A review on Steganalysis in high dimensions by fusing classifiers

مروری برروش استگآنالیز در ابعاد بالا به کمک هم جوشانی کلاسیفایرها

مینا اتحادی ابری [1] ،دکتر مریم امیرمزلقانی

دانشگاه صنعتی امیرکبیر، استادیار دانشگاه امیرکبیر

چکیده

به دلیل پیچیده شدن روز افزون متدهای استگانوگرافی (اختفای اطلاعات) در ابعاد بالای فضای ویژگی که سعی در افزایش امنیت و اختفای اطلاعات به طور کاملا نامحسوس در کاور را دارند، نیاز به متدهای جدید استگآنالیز(تشخیص اطلاعات نهان) جهت جدا کردن شی کاور از شی استگانوگرافی حاوی اطلاعات پنهان، حس می شود. در کار کردن با کاورهای با ابعاد بالا، اگر بخواهیم از متدهای قدیمی استگآنالیز استفاده کنیم می بینیم که محاسبات بسیار پیچیده می شود و برای محاسبات، وابستگی شدیدی به المان های متعدد کاور به وجود می آید. به طور کل می توان گفت در استگآنالیز در ابعاد بالا دو موضوع مطرح است: اولاانتخاب ویژگی های مناسب از میان کل ویژگیهای موجود در ابعاد بالا که باید برای کلاسیفایر در نظر گرفته شوند، دوما استفاده از یک الگوریتم یادگیری ماشینی که مقیاس پذیر با ابعاد بالا باشد. در کلاسیفایر کردن در ابعاد بالا مشکلاتی مانند کمبود داده های مجموعه آموزش، پیچیده بودن محاسبات و مرحله ی آموزش داده ها، غیر قابل تعمیم بودن، روباست پایین و کاهش کارایی وجود دارد. به مشکلات مذکور، پدیده نفرین بعدگفته می شود. در جهت رفع این مشکل در روش استگانالیز نوین بر خلاف تکنیک های قدیمی به جای استفاده از تعداد متعددی ماشین بردار پشتیبان، از روش کلاسیفایرها [2] استفاده شده است. در واقع در این روش تعدادی کلاسیفایر ضعیف (زیر کلاسیفایر[3])داریم که برمبنای انتخاب رندم تعدادی از ویژگی ها از میان کل ویژگیهای موجود ساخته شده اند و کلاسیفایر نهایی از ترکیب و هم جوشانی این کلاسیفایرهای ضعیف بدست می آید. مزایای استفاده از این نوع کلاسیفایر، قابل تعمیم بودن، سادگی، سریعبودن، عدم پیچیدگی محاسباتی و افزایش کارایی است. لازم به ذکر است که این روش برای اولین بار در مقاله [۱,2] مطرح شده و در کارهای بعدی تغییرات و اصلاحات جدیدی در جهت افزایش کارایی آن برای استگانالیز درابعاد بالا صورت پذیرفته است.

کلمات کلیدی: استگآنالیز، استگانوگرافی، کلاسیفایر هم جوشان، نفرین بعد

۱.مقدمه

[1] **Corresponding author: Mina etehadi Abari**
Email: minaetehadi@aut.ac.ir
[2] **Classifier fusion**
[3] **Sub-classifier**

در متدهای قدیمی، استگآنالیزهابه صورت کلاسیفایرهای باینری تعریف می شوند که بر روی منابع اعمال می گردند و تعیین می کنند که کدام منبع کاور و کدام یک حاوی استگانوگرافی می باشد. این استگآنالیزها توسط ویژگی های کاور و استگانوگرافی، آموزش داده می شوند. همچنین ویژگی ها به طور دستی برای آموزش و کلاس بندی و تعیین تاثیر متد استگآنالیز مطرح شده، استخراج می شدند. با پدید آمدن متدهای جدید استگانوگرافی غیر قابل تشخیص مانند HUGO[7] تشخیص اطلاعات نهان (استگآنالیز) با کمک کلاسیفایرهای باینری قدیمی سخت تر شد چراکه در متد HUGO استگانوگرافی بر روی کاورهای با ابعاد بالا ، که وابستگی پیچیده به پیکسل ها و المان های متعدد دارند، اعمال شده است. درنتیجه برای تشخیص اطلاعات استگانوگرافی شده با ابعاد بالا، روش استگآنالیز با همجوشانی کلاسیفایرها ارائه شده است[1]. در این مقاله مسئله فضاهای ویژگی در ابعاد بالا بررسی می شود، به علاوه اینکه به مسئله ماشین های یادگیری مقیاس پذیر می پردازیم .به طور کلی هدف از این مقاله، مروری بر روش هم جوشانی کلاسیفایرها واثبات مفید بودن آن در استگآنالیز با ابعاد بالا طی آزمایش های انجام شده می باشد.

۲.چالش های پیش رو

سه فاکتور اصلی که بر روی روش های یادگیری ماشین و کارایی کلاسیفایر ها تاثیر منفی می گذارند عبارتند از:۱- کمبود نمونه داده های مجموعه آموزش ۲- پایین بودن قدرت تمایز و تشخیص ضعیف کلاسیفایر مورد استفاده ۳- افزایش بعد که نفرین بعد را پدید می آورد.

