انجمن رمز ایران
Iranian Society of Cryptology
چهارمین کنفرانس انجمن رمز ایران
۲۴ الی ۲۶ مهر ماه ۱۳۸۶ دانشگاه علم و صنعت ایران

ارزیابی و مقایسه چهار روش کاهش بعد ویژگیها برای سیستم تشخیص نفوذ مبتنی بر ماشین بردار پشتیبان

محمد مهدی همایونپور
دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر
homayoun@aut.ac.ir

حمید رضا شجاع مودب
پژوهشکده پردازش هوشمند علائم
hr_moaddab@yahoo.com

چکیده: در این مقاله چهار روش تبدیلی کاهش بعد ویژگیها برای سیستم تشخیص نفوذ مبتنی بر SVM مقایسه و ارزیابی می‌شود. این چهار روش شامل تحلیل مولفه‌های اساسی یا PCA، تحلیل مولفه‌های مستقل یا ICA، تحلیل الگوی متمایز خطی یا LDA و نهایتا شبکه عصبی MLP می‌باشد. در این مقاله ما از داده‌های برنامه ارزیابی تشخیص نفوذ DARPA استفاده می‌کنیم که هر یک از رکوردهای این پایگاه داده شامل ۴۱ ویژگی می‌باشد. روش کار به این صورت است که ابتدا با استفاده از هر چهار روش، بعد ویژگیهای رکوردها را به ۱۰ کاهش می‌دهیم و سپس مدت زمان آموزش، آزمایش و درصد شناسائی حملات توسط سیستم تشخیص نفوذ مبتنی بر SVM برای ویژگیهای مربوط به هر یک از چهار روش و نیز حالت ۴۱ ویژگی را بدست آورده و با هم مقایسه می‌کنیم. مقایسه نتایج نشان داد که تکنیک LDA زمان آزمایش و آموزش کمتری دارد و نسبت به حالت ۴۱ مولفه‌ای زمان آزمایش با روش مذکور حدود ۲۱٪ کمتر می‌باشد ضمنا با LDA سیستم تشخیص نفوذ به درصدهای بالاتری در شناسائی حملات دست یافت. بنابراین ۱۰ ویژگی که به روش LDA بدست آمده از روشهای دیگر دارای اطلاعات مفید بیشتری می‌باشند و میزان اطلاعات مفیدی که در این روش کاهش بعد از دست می‌رود کمتر از سه روش دیگر است.

واژه های کلیدی: سیستمهای تشخیص نفوذ، شبکه‌های کامپیوتری، کاهش بعد داده‌ها، تحلیل مولفه‌های اساسی، تحلیل مولفه‌های مستقل، تحلیل الگوی متمایز خطی

۱- مقدمه

انتخاب ویژگیها یکی از موضوعات مهم در تشخیص نفوذ می‌باشد. از میان انبوه ویژگیهائی که می‌توان برای تشخیص نفوذ در نظر گرفت باید مشخص کرد که کدام یک از آنها واقعا مفید هستند. حذف ویژگیهای بی‌فایده در تشخیص نفوذ که عملا منجر به کاهش بعد آنها می‌شود باعث بالا رفتن سرعت محاسبات و نهایتا کارائی سیستم تشخیص نفوذ می‌شود. بطورکلی می‌توان روشهای کاهش بعد ویژگی را به دو گروه روشهای انتخابی و روشهای تبدیلی تقسیم‌بندی کرد. در

روشهای انتخابی، ویژگیهائی از یک بردار انتخاب می‌شود که در شناسائی یک حمله یا دسته‌ای از حملات موثرتر باشند و از اعمال همه ویژگیهای بردار مذکور برای شناسائی آن حمله اجتناب می‌شود[۲]. در روشهای تبدیلی سعی می‌شود با روشهای ریاضی اطلاعات مفید بردار ویژگی برای شناسائی حمله مورد نظر استخراج شود. در این مقاله ما چهار روش کاهش بعد تبدیلی شامل تحلیل مولفه‌های اساسی[۱] یا PCA، تحلیل مولفه‌های مستقل[۲] یا ICA، تحلیل الگوی متمایز خطی[۳] یا LDA و نهایتا MLP[4] را مورد بررسی قرار می‌دهیم و با اعمال داده‌های کاهش بعد داده شده با هر یک از چهار روش مذکور به یک سیستم تشخیص نفوذ مبتنی بر ماشین بردار پشتیبان میزان دقت و سرعت سیستم مذکور را در شناسائی انواع مختلف نفوذ با حالتی که داده‌ها کاهش بعد داده نشده‌اند مقایسه می‌کنیم و در نهایت بهترین روش تبدیلی کاهش بعد را از بین این چهار روش انتخاب کرده و با روش کاهش بعد انتخابی ارائه شده در مرجع [۲] مقایسه می‌کنیم. به منظور بررسی و ارزیابی اثر روشهای کاهش بعد بر روی سیستم تشخیص نفوذ مبتنی بر SVM نیاز به داده‌هایی است که بخشی از آن برای آموزش و بخش دیگر برای آزمایش بکار رود، بدین منظور از پایگاه داده KDDCUP99 [۳] استفاده شده که عملا نسخه‌ای از پایگاه داده تهیه شده بوسیله برنامه ارزیابی تشخیص نفوذ DARPA می‌باشد.

