف بینابی را دارد. بیناب ثبت شده از میانگین ۱۰ مرتبه اندازه‌گیری بر روی هر نمونه حاصل شده است. نمونه‌های استفاده شده در این تحقیق حاوی عناصر اصلی کروم(Cr)، منگنز(Mn)، سیلیکون(Si)، کربن(C)، نیتروژن(Ni)، گوگرد(S) و فسفر(P) بوده و نیکل(Ni)، مولیبدن(Mo)، تیتانیوم(Ti)، نیوبیم(Nb)، زیرکونیم(Zr)، مس(Cu)، تنگستن(W)، و برخی عناصر جزئی دیگر است. با توجه به اهمیت عنصر کروم در تعیین مقاومت در برابر خوردگی، حرارت و اکسیداسیون، طبقه-بندی آلیاژهای پایه کروم از اهمیت فوق‌العاده‌ای برخوردار است. درصد عناصر تشکیل دهنده آلیاژهای مورد استفاده در جدول۱ نشان داده شده است.

جدول۱: درصد عناصر تشکیل‌دهنده نمونه‌های آلیاژی

عنصر	نمونه۱	نمونه۲	نمونه۳	نمونه۴
C	۰/۰۴	۰/۰۳	۰/۰۵	۰/۰۶
Si	۰/۵۳	۰/۵۴	۰/۳۸	۰/۷۷
Mn	۱/۷۲	۱/۱۸	۰/۵۹	۲/۰۰
Cr	۱۶/۶۰	۲۲/۳۰	۱۷/۱۰	۱۷/۵۰
Ni	۱۰/۸۰	۴/۰۲	۸/۵۲	۱۰/۰۰
Mo	۲/۰۰	۰/۱۸	۲/۴۶	۰/۳۰
Ti	-	-	-	۰/۵۵
Cu	۰/۱۹	۰/۳۶	۰/۳۵	۰/۴۰

بیناب LIBS یکی از نمونه‌ها در شکل۱ نشان داده شده است.



شکل۱: بیناب LIBS نمونه چهارم در بازه بینابی ۳۹۶-۳۸۹ نانومتر

۴- روش کار و تحلیل نتایج

برای استفاده از روش‌های آماری فوق از زبان برنامه‌نویسی پایتون استفاده شد. مجموعه داده تهیه شده ابتدا مورد نرمال‌سازی قرار و بعد از تعیین ویژگی‌ها(X) و برچسب‌ها(Y) که شامل نمونه‌های آلیاژی با درصد عناصر مختلف بود به ترتیب با شماره۱، ۲، ۳ و ۴ نشان داده شده است، داده‌ها به دو مجموعه داده آموزشی و آزمایشی به وسیله‌ی روش ارزیابی Stratified K-Fold CV با K= ۵ تقسیم شدند. تفاوت این تکنیک با روش K-Fold CV حفظ نمودن نسبت بین داده‌های آموزشی و آزمایشی است. در ادامه مجموعه داده آموزشی یک بار توسط روش PCA و بار دیگر توسط روش LDA مورد کاهش ابعاد قرار گرفت و در هر مورد بعد از کاهش ابعاد، الگوریتم ماشین بردار پشتیبان یک بار با تابع کرنل خطی و یک بار هم با تابع کرنل شعاعی(RBF) برای طبقه‌بندی داده‌ها استفاده شد که نتایج آن‌ها در جدول۲ نشان داده شده است. در روش PCA تعداد مولفه‌های انتخابی، ۵ مولفه با درصد واریانس ۴۵٪، ۲۵٪، ۱۳٪، ۷٪ و ۲٪ بود که به ترتیب برای ایجاد PC1، PC2، PC3، PC4 و PC5 استفاده شد.

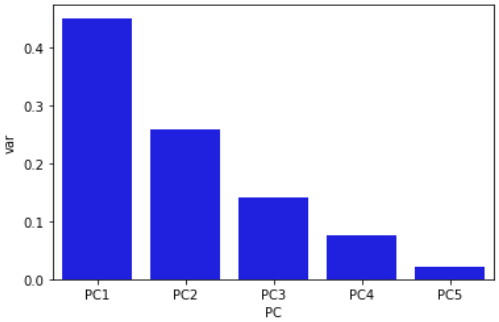
جدول۲: نتایج کاهش ابعاد به روش LDA و PCA

PCA	LDA	روش‌ها
٪۱۰۰	٪۱۰۰	Linear SVC
٪۹۸	٪۶۸	RBF SVC

درصد واریانس و تاثیر هر کدام از مولفه‌ها در طبقه‌بندی داده‌ها در نمودار شکل۲ نشان داده شده است. همانطور که