فرض کنید N تا داده ی آموزشی با d بعد و دو کلاس X,Y به شرح زیر داریم:

$x^m, y^m \in R^d$ و $Y=\{y^1,\dots,y^{N/2}\}$ و $X=\{x^1,\dots,x^{\frac{N}{2}}\}$

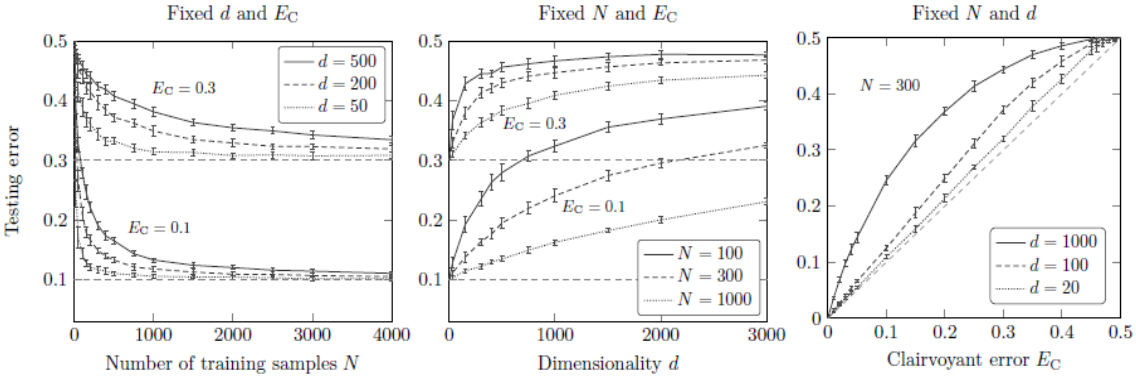
که هر کدام از نمونه ها داری توزیع گوسی یکنواخت و یکسان مستقل[4] با میانگین صفر می باشند (در کلاس x) و s>0 برای کلاس y. فرض می کنیم برای هر نمونه تست میانگین $\bar{z}$ است و حد آستانه s/2 است. به این کلاسیفایر، کلاسیفایر clairvoyant می گویند. در اینجا خطای تست و قدرت تشخیص کلاس به شرح زیر است:

$$E_c = 1 - \emptyset(s\sqrt{d/2}) \qquad (1)$$

$$D_c = 1 - 2E_c \qquad (2)$$

در فرمول فوق، $\emptyset$ ، c.d.f متغیر نرمال استاندارد است، E_c خطای تست کلاسیفایر و D_C قدرت تشخیص و تمایز کلاسیفایر مورد استفاده می باشد. شکل ۱، آزمایش های انجام شده در مقاله [1] را نشان می دهد که در هر مرحله، یکی از سه تا پارامتر d,N, E_c را تغییر داده و دو تای دیگر را ثابت نگه داشته است. به طور کل می توان گفت که شکل ۱ تاثیر سه پارامتر نمونه داده های مجموعه آموزشN ، قدرت تمایز و تشخیص کلاسیفایر مورد استفادهD_C و تعداد بعد dرا نشان می دهد.

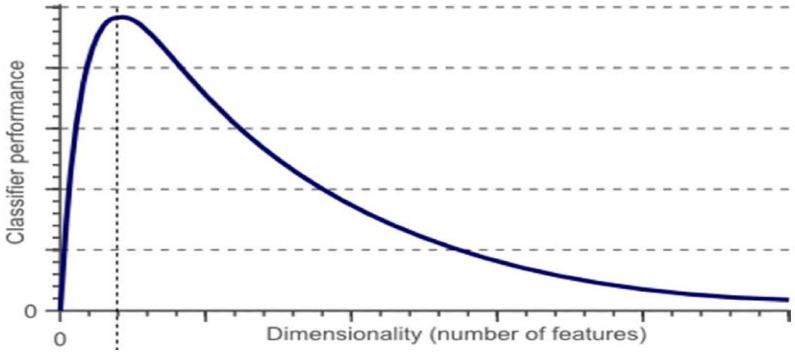
[4] IID



شکل۱- تاثیرتغییر پارامترهای N,d,E_C بر روی عملکرد کلاسیفیکیشن-هر آزمایش ۵۰ بار تکرار شده است.

قبل از تفسیر سه شکل شماره ۱ باید این نکته را یادآور کرد که E_c (خطای تست کلاسیفایر) و D_C (قدرت تشخیص کلاسیفایر) با یکدیگر رابطه مستقیم دارند. این موضوع در فرمول(2) مشهود است. شکل ۱ سمت چپ، d یعنی بعد و D_C یعنی قدرت تشخیص کلاسیفایر ثابت است و با افزایش تعداد داده های مجموعه آموزشN ، خطای تست در L-SVM ،که با +I نشان داده شده، کاهش می یابد. همچنین از شکل یک سمت چپ می توان دریافت که اگر تعداد ابعاد d زیاد باشد و یا اگر قدرت تشخیص کلاس D_C کم باشد، نیاز به تعداد داده های مجموعه آموزش N بیشتری داریم. این آزمایش با دو مقدار متفاوت $E_c = 0.1$,$E_c = 0.3$ انجام شده است هرچقدر E_c کمتر باشد، یعنی خطای تست کلاسیفایر کمتر است. در شکل ۱ وسط، تعداد داده های آموزشی Nو قدرت تشخیصD_C ثابت است و با افزایش بعد d، مقدار خطا افزایش پیدا می کند. یعنی افزایش بعد تاثیر منفی دارد. این آزمایش با سه مقدار متفاوت داده های آموزشی N=100,300,1000 انجام شده است. طبیعی است که هرچه تعداد داده های آموزشی افزایش یابد خطای تست کمتر می شود. در شکل ۱ سمت راست N, d ثابت است و خطای تست کلاسیفایرE_c تغییر می کند. اگر E_c =0 باشد یعنی خطای تست مینیمم شود می بینیم که L-SVM به خوبی عمل می کند و اگر E_c =1 باشد یعنی خطای تست به حداکثر میزان خود برسد، انتخاب های کلاسیفایر L-SVM فقط به صورت رندم می باشد که این بدترین حالت ممکن را نشان می دهد. در اینجا نیز سه بعد متفاوت d=20,100,1000 در نظر گرفته شده است، افزایش بعد، خطا را افزایش می دهد.