در بخش دوم به معرفی چهار تکنیک تبدیلی کاهش بعد می‌پردازیم. در بخش سوم اقدام به پیاده‌سازی چهار روش کاهش بعد و مقایسه آنها با یکدیگر می‌کنیم و در بخش چهارم به مقایسه تکنیکهای کاهش بعد انتخابی و تبدیلی می‌پردازیم.

۲- روشهای کاهش بعد

در این بخش به توضیح مختصری در مورد چهار نوع از روشهای کاهش بعد داده‌ها می‌پردازیم که این چهار روش عبارتند از PCA، ICA، LDA و MLP.

۲-۱- PCA یا تحلیل مولفه‌های اساسی

بردار x را در نظر می‌گیریم بدون اینکه مشکلی به طور عام ایجاد شود می‌توان فرض کرد که مولفه‌های آن دارای میانگین صفر هستند و اگر نباشد به راحتی میانگین را حساب و از آنها کم می‌کنیم. حال اگر فرض کنیم Y ماتریس کوواریانس x است و A ماتریس بردارهای ویژه و B ماتریس مقادیر ویژه محاسبه شده از ماتریس کوواریانس Y باشد و هر عضو A را با ϕ و هر عضو B را با λ نمایش می‌دهیم آنگاه ماتریس مولفه‌های اصلی P شامل بردارهای ویژه متناظر با m مقدار ویژه بزرگتر می‌باشد.

$$\lambda_1 > \lambda_2 > \ldots > \lambda_m > \ldots > \lambda_n$$
$$P = [\phi_1, \phi_2, \ldots, \phi_m] \qquad (۱)$$

بردارهای ویژگی جدید به کمک رابطه زیر بدست می‌آیند:

$$C = P.x \qquad (۲)$$

۲-۲- ICA یا تحلیل مولفه‌های مستقل

فرض کنیم که $x_1, \ldots, x_n$ n ترکیب خطی از n مولفه مستقل را داریم.

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \ldots + a_{jn}s_n \qquad (۳)$$

که همه متغیر با زمان t هستند[۵]. فرض کنیم که x_j و s_k متغیرهای تصادفی باشند. در نتیجه $x_j(t)$ یک نمونه از این متغیر تصادفی است. بدون اینکه مشکلی به طور عام ایجاد شود می‌توانیم فرض کنیم که هر دو دارای میانگین صفر هستند و اگر نباشد به راحتی میانگین را حساب و از آنها کم می‌کنیم. فرض کنید x برداری شامل $x_1, \ldots, x_n$ و s بردار شامل $s_1, \ldots, s_n$ و A ماتریس ضرائب a_{ij} می‌باشد. فرض بر این است که ماتریسها ستونی هستند از این رو ترانهاده x، یک ماتریس سطری است. بنابراین عبارت بالا را می‌توان به این صورت نشان داد:

$$x = As \qquad (۴)$$

اگر ستون ماتریس A را بخواهیم، **aj** به عنوان یک ستون ماتریس A خواهد بود که داریم:

$$x = \sum_{i=1}^{n} a_j s_i \qquad (۵)$$

مدل آماری معادله ۳ مدل آنالیز مولفه‌های مستقل یا ICA خوانده می‌شود که بیان کننده این است که چگونه داده‌های بدست آمده از روی مولفه‌های مستقل به صورت ترکیبی ساخته می‌شوند. مولفه‌های مستقل به عنوان متغیرهای پنهان هستند، یعنی اینکه بطور مستقیم دیده نمی‌شوند. همچنین ماتریس ترکیب معلوم نیست و تنها چیزی که ما در دست داریم مقادیر مشاهده x می‌باشد که مقادیر A و s از روی x تخمین زده می‌شوند. نقطه شروع برای تخمین A و s این فرض ساده است که مولفه‌های s_i از لحاظ آماری مستقل هستند. هنگامی که ماتریس A تخمین زده شد معکوس آن محاسبه می‌شود و با استفاده از آن مولفه‌های مستقل به صورت زیر محاسبه می‌شود.

$$\mathbf{s=Wx} \qquad (۶)$$

که W معکوس A است. حال می‌توان ماتریس W را به گونه‌ای انتخاب کرد تا کاهش بعد نیز داشته باشیم.

