Archive of SID

تعمیم عصبی یادگیری چندگانه هسته

احمد نوید غنی‌زاده[۱]،دانشجوی کارشناسی ارشد، رضا منصفی،دانشیار، سید کمال الدین غیاثی شیرازی[۳]، استادیار

۳،۲،۱- دانشکده مهندسی- دانشگاه فردوسی مشهد-مشهد-ایران

ghanizadeh@stu-mail.um.ac.ir[۱] monsefi@um.ac.ir[۲] k.ghiasi@um.ac.ir[۳]

چکیده

یادگیری چندگانه هسته (MKL)، خانواده‌ای از الگوریتم‌های یادگیری ماشین می‌باشد که با استفاده از چندین تابع هسته متفاوت در انواع کاربردهای داده‌کاوی مورد استقبال واقع‌شده است. اما بیش‌تر روش‌های مطرح‌شده در این زمینه، سعی در بهبود کارایی محاسباتی دارند و مقالات کم‌تری به صحت طبقه‌بندی توجه داشته‌اند. در این مقاله سعی شده است با ارائه قالبی متفاوت از فرم متعارف یادگیری چندگانه هسته، خروجی ماشین‌های یادگیر (SVM) به‌صورت غیرخطی با استفاده از مدل شبکه عصبی ترکیب شوند. به عبارتی در این روش، مدل جدیدی برای تابع هسته ارائه نمی‌شود بلکه قالب جدیدی برای ترکیب طبقه‌بندها مطرح می‌شود. درنهایت، روش پیشنهادی با انواع روش‌های مطرح یادگیری چندگانه هسته بر روی تعدادی از مجموعه‌داده‌های پرکاربرد مقایسه می‌شود و مشاهده می‌شود صحت این روش نسبت به سایر روش‌ها بهبود چشم‌گیری داشته است.

واژه‌های کلیدی

یادگیری چندگانه هسته, شبکه‌های عصبی , طبقه‌بندی, روش‌های هسته

۱- مقدمه

در سال‌های اخیر، یادگیری هسته[۱] دست‌خوش تغییرات سریع و گسترده‌ای درزمینه‌ی یادگیری ماشین شده است. ایده اصلی این روش‌ها جایگزینی تابع هسته با تابع ضرب داخلی است که سبب اجرای ضمنی الگوریتم در فضای ویژگی متناظر با آن تابع هسته می‌شود. اما با وجود پیشرفت‌های قابل‌ملاحظه در این زمینه، هنوز هم مشکلات متعددی وجود دارد. ازجمله چالش‌های اصلی در توابع هسته می‌توان به چگونگی انتخاب تابع هسته مناسب برای مسئله خاص و تنظیم پارامترهای آن تابع هسته اشاره نمود. هم‌چنین نداشتن توانایی و انعطاف کافی برای کنار آمدن با داده‌های ناهمگون[۲] از محدودیت‌های تابع هسته منفرد است.

برای برطرف نمودن مشکلات فوق، یادگیری چندگانه هسته[۳] (MKL) مطرح‌شده است. مدل مطرح‌شده توسط توابع چندگانه هسته باعث افزایش توانایی و انعطاف‌پذیری در چالش‌های واقعی گردیده است [۱, ۲]. MKL اولین بار توسط Bach و همکاران [۲] به‌عنوان مسئله بهینه‌سازی توأم مطرح گردید. به دلیل عملکرد تجربی مناسب، روش یادگیری فوق به‌شدت مورد استقبال واقع‌شده است. در دهه اخیر، مطالعات گسترده‌ای نشان داده‌اند که روش‌های MKL قادر به حل بسیاری از چالش‌های موجود از قبیل شناسایی صدا [۳]، شناسایی اشیاء [۴]، تشخیص نفوذ شبکه [۵]، انتخاب ویژگی [۶] و غیره می‌باشند.

درحالی‌که استراتژی MKL سبب بهبود کارایی شده است، اما دارای محدودیت‌هایی نیز می‌باشد. در دهه اخیر، بیش‌تر مطالعات انجام شده با ارائه تکنیک‌های مختلف بهینه‌سازی سعی در بهبود کارایی و افزایش سرعت الگوریتم داشته‌اند، درحالی‌که مسئله اصلی افزایش صحت طبقه‌بندی است. در فرم متعارف یادگیری چندگانه هسته، هسته‌ی نهایی یاد گرفته‌شده، از ترکیب خطی وزنی چندین هسته شکل‌گرفته است. به عبارتی می‌توان معماری MKL را از دیدگاه شبکه عصبی به‌صورت شبکه‌ی عصبی شعاعی پایه[۴] با هسته‌های متفاوت دید. این طبقه‌بند منفرد مبتنی بر هسته توانایی لازم برای تشخیص مرزهای تصمیم پیچیده و قدرت کافی در برخورد با الگوهای ناهمگون را ندارد.