استگانوگرافی های ضعیف با ابعاد پایین به راحتی قابل تشخیص هستند چراکه اطلاعات نهان اعمال شده در کاور اولیه (به عنوان مثال یک تصویر jpeg)، باعث تغییر در ویژگیها و المنت های اولیه کاور می شوند.این تغییر ویژگیها در ابعاد پایین حتی با داشتن تعداد کم نمونه های مجموعه آموزش به راحتی قابل شناسایی هستند ولی با افزایش بعد و متعاقبا افزایش تعداد ویژگیهای کاور، استگآنالیز پیچیده تر و دشوارتر می شود و نیاز به نمونه داده های آموزشی بیشتری برای آموزش کلاسیفایر داریم. از طرفی دیگر افزایش تعداد نمونه های مجموعه آموزش باعث کاهش سرعت اجرا و افزایش فضای حافظه اشغالی می گردد. به علاوه در اکثر استگانالیزها در دنیای واقعی و کاربردی، تعداد کاورهای منبع که می توانند به عنوان نمونه آموزشی اولیه در نظر گرفته شوند بسیار محدود است. در نتیجه می توان گفت مهم ترین چالش در استگآنالیزدر بعد بالا، مسئله نفرین بعد می باشد. شکل۲ یک نمونه از پدیده نفرین بعد را نمایش می دهد. در ادامه راهکارهای ارائه شده برای رویایی با چالش فوق مطرح می گردد.



شکل2- با افزایش ابعاد و متعاقبا افزایش تعداد ویژگی ها عملکرد کلاس بند کاهش می یابد

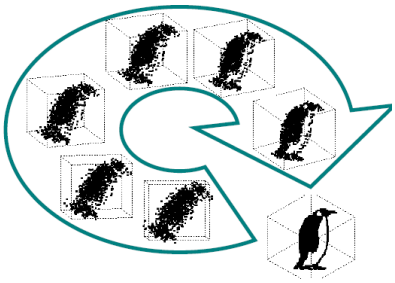
۳.اصلی ترین رویکردهای کلاسیفیکیشن در ابعاد بالا

اولین سوال این است که بهترین روش بهینه سازی وانتخاب مناسب ترین ویژگیها از میان کل فضای ویژگی ها در ابعاد بالا در هنگامی که تعداد نمونه های مجموعه آموزش محدود است برای استگآنالیز چیست؟ اولین و سطحی ترین راهکار، استفاده مستقیم از ابزارهای کلاسیفیکیشن است. در برخی روش ها استفاده از ماشین بردار پشتیبان به دلیل مقاوم و روباست بودن نسبت به مسئله ی نفرین بعد و داشتن محاسبات امکان پذیر، پیشنهاد می شود ولی در اینجا بیشتر هدف بررسی حالت هایی است که تشخیص الگوریتم استگانوگرافی استفاده شده دشوار است و کلاسیفیکیشن ها به طور مستقیم نمی توانند بر روی فضای ویژگی ها اعمال شوند، چرا که حجم و پیچیدگی محاسبات بسیار بالا می رود. اصلی ترین استراتژی های پیشین استگآنالیز برای کلاس بندی و تشخیص اطلاعات نهان در ابعاد بالا به شرح زیر است:

۱- ابتدا کاهش بعد و سپس کلاس بندی:

کاهش بعد می تواند با نظارت یا بی نظارت باشد. یک تکنیک کاهش بعد با نظارت، تکنیکی است که در مقاله[3] معرفی شده و تکنیک انتخاب ویژگی نام دارد. تکنیک PCA نیز یکی از اصلی ترین تکنیک با نظارت استفاده شده برای کاهش بعد است.در واقع تبدیل PCA داده ها را به فضای با ابعاد کمتری می برد.شکل ۳ یک نمونه از تبدیل PCA را نمایش می دهد. هدف کاهش تاثیر نفرین بعد در کلاس بندی است. در متدهای استگآنالیز قدیمی، کاهش بعد بیشتر بصورت دستی و با ابتکارهای[5] خود انسان بدست آمده است.