۲-۳- مقدمه‌ای بر طرح MLP به عنوان یک روش کاهش بعد

یک شبکه MLP را در نظر بگیرید که دارای R_1 ورودی، S_1 نرون در لایه اول، S_2 نرون در لایه دوم و S_3 نرون در خروجی می‌باشد. حال اگر از الگوریتم پس انتشار خطا استفاده کنیم و داده‌های ورودی و خروجی را یکسان و برابر R_1 قرار دهیم و خروجی لایه دوم نیز دارای بعد کمتری از ورودی باشد در این صورت می‌توانیم خروجی لایه دوم شبکه آموزش دیده مفروض را با فرضهای مذکور به عنوان ویژگی کاهش بعد داده شده در نظر بگیریم و به این ترتیب از MLP برای کاهش بعد استفاده کنیم.

۲-۴- LDA یا تحلیل الگوی متمایز خطی

فرض کنید که S_b ماتریس پراکندگی بین کلاسها و S_w ماتریس پراکندگی درون کلاسی برای مساله شناسایی الگو M کلاسه باشد. در اینصورت S_b بیانگر تغییرات بردارهای امید برای هر زوج کلاس می‌باشد در حالیکه S_w پراکندگی نمونه‌ها حول بردار امید کلاس آنها می‌باشد. در روش LDA سعی می‌شود بطور همزمان $Det(S_w)$ حداقل و $Det(S_b)$ حداکثر گردد.

$$Max\frac{|S_b|}{|S_w|} = Max\left|S_w^{-1}S_b\right| \qquad (۷)$$

یکی از مناسبترین معیارها جهت بیان تفکیک‌پذیری کلاسها استفاده از تریس[۵] ماتریس تمایز $S_w^{-1}S_b$ می‌باشد که بصورت زیر تعریف می‌شود:

$$T(n) = tr(S_w^{-1}S_b) \qquad (۸)$$

هدف در آنالیز LDA کاهش بعد بردارهای ویژگی از بعد اولیه (n) به بعد جدید (m) به کمک ماتریس انتقال $m\times n$ بعدی A می‌باشد. برای دستیابی به ماتریس A بعد جدید به گونه‌ای تعیین می‌شود که $T(m)$ حداکثر گردد. می‌توان نشان داد که این مطلب معادل بدست آوردن m بردار ویژه اول ماتریس تمایزدهندگی $S_w^{-1}S_b$ می‌باشد. جهت دستیابی به این بردارها کافیست مقادیر ویژه $S_w^{-1}S_b$ را به ترتیب نزولی مرتب نمود. بردارهای ویژه متناظر با m مقدار ویژه اول بردارهای مدنظر می‌باشند.

$$\lambda_1 > \lambda_2 > \ldots > \lambda_m > \ldots > \lambda_n$$
$$A = [\phi_1, \phi_2, \ldots, \phi_m] \qquad (۹)$$

بردارهای ویژگی جدید به کمک رابطه زیر بدست می‌آیند:

$$y = Ax \qquad (۱۰)$$

ماتریس پراکندگی نمونه‌های جدید $\tilde{S}_w$ و $\tilde{S}_b$ هر دو قطری می‌باشند و بیانگر آن می‌باشند که ضرایب جدید غیرهمبسته می‌باشند.

۳- پیاده‌سازی و مقایسه تکنیکهای کاهش بعد

در این بخش با پیاده‌سازی چهار الگوریتم PCA، ICA، LDA و MLP بعد داده‌ها را کاهش می‌دهیم و داده‌های حاصل را به یک سیستم تشخیص نفوذ مبتنی بر SVM اعمال می‌کنیم و نتایج را با هم مقایسه می‌کنیم. برای سهولت کار ابتدا داده‌های آموزشی و آزمایشی را طبق جدول ۱ از مجموعه داده‌های آموزشی و آزمایشی پایگاه داده KDDCUP99 بصورت اتفاقی انتخاب می‌کنیم[۳].

همانطور که در جدول نیز نشان داده شده است در این پایگاه داده حملا ت به چهار دسته حملات از کار انداختن

سرویس (Denial of Service: DOS)، حملاتی که برای شناسائی نقاط آسیب‌پذیر هدف انجام می‌شود و به آن حملات پویش (Probe) می‌گویند، حملات دسترسی غیر مجاز از ماشین دور (Remote To Local: R2L) و نهایتا دسترسی غیر مجاز به اختیارات ویژه (User To Root) تقسیم می‌شوند. ضمنا بخشی از داده‌ها نیز مربوط به رفتارهای عادی می‌باشند. هر رکورد شامل ٤١ ویژگی مربوط به یک ارتباط می‌باشد. از جمله این ویژگیها می‌توان به طول زمان ارتباط، نوع پروتکل مورد استفاده، تعداد بایتهای منتقل شده، پرچم نشان دهنده وضعیت عادی یا غیر عادی ارتباط، تعداد ارتباط به یک میزبان توسط اتصال جاری در طی ۲ ثانیه گذشته و مانند اینها اشاره کرد[۴].