در میان روش‌های مختلف مطرح‌شده MKL، مقاله MKBoost [۷] دارای صحت مناسب‌تری نسبت به سایر روش‌های مطرح مرزهای دانش است. به همین دلیل به بررسی دقیق‌تر مدل ارائه‌شده توسط این روش پرداختیم. مشاهده می‌شود ریشه صحت طبقه‌بندی بالاتر MKBoost بهینه‌سازی نیست زیرا مسائل MKL محدب هستند و پاسخ بهینه یکتا است. باکمی دقت متوجه خواهیم شد که علت برتری این الگوریتم، ترکیب غیرخطی طبقه بندهای SVM می‌باشد. چالش‌های روش‌های موجود و موفقیت‌های مناسب MKBoost، این ایده را به ذهن می‌رساند که برای داشتن مدلی توانمند و کارآمد، ترکیب طبقه‌بندها در چارچوب[۵] معماری عمیق معرفی گردد. به عبارتی خروجی ماشین‌های یادگیر (SVM) به‌صورت غیرخطی در بستر شبکه عصبی با یکدیگر ترکیب شوند.

نگارش مقاله از قسمت‌های زیر تشکیل‌شده است: در قسمت ۲ به مرور ادبیات در این زمینه پرداخته می‌شود. جزئیات روش پیشنهادی در

[1] Kernel learning

[2] heterogeneous data

[3] Multiple kernel Learning, MKL

[4] Radial basis Function (RBF)

[5] framework

www.SID.ir

Archive of SID

قسمت ۳ داده‌شده است. آزمایش‌ها و نتیجه‌گیری در قسمت ۴ و ۵ بیان می‌شود.

۲- پیشینه

در حال حاضر شبکه‌های عصبی و یادگیری هسته دو موضوع موفق و کاربردی در انواع زمینه‌ها می‌باشند، درنتیجه به‌طور مختصر به بررسی این دو موضع مهم در این قسمت می‌پردازیم. در ادامه این بخش، تنظیمات یادگیری چندگانه هسته و تعدادی از کارهای گذشته بیان می‌شود.

۲-۱- شبکه‌های عصبی

شبکه‌های عصبی با الهام از ساختار عصبی مغز انسان برای تشخیص توابع غیرخطی و الگوهای پیچیده، در انواع زمینه‌های مختلف ازجمله استخراج ویژگی، تحلیل الگوها و طبقه‌بندی دارای صحت و عملکرد مناسبی می‌باشند.

۲-۲- یادگیری هسته

از دیرباز یادگیری هسته به‌طور گسترده موردمطالعه قرارگرفته است و در انواع روش‌های مبتنی بر هسته ازجمله ماشین بردار پشتیبان[۶]، به موفقیت دست‌یافته است. تمرکز مطالعات اخیر بر روی یادگیری خودکار تابع هسته بر روی مجموعه آموزشی است که به نحو مناسبی با داده‌های ناهمگون حاصل از منابع مختلف کنار بیاید. برای رسیدن به این هدف، استفاده از تابع هسته منفرد از پیش تعریف‌شده نامعقول می‌باشد.

یکی از روش‌های امیدوارکننده برای ارتقاء روش‌های مبتنی بر هسته، یادگیری چندگانه هسته می‌باشد. هدف اصلی این روش، یادگیری بهینه ترکیب هسته‌های مختلف است.

۲-۲-۱- تعریف مسئله

فرض کنید $D = \{x_1, x_2, .., x_N\}$ مجموعه داده‌های آموزشی به‌طوری‌که $X \subseteq \mathbb{R}^D$ باشد و $Y = \{y_1, y_2, \dots, y_N\}^T \in \{-1,1\}^n$ برچسب این داده‌های آموزشی باشد. در این تنظیمات، تعداد توابع هسته محدود می‌باشند ($K_\beta = \{k_1, k_2, \dots, k_s\}$) که توسط ترکیب محدب مطابق زیر با یکدیگر ترکیب می‌شوند:

$$k(x, y, \beta) = \sum_{j=1}^{s} \beta_j k_j(x, y) \quad with\ \beta_j \geq 0,\ \sum_{j=1}^{s} \beta_j = 1 \qquad (۱)$$

Where $x, y \in \mathbb{R}^n$.