[5] **heuristic**



شکل ۳- یک نمونه از تبدیل PCA

۲- کاهش بعد و کلاس بندی همزمان با هم

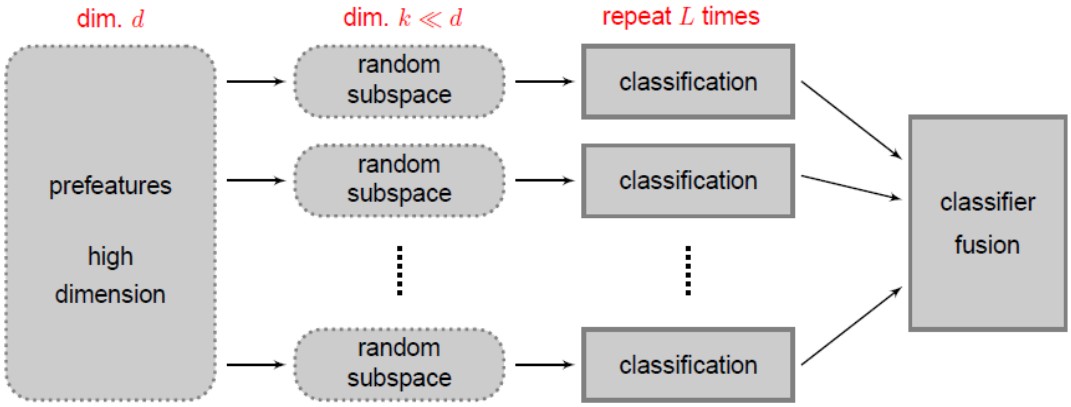
در این روش به صورت الگوریتم تکرار شونده در هر مرحله هم بعد را کاهش می دهیم و هم فیدبکی از کلاسیفیکیشن را مشاهده می کنیم. در یادگیری ماشین به این متدها متدهای توکار[۶] و رپر[۷][9] و [3]می گوییم.

۳- هم جوشانی کلاسیفایرها

در این روش به طور رندم ابعاد را کاهش می دهیم و یک کلاسیفایر درست می کنیم. در نتیجه زیرکلاسیفایرهای [۸]متعدد از زیر فضاهای مختلف خواهیم داشت که به آنها کلاسیفایر ضعیف نیز گفته می شود. این کار را مدام تکرار می کنیم و در نهایت کلاسیفایرهای گوناگون را با یکدیگر ادغام و هم جوشانی می کنیم. برای درست کردن کلاسیفایرهای هم جوشان با ناظر، باید زیر کلاسیفایرهای متنوع زیادی داشته باشیم که انواع داده های دیده نشده را پوشش دهند و خطاهای تست گوناگونی برای داده های دیده نشده به ما بدهند. در اینجا تنوع بسیار مهم تر از دقت کلاسیفایرها است. هنگامی که به طور رندم زیر فضاهایی از فضای کل ویژگی ها انتخاب می کنیم که منجر به کاهش بعد و از بین بردن مسئله نفرین بعد می شود، ماشین بردار پشتیبان فضای مارجینی بین کلاس های مختلف ایجاد می کند که این فضای مارجین نسبت به حالتی که بعد کاهش پیدا نکرده است، فضای مارجین بهتری دارد. به بیان دیگر در این حالت قدرت تشخیص کلاسیفیکیشن ها افزایش می یابد. می توان گفت این روش الهام گرفته شده از روش کلاسیفایر افزایشی[۹] می باشد[4].دیاگرام ۱ چگونگی عملکرد متد کلاسیفایر هم جوشان را نمایش می دهد. نتایج بدست آمده در مقاله [1,2] و همچنین آزمایشات انجام پذیرفته توسط ما، جهت ارزیابی استگانالیز در ابعاد بالا، نشان داده که این روش نه تنها برای jpeg بلکه برای تمامی دامنه های فضایی و سایر کاورهای مختلف در ابعاد بالا نتایج خوبی می دهد.

---

[6] **Embedded method**
[7] **Wrapper method**
[8] **Sub-classifier**
[9] **Boosting (aggregation of weak classifiers)**



دیاگرام ۱- ساختار کلی متد کلاسیفایر هم جوشان

۴.پیاده سازی کلاسیفایر هم جوشان

الگوریتم۱ مربوط به کلاسیفایر همجوشان است که در آن پارامترها به شرح زیر می باشند:

d ابعاد زیر ویژگی ها[10] است.

N^{trn}تعداد نمونه های مجموعه آموزش در هر کلاس

N^{tst} تعداد نمونه های مجموعه تست در هر کلاس

L تعداد زیر کلاسیفایر های متعدد که از کاهش ابعاد به طور رندم در هر زیر فضا ساخته شده است.

کلاسیفایر F1 شامل ۰ و ۱ است که ۰ به کاور اشاره می کند و ۱ به استگانوگرافی اشاره می کند. کلاسیفایر نهایی F1 از ترکیب چند تا زیر کلاسیفایر بدست آمده است. زیر کلاسیفایرها در ابتدا از نوع FLD[11] هستند چرا که پیچیدگی کم و کلاسیفایرهای کم ثبات[12]، باعث افزایش تنوع در زیر کلاسیفایرها می شوند و باعث می شوند انواع داده های دیده نشده و جدید را در بر گیرند[8]. در نهایت L تا زیرکلاسیفایر را با یکدیگر ترکیب و هم جوشانی می کنیم تا به یک استراتژی مناسب برسیم. باید توجه کرد که در اینجا زیر کلاسیفایرها بر اساس فضاهای ویژگی در ابعاد محدود[13] d_{red} (تعداد ابعاد انتخاب شده، نسبت به بعد کلی سیستم) آموزش داده می شوند که ممکن است d_{red} ابعادشان کمتر از ابعاد کل سیستم اولیه باشد، در نتیجه زیر کلاسیفایرها به تنهایی هرکدامشان، دقت اولیه خوبی نخواهند داشت ولی با هم جوشانی این زیرکلاسیفایرها، به دقت قابل توجهی خواهیم رسید.