جدول ۱: تعداد رکوردهای آزمایشی و آموزشی

	داده‌های آموزشی	داده‌های آزمایشی
Normal	۲۰۰	۲۰۰
Probe	۲۰۰	۲۰۰
DOS	۲۰۰	۲۰۰
R2L	۲۰۰	۲۰۰
U2R	۵۲	۲۰۰
جمع	۸۵۲	۱۰۰۰

با استفاده از مجموعه آموزش و آزمایش جدید طبقه‌بندی کننده‌های بهینه SVM مربوط به هر دسته را با تمام ۴۱ ویژگی مربوط به هر رکورد پیدا می‌کنیم سپس با کاهش بعد داده‌ها بوسیله هر یک از چهار الگوریتم فوق‌الذکر طبقه‌بندی کننده بهینه SVM را برای داده‌های کاهش بعد داده شده نیز پیدا کرده و اقدام به مقایسه نتایج آنها می‌کنیم. هدف ما از مقایسه چهار روش کاهش بعد مورد نظر تعیین میزان اطلاعات مفیدی است که بعد از کاهش بعد برای شناسائی نفوذ وجود دارد و این مقدار بر اساس نرخ شناسائی نفوذ اندازه‌گیری می‌شود. بنابراین با چهار روش فوق‌الذکر بعد بردارهای ویژگی را از ۴۱ به ۱۰ کاهش می‌دهیم کاهش بعد یکسان برای چهار روش کار مقایسه میزان اطلاعات مفید در ویژگیهای مربوط به هر روش کاهش بعد، را ساده‌تر می‌کند. علت کاهش بعد به ۱۰ این است که در تحقیقات مرتبط با کاهش بعد انتخابی [۲] بطور متوسط بعد به ۱۳ مولفه کاهش پیدا کرده است بنابراین و با توجه به تئوری حاکم بر روشهای تبدیلی کاهش بعد می‌توانیم بگوئیم که کاهش بعد با روشهای تبدیلی نیز نباید با این مقدار تفاوت چندانی داشته باشد بر این اساس کاهش بعد تا ۱۰ مولفه را مورد توجه قرار دادیم. البته نتایج آزمایشات انجام شده نیز صحت این ایده را تائید می‌کند. با استفاده از داده‌های با بعد کمتر نرخ شناسائی پنج دسته (Normal, DOS, R2l, U2R, Probe) و نیز زمان آموزش و آزمایش SVM مربوطه را بدست می‌آوریم.

۳-۱- شرح پیاده‌سازی

سیستم تشخیص نفوذ ما مبتنی بر SVM است که بدین منظور از جعبه‌ابزار OSU SVM Classifier (ver 3.00) استفاده کردیم. ضمنا با استفاده از تابع Scale جعبه‌ابزار مذکور ویژگیهای مشابه بردارهای ویژگی را در بازه (۰،۱) نرمال کردیم.

در ابتدا ۸۵۲ رکورد آموزش و ۱۰۰۰ رکورد آزمایش با ۴۱ ویژگی مربوط به هر رکورد در آموزش و آزمایش پنج SVM مربوط به پنج دسته مختلف مورد استفاده قرار گرفت. مثلا SVM مربوط به دسته DOS دو دسته DOS و غیر آن را طبقه‌بندی می‌کند و عملا یک طبقه‌بندی کننده دو کلاسه است چهار SVM مربوط به چهار دسته دیگر نیز به همین ترتیب است.

جدول ۲: نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل ۴۱ ویژگی

دسته	شناسائی دسته (٪)	شناسائی غیر دسته(٪)	نرخ شناسائی(٪)	مدت زمان آموزش(ثانیه)	مدت زمان آزمایش(ثانیه)
Normal	۷۳	۹۲/۵	۸۸/۶	۰/۱۴	۰/۰۹
Probe	۸۷/۵	۹۹/۶۲	۹۷/۱۹	۰/۱۲	۰/۱۳
DOS	۹۶	۹۹/۳۸	۹۸/۷	۰/۱۲	۰/۱۳
R2L	۳/۵	۹۹/۲۵	۸۰/۱	۰/۰۹	۰/۰۶
U2R	۱۹/۵	۹۸/۵	۸۲/۷	۰/۰۷	۰/۰۵

جدول ۲ نتایج مربوط به طبقه‌بندی کننده SVM را با ٤١ ویژگی نشان می‌دهد ضمنا زمانها میانگین ۱۰ بار اجرا می‌باشند. با توجه به جدول، میانگین زمان آموزش برابر است با ۰/۱۰۸ ثانیه

و میانگین زمان آزمایش برابر است با ۰/۰۹۲ ثانیه است. ضمنا صحت تشخیص دسته عبارتست از تعداد حملات مربوط به دسته که درست شناسائی شده تقسیم بر کل حملات مربوط به آن دسته، خطای مثبت عبارتست از رفتارهای نرمال که حمله شناسائی شده‌اند تقسیم بر کل رکوردهای نرمال. نرخ آشکارسازی نیز برابر است با حملات مربوط به دسته و غیر دسته که درست شناسائی شده تقسیم بر کل حملات دسته و غیر دسته.