به‌طوری‌که β_j وزن هریک از طبقه‌بندها در ترکیب نهایی می‌باشد. درنتیجه فرم اولیه یادگیری چندگانه هسته مطابق زیر است:

$$\max_{\beta} \min_{\alpha} \quad \sum_{k=1}^{s} \beta_k J_k(\alpha)$$
$$w.r.t. \quad \alpha \in \mathbb{R}^n,\ \beta \in \mathbb{R}^k \qquad (۲)$$
$$\text{s.t.} \quad 0 \leq \alpha \leq C,\ 0 \leq \beta,\ \sum_{i=1}^{N} \alpha_i y_i = 0\ and\ \sum_{k=1}^{S} \beta_k = 1.$$

که داریم:

$$j_k(\alpha) = \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j k_k(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \qquad (۳)$$

که در اینجا α ضرایب بردارهای ماشین پشتیبان است.

به‌طور خاص، روش‌های معرفی‌شده برای یادگیری چندگانه هسته از دو دیدگاه قابل‌بررسی است: ۱) روش‌هایی که با ترکیب خطی هسته‌ها، هسته جدیدی معرفی می‌نمایند، یا ۲) روش‌هایی که هسته پیچیده‌ای مطرح می‌نمایند. در فرم متعارف MKL، شاهد ترکیب خطی وزنی چندین هسته هستیم که این ترکیب خطی، پیچیدگی لازم را برای تقریب توابع پیچیده ندارد. البته در مطالعاتی سعی شده است با معرفی هسته پیچیده، صحت طبقه‌بندی افزایش یابد. به‌طور مثال در [۸] نویسنده سعی داشته است که دو مفهوم یادگیری عمیق[۷] و یادگیری چندگانه هسته را با یکدیگر تلفیق نماید. در این مقاله با معرفی تابع هسته arc-cosine نگاشت‌های متعددی بر روی داده اعمال می‌شود که ساختار آن به‌گونه‌ای است که مفهوم شبکه عصبی بی‌نهایت بعدی در آن گنجانیده شده است. در این حالت مفهوم یادگیری عمیق در هسته‌های تودرتو نشان داده‌شده است. به عبارتی:

$$k^{(l)}(x, y) = \phi^{(l)}\left(\dots \phi^{(1)}(x)\right).\phi^{(l)}\left(\dots \phi^{(1)}(y)\right) \qquad (۴)$$

در اینجا x و y داده‌های ورودی هستند و $\varphi^{(l)}$ تابع ویژگی در مرحله lام است. هم‌چنین $k^{(l)}$ هسته نهایی در لایه l می‌باشد. روش فوق دارای محدودیت‌هایی است، ازجمله ، استفاده از تنها نوع خاصی از تابع هسته (تابع *arc-cosine*) و معماری ثابت آن است.

روش دیگری که با استفاده از ایده یادگیری عمیق، هسته جدیدی معرفی می‌کند، در [۹] معرفی‌شده است. در این روش برای بهینه‌سازی مسئله از تابع هدف دوگان[۸] استفاده‌شده است. در این مقاله، نویسنده به علت پیچیدگی محاسباتی نتوانسته است بیش از دو لایه بسط دهد، همچنین در لایه دوم، فقط از یک تابع شعاعی پایه‌ای استفاده‌شده است که ماهیت چندگانه بودن هسته را ندارد.

درنهایت به بررسی روش MKBoost [۷] که حاصل از اعمال ایده boosting در یادگیری چندگانه هسته است، پرداخته می‌شود. روشی که برخلاف روش‌های اشاره‌شده در بالا، به‌جای معرفی هسته جدید از قالب متفاوتی برای ترکیب طبقه‌بندهای مبتنی بر هسته استفاده نموده است. در این الگوریتم، تابع تصمیم نهایی از ترکیب غیرخطی طبقه‌بندها مطابق زیر حاصل می‌شود:

[6] Support Vector Machine (SVM)

[7] Deep learning

[8] Dual

www.SID.ir

Archive of SID

***Algorithm 1* (*BNGMKL*)**

1: INPUT
- training data: $(x_1, y_1), \dots, (x_n, y_n)$
- kernel function: $k_j(.,.): X \times X \to \mathbb{R}, j = 1, \dots, M$
- initial distribution $D_1(i) = \frac{1}{N}, i = 1, .., N$