[10] **Prefeatures**
[11] **fisher linear discriminants**
[12] **unstable**
[13] $d_{reduction}$

Algorithm 1 Ensemble classifier.

1: **for** l=1 to L **do**
2: Randomly select $\mathcal{D}_l \subset \{1,\ldots,d\}, |\mathcal{D}_l| = d_{\text{red}} < d$
3: Train a classifier F_l on cover features $\mathbf{x}_m^{(\mathcal{D}_l)}$ and stego features $\mathbf{y}_m^{(\mathcal{D}_l)}$, $m = 1,\ldots,N^{\text{trn}}$. Each classifier is a mapping $F_l : \mathbb{R}^{d_{\text{red}}} \to \{0,1\}$.
4: Make decisions using F_l:

$$F_l(\mathbf{b}) \triangleq [F_l(\mathbf{b}^{(1)})),\ldots,F_l(\mathbf{b}^{(N^{\text{tst}})})] \in \{0,1\}^{N^{\text{tst}}}. \quad (1)$$

5: **end for**
6: Fuse all decisions by voting for each test example $k \in \{1,\ldots,N^{\text{tst}}\}$:

$$F(k) = \begin{cases} 1 & \text{when } \sum_{l=1}^{L} F_l(\mathbf{b}^{(k)}) > L/2 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

7: **return** $F(k), k = 1,\ldots,N^{\text{tst}}$

الگوریتم ۱- الگوریتم کلاسیفایر هم جوشان

بهترین حسن الگوریتم هم جوشان عدم پیچیدگی محاسباتی می باشد و اگر درست پیاده سازی شود، داده های مجموعه آموزش بستگی به ابعاد زیر ویژگی ها ندارند. درواقع طبق گام ششم الگوریتم ۱، در هر زیر کلاسیفایر به طور رندم تنها بخشی از زیرویژگی ها (و نه کل ابعاد استگانوگرافی) در نظر گرفته می شوند و نهایتا اگر بیش از نصف زیر کلاسیفایر ها رای بر وجود اطلاعات نهان بر روی شی مورد بررسی دهند، خروجی ۱ (استگانوگرافی) و در غیر این صورت خروجی صفر(کاور) خواهد بود. کلاسیفایر نهایی هم جوشان بستگی به دو پارامتر d_{red} ، L دارد که L تعداد زیرکلاسیفایرها است و d_{red} تعداد ابعاد نسبت به بعد کلی سیستم d می باشد. دقت کلاسیفایر بستگی به L دارد و پیچیدگی کلاسیفایر بستگی به تعداد ابعاد، d ، دارد. برای افزایش سرعت باید L را کوچک انتخاب کنیم. d_{red} را به طور دستی و تجربی با آزمایش های متعدد انتخاب شده است.

۵.آزمایشات در حوزه jpeg

در مرحله ی تست، متد استگآنالیز مذکور تجزیه و تحلیل شده و تاثیر تغییر پارامترهای مختلف بر روی الگوریتم nsF5 مشاهده می شود. الگوریتمnsF5 امن ترین الگوریتم رایجی است که با تغییر و دستکاری مستقیم ضرایب همبستگی DCT کار می کند[5]. تست بر روی ۶۵۰۰ تا تصویر jpeg که از بیست دوربین مختلف بدست آمده اند و با فاکتور کیفیت ۷۵ فشرده سازی شده اند، انجام پذیرفته است [1,2]و برای کاهش سایز، تمامی تصاویر را به خاکستری[۱۴]تبدیل شده اند. برای کلاس بندی از libsvm ،که یک روش عمومی ماشین بردار پشتیبان برای بیش از دو نوع کلاس است، استفاده شده است[6]. به طور رندم نیمی از تصاویر بعنوان مجموعه آموزش و نیمه ی دیگر بعنوان مجموعه تست در نظر گرفته شده اند. آستانه تصمیم گیری کلاسیفایرها بر اساس مینیمم خطای کلاس بندی، تعیین می شود. برای سنجش دقت متد پیشنهاد شده نیاز به یک ارزیاب استاندار داریم که به شرح زیر است:

[14] **grayscale**

$$P_E = \min_{P_{FA}} \frac{1}{2}(P_{FA} + P_{MD}(P_{FA})) \qquad (3)$$

P_{FA} احتمال هشدار اشتباه یعنی یک کاور ساده را اشتباها حاوی پیام مخفی، تشخیص دهد.

P_{MD} احتمال عدم تشخیص استگانوگرافی

P_E[15] احتمال خطای کلی است که از آن برای سنجش دقت تشخیص استگآنالیز استفاده می شود.

در این آزمایشات تاثیر پارامترهای مختلف بر روی الگوریتم nsF5 بررسی شده است و به طور رندم ده بار ، مجموعه داده به دو بخش مجموعه تست و مجموعه آموزش تقسیم گردیده.جدول ۱ نتایج P_E کلاسیفایر[16]گوسین G-SVMو کلاسیفایر هم جوشان را برای زیر ویژگی های مشابه را نمایش می دهد.. G-SVM با 5-fold cross validation به شرح زیر آموزش دیده است:

$$C = 10^a,\ \gamma = \frac{1}{N_F} \cdot 2^b,\ a \in \{-3, \ldots, 4\},\ b \in \{-5, \ldots, 2\}, \qquad (٤)$$