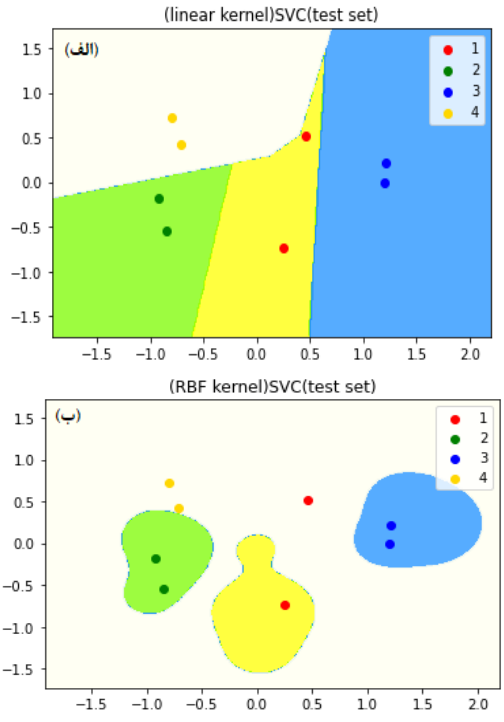
[Downloaded from opsi.ir on 2024-11-24]

در شکل۲ دیده می‌شود، سه مولفه اول بیش از ۷۰درصد واریانس را شامل می‌شوند که باعث انتخاب سه PC نخست می‌شود.

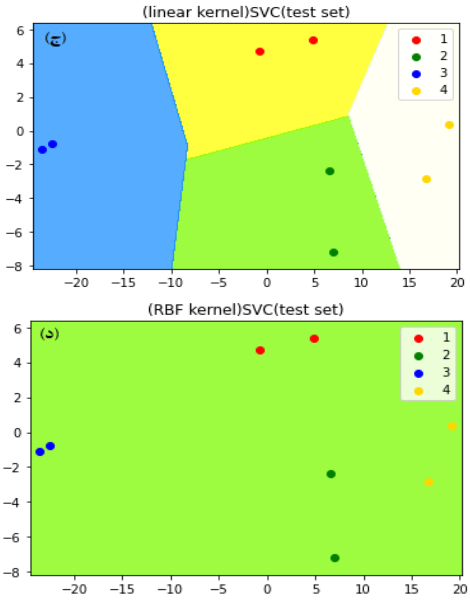


شکل۲: درصد واریانس هر کدام از PCها

نمودار پراکندگی طبقه‌بندی داده‌های حاصل از SVM پس از کاهش ابعاد به وسیله‌ی روش PCA در شکل‌های ۳ الف و ب نشان داده شده است. در به‌کارگیری روشLDA، با توجه به این که تعداد مولفه برای آن روش بایستی کمینه‌ی بین تعداد ویژگی‌ها و تعداد کلاس(برچسب) منهای یک باشد، تعداد ۳ مولفه در نظر گرفته شد. نمودار پراکندگی طبقه‌بندی داده‌های حاصل از SVM پس از کاهش ابعاد به وسیله‌ی روش LDA در شکل‌های ۳ ج و د نشان داده شده است.



شکل۳: نمودار طبقه‌بندی داده‌ها همراه کاهش ابعاد به روش PCA بر اساس الف) تابع کرنل خطی SVM ، ب) تابع کرنل شعاعی (RBF) SVM



ج) نمودار طبقه‌بندی داده‌ها همراه کاهش ابعاد به روش LDA بر اساس تابع کرنل خطی SVM و د) تابع کرنل شعاعی (RBF) SVM

۵- نتیجه‌گیری

نتایج این تحقیق نشان داد که تلفیق روش‌های کاهش ابعاد PCA و LDA با روش طبقه‌بندی ماشین بردار پشتیبان(SVM)، می‌تواند تاثیر مطلوبی بر دقت طبقه‌بندی داشته باشد. نتایج حاصل از روش PCA برای SVM با توابع کرنل خطی و شعاعی(RBF) به ترتیب برابر ۱۰۰٪ و ۹۸٪ و برای LDA به ترتیب ۱۰۰٪ و ۶۸٪ است. با توجه به نتایج بدست آمده تکنیک PCA روش مناسب و کارآمدی برای کاهش ابعاد در مسائل کمومتری مربوط به LIBS است.

۶- مرجع‌ها

[1].T. Zhang, H. Tang, and H. Li, *Chemometrics in laser-induced breakdown spectroscopy* ,Journal of Chemometrics, 2018. **32**(11): p. e2983.
[2].S.R Gunn,.,Support vector machines for classification and regression. ISIS technical report,1998.**14**(1): p.5-16.
[3].Aberkane, S.M., et al., *Sorting zamak alloys via chemometric analysis of their LIBS spectra*. Analytical Methods, 2017. **9**(24): p. 3696-3703.
[4]. A.P, Rao, Rapid quantitative analysis of trace elements in plutonium alloys using a handheld laser-induced breakdown spectroscopy (LIBS)device coupled with chemometrics and machine learning. Analytical Methods, 2021. **13**(30): p. 3368-3378.
[5].P, Xanthopoulos, P.M. Pardalos, *Linear discriminant analysis*,in *Robust data mining*. 2013, Springer. p. 27-33.

[Downloaded from opsi.ir on 2024-11-24]