2: **for** $t = 1, \dots, T$ **do**
3: sample a set of n example based on weights
4: **for** $j = 1, \dots, S$ **do**
5: train weak classifier with kernel k_j:
6: compute the training error over example weights
7: **end for**
8: Select the best kernel function corresponding classifier with minimal error rate
9: Update the weights of examples and classifiers.
10: **end for**

Train a Neural Generalization of Multiple Kernel Learning

11: INPUT
- initialize network weights (small random values)
- Choose the best kernel function at each boosting trial:

$$k_j(.,.): X \times X \to \mathbb{R}, \quad j = 1, \dots, T$$

12: Training the neural network

$$f(x) = \text{sign}\left(\sum_{t=1}^{T} \beta_t \left(\text{sign}\left(\sum_{k=1}^{S} \alpha_t^k f_t^k(x)\right)\right)\right) \qquad (۵)$$

در اینجا T تعداد گام‌های روش boosting می‌باشد و α_t^k وزن طبقه‌بند kام در گام tام است. یادگیری طبقه‌بند f_t^k با هسته k_k از طریق اجرای الگوریتم SVM صورت می‌گیرد. در این روش با استفاده از عملگر sign در لایه میانی، طبقه‌بندها با یکدیگر ترکیب غیرخطی می‌شوند. از مقایسه آزمایش‌های الگوریتم فوق با سایر روش‌های موجود، مشاهده می‌شود که این روش دارای صحت طبقه‌بندی بالاتری است. علت این برتری، معرفی مدل پیچیده‌تر در ترکیب غیرخطی طبقه‌بندها در قالب معماری دو لایه است. درنتیجه ایده مقاله این است که به‌جای معرفی هسته جدید، قالب جدیدی برای ترکیب طبقه‌بندها ارائه شود. به عبارتی، خروجی طبقه‌بندهای هسته‌ای به‌صورت غیرخطی در بستر شبکه عصبی با یکدیگر ترکیب شوند. با استفاده از شبکه عصبی، پیچیدگی مدل افزایش پیدا می‌کند و نگاشت قوی‌تری در نمایش داده‌ها به دست می‌آید.

name	source	DN	DT	D
a1a	UCI	802	803	123
a3a	UCI	1592	1593	123
a4a	UCI	2391	2392	123
german.numer	Statlog	500	500	24
ionosphere	UCI	175	176	34
sonar	UCI	104	104	60
thyroid	UCI	140	75	5

جدول ۱- خلاصه‌ای از مجموعه‌داده‌های آموزش و آزمون

(DN تعداد مجموعه‌داده‌های آموزشی، DT تعداد داده‌های آزمون و D تعداد ویژگی‌های هر یک از مجموعه‌داده‌ها را نشان می‌دهد).

۳- روش پیشنهادی

در اینجا قالبی جدید برای ترکیب طبقه‌بندها مطرح می‌شود. در این مقاله، الگوریتم پیشنهادی برای طبقه‌بندی دو کلاسه توضیح داده‌شده است و به‌راحتی می‌توان آن را برای مسائل چند کلاسه نیز بسط داد.

درروش پیشنهادی، خروجی ماشین‌های یادگیر (S) به‌صورت غیرخطی با یک مدل شبکه عصبی ترکیب می‌شوند. توجه شود روش پیشنهادی ازاین‌جهت که تابع هسته را عمیق نمی‌کند و هم‌چنین مدل جدیدی برای تابع هسته ارائه نمی‌کند با روش‌های قبلی متفاوت است.

دو نوآوری مهم به کار گرفته‌شده درروش پیشنهادی عبارت‌اند از:

۱) ارائه قالبی متفاوت برای ترکیب هسته‌ها
۲) چگونگی انتخاب هسته‌ها در قالب پیشنهادی

۳-۱- تعمیم عصبی یادگیری چندگانه هسته

در این قسمت، تعمیم عصبی یادگیری چندگانه هسته مطرح می‌شود. هدف از پیشنهاد این روش، بالا بردن صحت طبقه‌بندی و افزایش مقیاس‌پذیری روش‌های موجود یادگیری چندگانه هسته می‌باشد. این روش با ارائه مدلی پیچیده‌تر، بازنمایی بهتری از داده‌ها خواهد داشت. دو نوآوری مهم به کار گرفته، به شرح زیر است:

۳-۱-۱- ارائه قالبی متفاوت برای ترکیب هسته‌ها

قالب روش پیشنهادی در شکل ۱ نمایش داده‌شده است. همان‌طور که در شکل پیدا است، داده‌ها از S هسته انتخابی عبور داده می‌شوند و به‌عنوان لایه ورودی شبکه عصبی قرار می‌گیرند. درواقع هر داده ورودی با تمام داده‌های آموزشی، در فضاهای ویژگی متناظر با S هسته انتخابی، ضرب داخلی می‌شود و حاصل آن به‌عنوان لایه ورودی شبکه عصبی در نظر گرفته می‌شود. تعداد لایه‌های این شبکه دو لایه می‌باشد و به‌راحتی می‌توان تعداد لایه‌ها را افزایش داد. درنتیجه، می‌توان مفهوم یادگیری عمیق را در این ترکیب طبقه‌بندها جستجو نمود. در این قالب، وزن‌های هر

www.SID.ir

Archive of SID

یک از لایه‌ها طی عمل باز انتشار خطا از طریق جستجوی گرادیانی حاصل می‌شوند.

۱-۱-۳- چگونگی انتخاب هسته‌ها در قالب پیشنهادی

گام دوم روش پیشنهادی، نحوه انتخاب هسته‌های اولیه را بیان می‌کند مشاهده می‌شود نوع هسته‌های اولیه تأثیر بسزایی در صحت و سرعت نهایی دارد. برای انتخاب هسته‌های ورودی شبکه عصبی، از هسته‌های برگزیده روش MKBoost استفاده می‌شود. علت انتخاب هسته‌ها از این روش، کارایی مناسب MKBoost نسبت به سایر روش‌های موجود MKL می‌باشد و تأثیر مدل‌سازی به‌صورت شبکه عصبی بهتر دیده شود.

روش MKBoost برای ترکیب طبقه‌بندها از ایده boosting استفاده کرده است. در هر گامِ boosting، بر اساس میزان سختی داده‌ها به آن‌ها وزنی داده می‌شود. درعین‌حال وزن هریک از طبقه‌بندها، بر اساس میزان صحت طبقه‌بندی الگوهای آموزشی مقداردهی می‌گردد. در میان انواع مختلف ارائه‌شده، الگوریتم MKBoost-D1 انتخاب‌شده است. علت انتخاب این الگوریتم، صحت بالای آن نسبت به ۴ روش ارائه‌شده در این مقاله است. در این روش، در هر گام از میان تمامی S طبقه‌بند، بهترین طبقه‌بند با کم‌ترین خطا انتخاب می‌شود. در انتهای روش پیشنهادی، تمامی هسته‌های پایه انتخاب می‌شوند. علت آن‌هم، عمومیت بخشیدن به الگوریتم پیشنهادی است. هم‌چنین تأثیر شبکه عصبی به‌تنهایی مشخص شود. الگوریتم ۱ این قالب را نمایش می‌دهد.

۴- نتایج آزمایشات

در قسمت آزمایشات، عملکرد روش‌های پیشنهادی بر روی انواع مجموعه داده‌ها با سایر روش‌های جدید MKL مقایسه می‌شود. مشابه مقاله MKBoost ما دقیقاً همان تنظیمات استفاده می‌نماییم. در تمامی آزمایشات، ما از ۱۷ هسته پایه‌ی مرسوم که در زیر تعریف‌شده است. استفاده می‌نماییم:

۱- سه تابع هسته چندجمله‌ای با درجه‌های $p = \{1,2,3\}$

$(xy)^p$

۲- ۱۴ تابع هسته نرمال با وزن‌های $\sigma = \{2^{-6}, 2^{-5}, \dots, 2^{7}\}$

$\exp\left(\frac{x - y}{\sigma}\right)$

مجموعه داده‌های مورد آزمایش در جدول ۱ نمایش داده‌شده است. به‌طورمعمول روش‌های جدید MKL از این مجموعه داده‌ها استفاده کرده‌اند. اندازه دادگان فوق خیلی بزرگ نیست، زیرا تعدادی از روش‌های مورد مقایسه مقیاس‌پذیر نمی‌باشند و تنها برای مجموعه داده‌های کوچک مناسب هستند.