در فرمول فوق N_F تعداد زیر ویژگی ها است. برای کلاسیفایر هم جوشان نیز d_{red} =400 و تعداد زیر کلاسیفایرها L=31 می باشد. شکل ۴، P_E را برای دو مقدار ظرفیت ترابری[17]مختلف نشان می دهد. در اینجا ظرفیت ترابری به معنای توانایی درج بیت مخفی است و واحد آن[18] bpbs است. در شکل ۴، هرچقدر تعداد زیرکلاسیفایرها افزایش یابد و تعداد بیشتری زیر کلاسیفایر با یکدیگر هم جوشانی داشته باشند، خطا کمتر می شود و دقت افزایش می یابد. خط افقی مستقیم در این شکل، عملکرد کلاسیفایر G-SVM را نشان می دهد. و خط نیمه کج با علامت Ī، عملکرد کلاسیفایرگروهی را نشان می دهد. محور افقی در شکل، تعداد زیرکلاسیفایرهای همجوشانی شده را نشان می دهد. زمان آموزش کلاسیفایر هم جوشان که با ویژگی CC-PEV,=400 ,d_{red} L=31 ، تعریف شده است حدود 70ثانیه طول می کشد. در حالیکه آموزش کلاسیفایر G-SVM با همان ویژگی مشابه CC-PEV و با بهینه ترین مقادیر C,λ حدود ۳,۵ برابر طول می کشد. این نشان می دهد که سرعت عملکرد و یادگیری کلاسیفایر هم جوشان مورد استفاده در استگآنالیز بسیار سریعتر از G-SVM است.

تفاوت بین عملکرد کلاسیفایر گروهی و کلاسیفایر G-SVM وقتی بیشتر می شود که تعداد ابعاد ویژگی ها افزایش یابد. به عنوان مثال آموزش با کلاسیفایرهم جوشان اگر تعداد زیر ویژگی ها برابر ۵۰۰۰۰ تا باشد با $d_{red} = 2000$, L=90, تنها بیست دقیقه طول می کشد، در حالیکه آموزش همان تعداد مشابه زیر ویژگی ها با کلاسیفایر G-SVM با مقادیر بهینه C,λ حدود هفت ساعت زمان می برد.شکل ۵ نیز برتری متد کلاسیفایر هم جوشان در مقایسه با G-SVM رانشان میدهد. همچنین می توان گفت که عملکرد کلاسیفایر همجوشان کاملا وابسته به d_{red} است و هرچه d_{red} بیشتر باشد خطا کمتر می شود.

[15] **Error**
[16] **Soft margin**
[17] **payload**
[18] **Bit per non-zero AC. DCT**

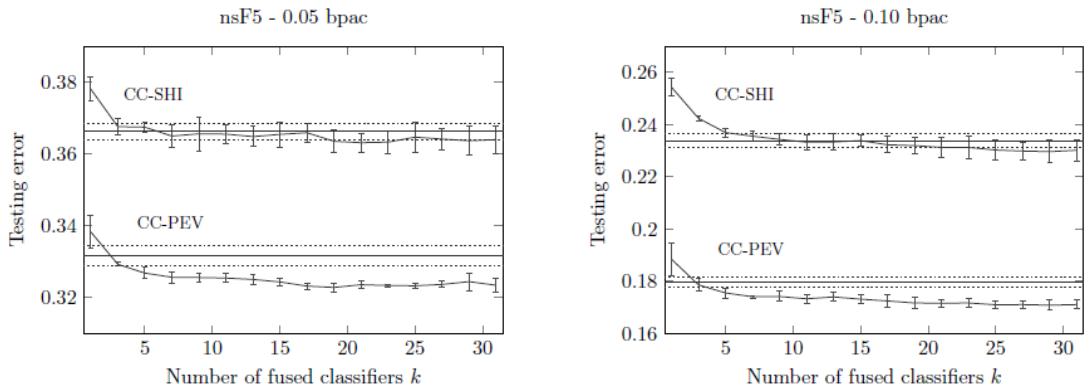
Prefeatures : CC-SHI

Payload	G-SVM	Ensemble class.
0.05	0.3662 ± 0.0022	0.3640 ± 0.0039
0.10	0.2339 ± 0.0028	0.2302 ± 0.0042
0.15	0.1194 ± 0.0035	0.1159 ± 0.0018
0.20	0.0595 ± 0.0014	0.0507 ± 0.0016

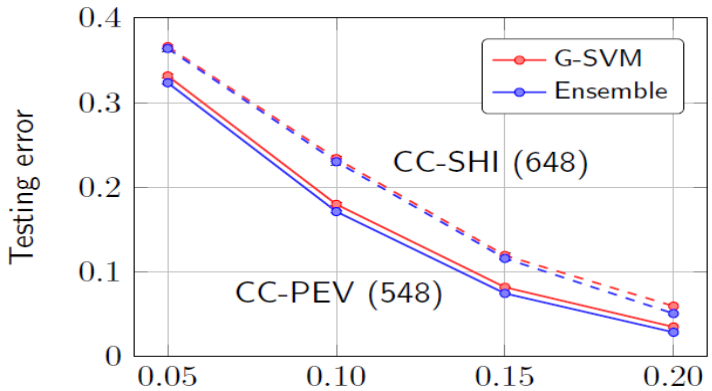
Prefeatures: CC-PEV

Payload	G-SVM	Ensemble class.
0.05	0.3316 ± 0.0028	0.3235 ± 0.0019
0.10	0.1798 ± 0.0018	0.1712 ± 0.0017
0.15	0.0818 ± 0.0008	0.0745 ± 0.0008
0.20	0.0349 ± 0.0011	0.0286 ± 0.0010

جدول ۱-خطای P_E بر روی الگوریتم nsF5 با کلاسیفیکیشن G-SVM و کلاسیفایر هم جوشان که $L=31, d_{red} = 400$ است.