برای پیاده‌سازیPCA از جعبه‌ابزار Netlab استفاده کردیم. روش کار به این صورت است که ماتریس کوواریانس بردارهای آموزشی یک دسته از حملات مثلا DOS را بدست آوردیم که یک ماتریس ۴۱×۴۱ شد و سپس بردارهای ویژه متناظر با ۱۰ مقدار ویژه بزرگتر مربوط به مقادیر ویژه ماتریس کوواریانس فوق را انتخاب کردیم که نتیجه آن یک ماتریس ۱۰×۴۱ شد و عملا ماتریس مولفه‌های اصلی ما بدست آمد و سپس ترانهاده آن را در ترانهاده کل داده‌های آموزشی که یک ماتریس ۸۵۲×۴۱ است ضرب کردیم و داده‌های کاهش بعد داده شده آموزشی که یک ماتریس ۱۰×۸۵۲ بود را به SVM آموزش دادیم. در فاز آزمایش نیز همان ماتریس مولفه‌های اصلی را به همان ترتیب فاز آموزش در مجموعه آزمایش ضرب کردیم و داده‌های کاهش بعد داده شده را با SVM آزمایش کردیم. در ادامه این کار را برای چهار دسته دیگر نرمال، Probe، R2L، U2R تکرار کردیم. جدول ۳ شامل نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شاملPCA با ۱۰ ویژگی می‌باشد. میانگین زمان آموزش برابر است با ۰/۱۰۲ ثانیه و میانگین زمان آزمایش برابر است با ۰/۰۷٤ ثانیه است.

جدول ۳: نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل ۱۰ ویژگی با PCA

دسته	شناسائی دسته(٪)	شناسائی غیر دسته(٪)	نرخ شناسائی(٪)	مدت زمان آموزش(ثانیه)	مدت زمان آزمایش (ثانیه)
Normal	۸۶	۷۷/۵	۷۹/۲	۰/۱۱	۰/۰۸
Probe	۸۱	۹۹/۶۲	۹۵/۸۹	۰/۱۲	۰/۰۸
DOS	۹۷	۹۸/۸۷	۹۸/۴۹	۰/۰۸	۰/۰۴
R2L	۳	۹۹/۸۷	۸۰/۴۹	۰/۱۴	۰/۱۳
U2R	۹/۵	۱۰۰	۸۱/۹	۰/۰۶	۰/۰۴

برای پیاده‌سازی ICA از جعبه‌ابزار Fastica استفاده کردیم. بدین منظور برای هر یک از پنج دسته نرمال، Probe، DOS، R2L، U2R ماتریس ویژگیهای مستقل را بدست آوردیم که پنج ماتریس به ازاء هر دسته حاصل شد و چون تعداد ویژگیهای مستقل را ۱۰ ویژگی تعیین کردیم این ماتریسها ۴۱×۱۰ می‌باشد که با استفاده از آنها اقدام به کاهش بعد داده‌های آموزشی و آزمایشی شبیه عملیاتی که در مورد PCA شرح داده شده کردیم.جدول ۴ شامل نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل ICA با ۱۰ منبع مستقل می‌باشد. همانطور که ملاحظه می‌شود میانگین زمان آموزش برابر است با ۰/۸۱٤ ثانیه و میانگین زمان آزمایش برابر است با ۰/۰۷٤ ثانیه است.

جدول۴: نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل ۱۰ ویژگی با ICA

دسته	شناسائی دسته (٪)	شناسائی غیر دسته(٪)	نرخ شناسائی(٪)	مدت زمان آموزش(ثانیه)	مدت زمان آزمایش (ثانیه)
Normal	۸۵/۵	۸۰/۶۲	۸۱/۵۹	۰/۴۴	۰/۰۸
Probe	۸۲/۵	۹۹/۶۲	۹۶/۱۹	۰/۷	۰/۰۸
DOS	۹۶	۹۸/۱۲	۹۷/۶۹	۱/۸۵	۰/۰۷
R2L	۶	۹۹	۸۰/۴	۰/۲	۰/۱
U2R	۱۰	۹۹/۸۷	۸۱/۸۹	۰/۸۸	۰/۰۴

برای بدست آوردن ماتریس کاهش بعد با استفاده از MLP یک شبکه پرسپترون با سه لایه طرح شد که ۴۱ نرون در لایه ورودی و خروجی و ۱۰ نرون در لایه پنهان آن بود. با استفاده از الگوریتم پس انتشار خطا اقدام به آموزش شبکه عصبی مذکور کردیم که داده‌های آموزشی و هدف مشابه و متعلق به هر یک از پنج دسته نرمال، Probe، DOS، R2L، U2R از مجموعه آموزش بودند. پس از آموزش شبکه عصبی حاصل ضرب ترانهاده ماتریس اوزان ورودی که یک ماتریس ۴۱×۴۱ بود در ترانهاده ماتریس اوزان لایه پنهان که یک ماتریس ۴۱×۱۰ بود را بدست آوردیم که این کار را برای هر یک از پنج دسته بطور جداگانه انجام دادیم که منتج به پنج ماتریس ۴۱×۱۰ کاهش بعد مربوط به هر یک از پنج دسته شد. در ادامه با ضرب ماتریس کاهش بعد مربوط به هر دسته در داده‌های آموزشی و