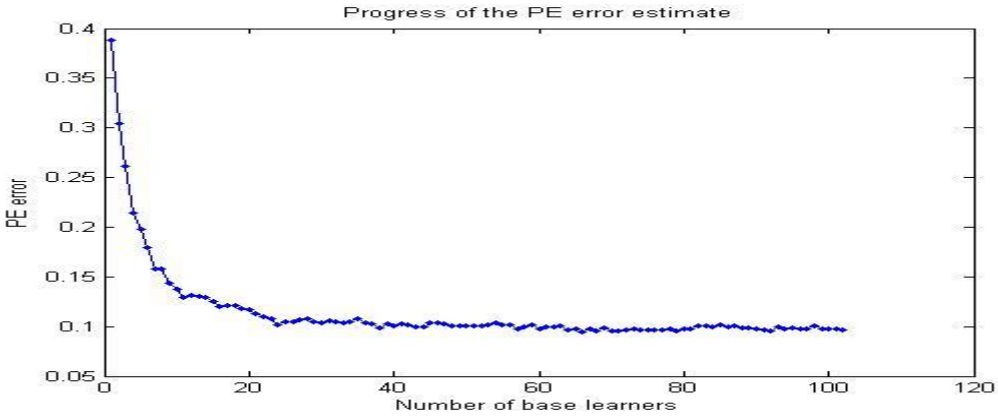
شکل ۴- سمت چپ: خطای p_E برای الگوریتم nsF5 با ظرفیت ترابری ۰٫۰۵bpac- - سمت راست: خطای p_E برای الگوریتم nsF5 با ظرفیت ترابری ۰٫۱۰bpac- کلاسیفایر همجوشان $d_{red} = 400$ در مقایسه با کلاسیفایر G-SVM – ویژگی ها: CC-PEV,CC-SHI-خط افقی مستقیم G-SVM است و خط نیمه کج با علامت I کلاسیفایر همجوشان است.



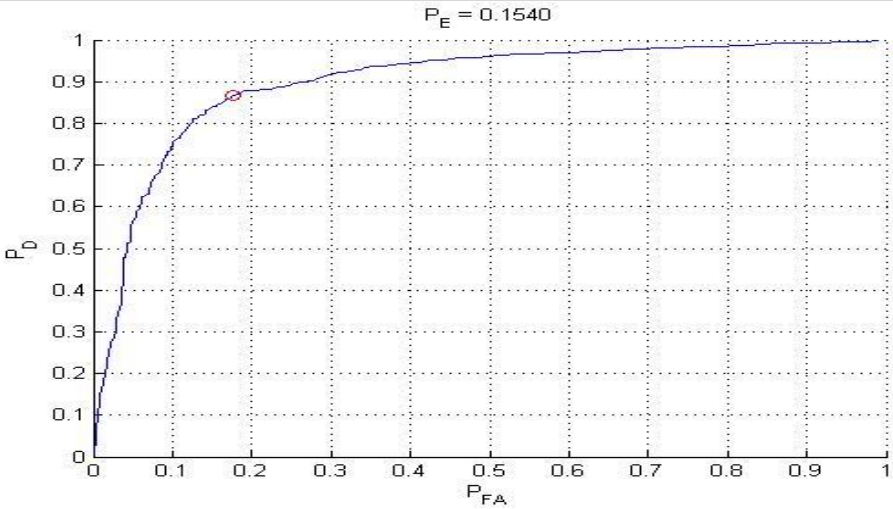
شکل۵- خطوط آبی و قرمز به ترتیب نمایانگر کلاسیفایر هم جوشان و G-SVMهستند.

علاوه بر آزمایشات انجام شده در مقاله [1,2]، آزمایش انجام شده توسط ما در محیط متلب۲۰۱۶ نشان می دهد که هرچقدر تعداد زیر کلاسیفایر ها بیشتر شود خطای PE کاهش می یابد. در این پیاده سازی هدف استگآنالیز تصاویر jpeg بوده است .ویژگی ها را از نوع cc-pev انتخاب کرده ایم و از الگوریتمnsF5 با ظرفیت ترابری ۰,۲ استفاده نموده ایم تا صحت و سقم نتایج بدست آمده در مقاله [1,2]را اثبات نماییم. در این آزمایش ۲۰۰۰ نمونه تصویر و ویژگی کاور و ۱۹۹۴ نمونه تصویر و ویژگی استگانوگرافی داریم. دو جفت ماتریس stego , cover داریم که حاوی تصاویر استگانوگرافی و تصاویر کاور هستند. در stego ، ماتریس names حاوی نام تصاویر با ابعاد ۱*۱۹۹۴ است. لازم به ذکر است که ۱۹۹۴ تصویر داریم که ۹۹۷ تای آنها در مجموعه تست و ۹۹۷ تای دیگر در مجموعه آموزش قرار دارند. ماتریس F ماتریس ویژگی ها با ابعاد ۵۴۸*۱۹۹۴ است که ردیف ها شامل نمونه تصاویر و ستون ها شامل ویژگی ها است که ویژگی ها در اینجا در فضای ۵۴۸ بعدی هستند.

در cover ، ماتریس names حاوی نام تصاویر با ابعاد ۱*2000 است. ماتریس F ماتریس ویژگی ها با ابعاد ۵۴۸*2000 است که ردیف ها شامل نمونه تصاویر و ستون ها شامل ویژگی ها است. نیمی از داده ها را به مجموعه تست و نیمی دیگر را به مجموعه آموزش اختصاص داده ایم و نهایتا الگوریتم کلاسیفایر هم جوشان را پیاده سازی میکنیم. در اینجا تعداد زیر کلاسیفایرها از بین ۱ تا ۱۲۰ تا تغییر می کند. شکل ۶ خروجی این آزمایش ما را نشان می دهند.به علاوه برای ارزیابی دقت متد فوق، نمودار احتمال تشخیص اشتباه- به نرخ تشخیص را رسم نموده ایم که بر روی الگوریتم nsF5 خروجی قابل قبولی دارد و بیانگر دقت بالای متد کلاسیفایر همجوشان است. شکل۷ این نمودار را نشان می دهد.



شکل 6- با فزایش زیر کلاسیفایرها در الگوریتم کلاسیفایر هم جوشان، خطا کاهش می یابد.



شکل ۷- نمودارنرخ تشخیص اشتباه- به نرخ تشخیص(اعمال شده بر الگوریتم ns F5)

۶.نتیجه گیری

در این مقاله روش های استگآنالیز مدرن در فضاهای ویژگی با ابعاد بالا را مرور کردیم. به دلیل داشتن ابعاد بالا و وابستگی های پیچیده بین پیکسل ها، طراحی استگآنالیزها در ابعاد بالا کار دشوارتری است. برای طراحی استگآنالیز دو موضوع وجود دارد. یکی اینکه چگونه از بین کل ویژگی ها، ویژگی هایی مناسب و کاربردی را انتخاب کنیم، دیگر اینکه با توجه به مشکل نفرین بعد، چگونه کلاسیفایر مناسب در فضای ویژگی با ابعاد بالا طراحی کنیم. افزایش بعد ، در یادگیری ماشین مشکل نفرین بعد را ایجاد می کند. نفرین بعد پیچیدگی یادگیری و آموزش را به سرعت افزایش می دهد و برای کلاسیفایر داده های دیده نشده موفق عمل نمی کند. یک روش برای پیاده سازی کلاسیفایر در ابعاد بالا استفاده از ماشین بردار پشتیبان کرنلی گوسین G-SVM بود که در این روش یادگیری بر روی یک مجموعه بزرگ از داده های کاور و استگانوگرافی انجام میشد. البته استفاده از SVM با کرنل گوسین، برای آموزش مجموعه ای از تصاویر در ابعاد بالا بسیار هزینه بر و کند بود. جهت حل این مشکل کلاسیفایر هم جوشان مطرح شد. تکیه ی اصلی ما در این مقاله، مرور این روش[1,2] به عنوان یک کلاسیفایر ساده و سریع و مقیاس پذیر بود. کلاسیفایر هم جوشان در اثر ترکیب چندین زیر کلاسیفایر به وجود می آمد که این زیر کلاسیفایر ها، خود کلاسیفایر های ضعیفی هستند که با انتخاب رندم تعدادی از ابعاد ویژگی از بین کل ابعاد فضای ویژگی ساخته شده اند. این زیر کلاسیفایرها دقت خوبی نداشتند ولی در عوض دارای تنوع زیاد در مجموعه داده بودند تا کل داده های دیده نشده را پوشش دهند. نتایج نشان داد که کلاسیفایر همجوشان پیچیدگی محاسباتی کمی دارد و مقیاس پذیر است و باعث می شود که بتوانیم از آن در فضاهای ویژگی ۴۰۰۰۰ بعدی، با مجموعه داده ای بالغ بر ۹۰۰۰۰ تصویر استفاده کنیم. همچنین آزمایشات انجام داده شده نشان داد که این کلاسیفایر بر روی الگوریتم nsF5عملکرد خوبی دارد.

مراجع

1. J.Kodovský and J.Fridrich. Steganalysis in high dimensions: Fusing classifiers built on random subspaces. The International Society for Optical Engineering · February 2011

2. J Kodovsky, Ensemble classification in steganalysis – Cross-validation and AdaBoost, Technical report, August 2011

3. T. N. Lal, O. Chapelle, J.Weston, and A. Elisseeff. Embedded methods. In I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, editors, Feature Extraction: Foundations and Applications, Studies in Fuzziness and SoftComputing, pages 137–165. Physica-Verlag, Springer, 2006.

4. Freund et al. A Short Introduction to Boosting, Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999.

5. J. Fridrich, T. Pevný, and J. Kodovský. Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In J. Dittmann and J. Fridrich, editors, Proceedings of the 9th ACM Multimedia & Security Workshop, pages 3–14, Dallas, TX, September 20–21, 2007.

6. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

7. Pevny et al. Using high dimensional image models to perform highly undetectable steganography, Springer, International Workshop on Information Hiding, Berlin, 2010

8. R. Cogranne et al. Theoretical model of the FLD ensemble classifier based on hypothesis testing theory, Workshop on Information Forensics and Security (WIFS), 2014 IEEE

9. F. Pereira and G. Gordon. The support vector decomposition machine. In Proceedings of the 23rd interna-tional conference on Machine learning, ICML ’06, pages 689–696, Pittsburgh, PA, 2006.