آزمایشی بعد داده‌های مذکور را کاهش دادیم و داده‌های جدید را برای آموزش و آزمایش ماشین بردار پشتیبان مورد استفاده قرار دادیم. جدول ۵ شامل نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای کاهش بعد داده شده به ۱۰ ویژگی با MLP می‌باشد. میانگین زمان آموزش برابر است با ٤/۳۸ ثانیه و میانگین زمان آزمایش برابر است با ۰/۰۷٦ ثانیه است.

جدول۵: نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل ۱۰ ویژگی با MLP

دسته	شناسائی دسته (٪)	شناسائی غیر دسته(٪)	نرخ شناسائی (٪)	مدت زمان آموزش (ثانیه)	مدت زمان آزمایش (ثانیه)
Normal	۸۴	۷۷/۳۷	۷۸/۶۹	۴/۶	۰/۱۷
Probe	۸۳/۵	۹۷/۵	۹۴/۷	۴/۴	۰/۰۴
DOS	۹۴	۹۹/۶۲	۹۸/۴۹	۴/۴	۰/۰۹
R2L	۱/۵	۹۸/۳۷	۷۸/۹۹	۴/۲	۰/۰۴
U2R	۲۲	۹۷/۶۲	۸۲/۴۹	۴/۳	۰/۰۴

برای پیاده‌سازی LDA از تابع ldatrace جعبه ابزار Voicebox استفاده کردیم به این ترتیب که ماتریس کوواریانس کل داده‌های آموزشی را به عنوان ماتریس کوواریانس بین کلاسها به تابع Ldatrace دادیم و برای آرگومانهای دیگر از مقادیر پیش فرض تابع مذکور استفاده کردیم و چون ماتریس کوواریانس درون کلاسها در مقادیر پیش‌فرض این تابع یک ماتریس همانی است بنابراین در آزمایشات انجام شده عملا از تکنیک PCA به یک طریق دیگر استفاده کرده‌ایم.

خروجی تابع شامل ماتریس بردارهای ویژه است که با توجه به مقادیر ویژه آنها مرتب شده‌اند که ما ۱۰ ستون اول این ماتریس را انتخاب کردیم. در ادامه ترانهاده ماتریس کاهش بعد بدست آمده را در ترانهاده ماتریس داده‌های آزمایشی و آموزشی ضرب کردیم و مجموعه‌های آموزش و آزمایش کاهش بعد داده شده حاصل شد. در مرحله بعد با داده‌های بدست آمده پنج SVM مربوط به پنج دسته نرمال، Probe، DOS، R2L و U2R را آموزش داده و سپس آزمایش کردیم که در جدول ۶ نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل LDA با ۱۰ ویژگی می‌باشد. نکته‌ای که در مورد LDA قابل ذکر است این است که عمل کاهش بعد داده‌ها برای هر پنج دسته تنها یک بار و روی کل داده‌های آموزشی انجام می‌شود و نه تک تک بر روی داده‌های مربوط به هر دسته که زمان آن ۰/۰۴ ثانیه است. بنابراین برای بدست آوردن میانگین زمان آموزش جمع زمانهای آموزش در جدول ۶ با این مقدار ۰/۰۴ ثانیه را تقسیم بر ۵ می‌کنیم که برابر با ۰/۱ ثانیه می‌شود و میانگین زمان آزمایش برابر با ۰/۰۷۲ ثانیه است.

جدول ۶: نتایج حاصل از طبقه‌بندی کننده‌های SVM با ورودیهای شامل ۱۰ ویژگی با LDA

دسته	شناسائی دسته(٪)	شناسائی غیر دسته (٪)	نرخ شناسائی (٪)	مدت زمان آموزش (ثانیه)	مدت زمان آزمایش (ثانیه)
Normal	۸۳/۵	۸۱/۲۵	۸۱/۷	۰/۱۴	۰/۰۸
Probe	۹۰/۵	۹۹/۶۲	۹۷/۸	۰/۰۷	۰/۰۸
DOS	۹۵	۹۹/۵	۹۸/۶	۰/۱	۰/۰۸
R2L	۲	۹۹/۸۷	۸۰/۳	۰/۰۹	۰/۰۸
U2R	۵۲/۵	۹۹/۳۷	۹۰	۰/۰۶	۰/۰۴

۳-۲- مقایسه روشهای کاهش بعد تبدیلی

در جدول ۷ نرخ آشکارسازی SVM با ورودیهائی که از روشهای مختلف کاهش بعد بدست آمده آورده شده است.

جدول ۷: مقایسه نرخ آشکارسازی SVM با هریک از روشهای کاهش بعد بر حسب درصد

	نرخ شناسائی با ۴۱ ویژگی اصلی	نرخ شناسائی با ICA	نرخ شناسائی با PCA	نرخ شناسائی با MLP	نرخ شناسائی با LDA
Normal	۸۸/۶	۸۱/۵۹	۷۹/۲	۷۸/۶۹	۸۱/۷
Probe	۹۷/۱۹	۹۶/۱۹	۹۵/۸۹	۹۴/۷	۹۷/۸
DOS	۹۸/۷	۹۷/۶۹	۹۸/۴۹	۹۸/۴۹	۹۸/۶
R2L	۸۰/۱	۸۰/۴	۸۰/۴۹	۷۸/۹۹	۸۰/۳
U2R	۸۲/۷	۸۱/۸۹	۸۱/۹	۸۲/۴۹	۹۰
میانگین	۸۹/۴۶	۸۷/۵۵	۸۷/۱۹	۸۶/۶۸	۸۹/۶۸

همانطور که ملاحظه می‌شود SVM با ۴۱ ویژگی بالاترین نرخ شناسائی را برای دسته نرمال و DOS دارد و بعد از آن بهترین نرخ شناسائی برای این دو دسته با ویژگیهای حاصل از LDA بدست آمده. در شناسائی دسته Probe و U2R بالاترین نرخ شناسائی با ویژگیهای حاصل از LDA حاصل شده است. در شناسائی R2L بهترین نتیجه با ویژگیهای حاصل از PCA بدست آمده. در مجموع ویژگیهای حاصل از تکنیک LDA در شناسائی چهار دسته نرمال، Probe، DOS، U2R موثرتر از ویژگیهای سه روش کاهش بعد دیگر است و در شناسائی R2L ویژگیهای حاصل از PCA موثرتر می‌باشد.

در جدول۸ میانگین زمان آموزش و آزمایش SVM با استفاده از پنج دسته ویژگی مذکور ارائه شده است. همانطور که ملاحظه می‌شود بهترین زمان آموزش و آزمایش برای LDA می‌باشد. بطور کلی با توجه به دو جدول۷ و ۸ می‌توان نتیجه‌گیری کرد که تکنیک LDA از روشهای دیگر کاهش بعد مناسب‌تر است. نکته‌ای که به خوبی از جدول ۸ حاصل می‌شود این است که زمان آزمایش برای هر چهار روش کاهش بعد تبدیلی تقریبا یکسان است و آن به این دلیل است که سیستم تشخیص نفوذ با رکوردهای با ۱۰ مولفه در هر چهار مورد کار آزمایش را انجام داده است. بنابراین منطقی است که زمان آزمایش هر چهار روش تقریبا یکسان باشد. ولی چون الگوریتم استخراج ۱۰ مولفه از میان ۴۱ مولفه مربوط به هر یک از چهار روش با هم از نظر حجم محاسباتی متفاوت می‌باشد بنابراین مدت زمان آموزش برای آنها نیز متفاوت شده است.

جدول ۸: مقایسه زمان آموزش و آزمایش SVM با هریک از روشهای کاهش بعد

	نرخ شناسائی با ۴۱ ویژگی اصلی (٪)	نرخ شناسائی بروش ICA (٪)	نرخ شناسائی بروش PCA (٪)	نرخ شناسائی بروش MLP (٪)	نرخ شناسائی بروش LDA (٪)
میانگین زمان آموزش(ثانیه)	۰/۱۰۸	۰/۸۱۴	۰/۱۰۲	۴/۳۸	۰/۱
میانگین زمان آزمایش(ثانیه)	۰/۰۹۲	۰/۰۷۴	۰/۰۷۴	۰/۰۷۶	۰/۰۷۲

۳-۳ تعداد بهینه ویژگیهای LDA

با آزمایشات بالا روشن شد که تکنیک LDA از سه روش دیگر کاهش بعد تبدیلی بهتر عمل می‌کند. در ادامه سیستم تشخیص نفوذ مبتنی بر SVM را برای هر پنج دسته نرمال، Probe، DOS، R2L و U2R مجددا با داده‌های کاهش بعد داده شده LDA مورد آزمایش قرار دادیم با این تفاوت که به جای ۱۰ ویژگی سعی کردیم تعداد بهینه را برای هر دسته پیدا کنیم که نتایج در جدول ۹ ارائه شده است. با توجه به جدول ۹ میانگین زمان آموزش ۰/۱ ثانیه است که حاصل جمع زمانهای آموزش جدول ۹ و مدت زمان تبدیل داده‌ها به ضرائب LDA که ۰/۰۴ ثانیه است تقسیم بر ۵ می‌باشد. میانگین زمان آزمایش نیز ۰/۰۷۲ ثانیه می‌باشد. نهایتا از جدول ۸ و ۹ می‌توان نتیجه گرفت که LDA باعث کاهش زمان آموزش به میزان ۷/۴٪ و کاهش زمان آزمایش به میزان ۲۱/۷۳٪ می‌شود و نرخ شناسائی با داده‌های مربوط به LDA در دسته‌های مربوط به Probe، U2R، R2L بهتر از ویژگیهای با ۴۱ مولفه است. ضمنا تعدا ویژگیها نیز بطور متوسط به ۹ مولفه کاهش یافته است.

جدول ۹: تعداد بهینه ویژگیهای LDA

دسته	تعداد ویژگی	نرخ شناسائی (٪)	مدت زمان آموزش (ثانیه)	مدت زمان آزمایش (ثانیه)
Normal	۱۲	۸۲/۷	۰/۱۴	۰/۰۷
Probe	۱۰	۹۷/۸	۰/۰۷	۰/۰۸
DOS	۷	۹۸/۷	۰/۱	۰/۰۹
U2R	۷	۸۰/۳	۰/۰۹	۰/۰۸
R2L	۱۰	۹۰	۰/۰۶	۰/۰۴

۴- مقایسه دو روش کاهش بعد تبدیلی و انتخابی

در جدول ۱۰ نتایج بدست آمده از روش کاهش بعد انتخابی که در مرجع ۲ پیاده‌سازی شده و روش کاهش بعد تبدیلی با LDA که در این مقاله پیاده‌سازی شده ارائه شده است. با مقایسه نتایج حاصل از دو روش تبدیلی و انتخابی کاهش بعد ملاحظه می‌کنیم که در روش تبدیلی میزان کاهش بعد و زمان آزمایش بهتر از روش انتخابی است ولی زمان آموزش بدتر است که

علت آن مدت زمانی است که سیستم عملیات محاسباتی برای تبدیل ۴۱ ویژگی به ۱۰ ویژگی کاهش بعد به روش LDA را انجام می‌دهد. ضمنا در روش تبدیلی، نرخ شناسائی و میزان کاهش بعد در مجموع بهتر است.

جدول۱۰: مقایسه کاهش بعد انتخابی در مرجع ۲ و LDA پیاده‌سازی شده در این تحقیق

روش کاهش بعد	تعداد ویژگیها بطور متوسط	دسته‌هائی که نسبت به ۴۱ ویژگی بهتر شناسائی می‌شوند	کاهش مدت زمان آموزش (٪)	کاهش مدت زمان آزمایش (٪)
LDA پیاده‌سازی شده در این تحقیق به عنوان روش تبدیلی	۹	Probe، U2R، R2L	۷/۴	۲۱/۷۳
روش انتخابی مرجع ۲	۱۳	Normal	۱۴/۱۶	۱۳/۴۶

۵- نتیجه‌گیری

از آنجائیکه بهترین درصدهای شناسائی نفوذ بوسیله سیستم تشخیص نفوذ با ویژگیهای LDA حاصل شده است، بنابراین می‌توان نتیجه‌گیری کرد که اطلاعات مفید بیشتری در ویژگیهای بدست آمده از LDA نسبت به سه روش دیگر وجود دارد. ضمنا با LDA زمان آموزش و آزمایش مناسب‌تری از سه روش تبدیلی دیگر حاصل می‌گردد. بنابراین می‌توان نتیجه گرفت که LDA مناسبتر از سه روش دیگر می‌باشد. ضمنا با مقایسه دو روش کاهش بعد انتخابی و تبدیلی ملاحظه کردیم که در روشهای کاهش بعد تبدیلی می‌توانیم به نرخ آشکارسازی و زمان آزمایش بهتری نسبت به حالتی که از روشهای انتخابی استفاده می‌کنیم دست یابیم، ولی زمان آموزش ما به دلیل محاسبات مربوط به تبدیل بردارهای ویژگی بیشتر می‌شود. کاهش مدت زمان این تبدیل می‌تواند باعث بهبود زمان آموزش گردد.

مراجع

[1] L. I Smith, "A Tutorial on Principal Component analysis", February 26, 2002. http://csnet.otago.ac.nz/cosc453

[2] S. Mukkamala, A. H. Sung, "Feature ranking and selection for Intrusion detection system using support vector machines", Presentations in Workshop on Statistical and Machine Learning Techniques in Computer Intrusion Detection, June 2002.

[3] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[4] Lee, W., Stolfo, S. J., Mok, K. W., "Mining in a Data-Flow Environment: Experience in Network Intrusion Detection", 5'th ACM SIGKDD, SanDiego, CA, 1999b, pp. 114-124.

[۵] علی عبداللهزاده میلانی، آنالیز مولفه‌های مستقل اصول و کاربردها، سمینار کارشناسی ارشد، دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر، ۱۳۸۱.

[1] Principle Component Analysis (PCA)
[2] Independent Component Analysis (ICA)
[3] Linear Discriminant Analysis (LDA)
[4] Multi Layer Perceptron (MLP)
[5] Trace