۱-۴- روش‌های مورد مقایسه

برای نشان دادن برتری روش‌های پیشنهادی، آزمایشات با سایر روش‌های جدید و برتر MKL مقایسه شده است. جزئیات روش‌های مورد مقایسه در زیر آمده است:

Lp-MKL: یکی از الگوریتم MKL که از نرم P استفاده کرده است [۱].

MKBoost: یکی از الگوریتم MKL که از ایده boosting استفاده کرده است [۷].

MKL-SD: یکی از الگوریتم MKL که از بهینه‌سازی نیمه‌معین[9] بهره برده است [۱۱].

BNGMKL: الگوریتم پیشنهادی وقتی هسته‌های ورودی شبکه عصبی از الگوریتم MKBoost انتخاب شوند.

NGMKL: الگوریتم پیشنهادی وقتی تمامی هسته‌های پایه انتخاب شوند.

برای پیاده‌سازی شبکه عصبی پیشنهادشده از ابزار Caffe [۱۰] استفاده‌شده است. برای پیاده‌سازی سایر روش‌های مورد مقایسه مطابق کارهای گذشته عمل می‌گردد. برای پیاده‌سازی MKL-SD از ابزار SimpleMKL [۱۱] استفاده‌شده است. در این روند حداکثر تعداد تکرارهای اجرای بهینه‌سازی ۵۰۰ و آستانه فاصله دوگان 0.01 می‌باشد. همچنین برای پیاده‌سازی MKBoost از ابزار موجود در مقاله [۷] استفاده‌شده است. در این روند تعداد گام‌های boosting ۱۰۰ و نرخ نمونه‌برداری 0.2 می‌باشد. برای پیاده‌سازی LP-MKL از ابزار موجود در مقاله [۱] استفاده‌شده است، و برای داشتن تعادل بین دقت و کارایی $p = 2$ در نظر گرفته‌شده است.

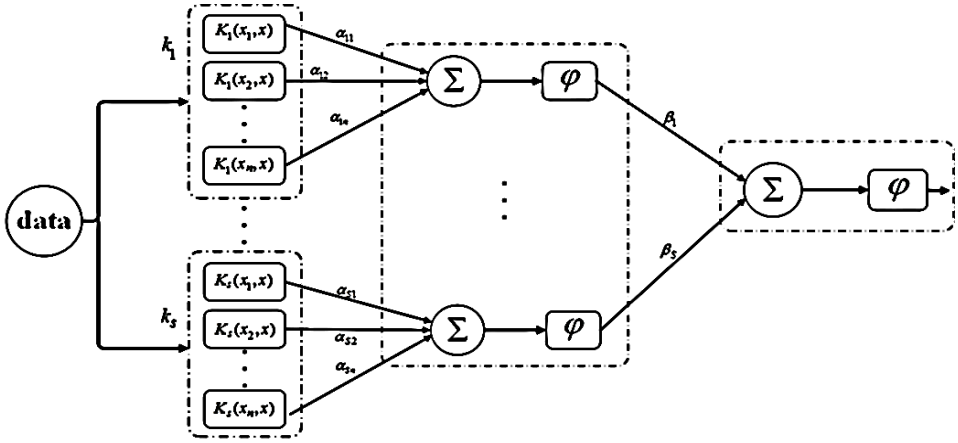
	Metric	MKL-SD	Mkboost	Lp-MKL	BNGMKL	NGMKL
D1	*error*	0.2453 ± 0.0009	0.1878 ±0.0077	0.2393 ± 0.006	**0.1739 ± 1.0e − 04**	0.1804 ± 4.4e − 05
	std					
D2	*error*	0.2424 ± 0.0003	0.1863 ±0.0051	0.2403 ± 0.004	**0.1708 ± 6.3e − 05**	0.1793 ± 9.0e − 06
	std					
D3	*error*	0.2485 ± 0.0001	0.189 ±.0052	0.2439 ± 0.003	**0.1732 ± 6.9e − 05**	0.1822 ± 1.9e − 05
	std					
D4	*error*	0.2999 ± 0.0004	0.2639 ±.0117	0.2935 ± 0.005	0.2579 ± 2.4e − 04	**0.2457 ± 1.0e − 04**
	std					
D5	*error*	0.0694 ± 0.0155	0.0574 ±0.0139	0.0663 ± 0.020	**0.0509 ± 3.5e − 04**	0.0699 ± 8.8e − 05
	std					
D6	*error*	0.2248 ± 0.0423	0.1817 ±0.0396	0.1976 ± 0.040	**0.1793 ±0.002**	0.2141 ± 8.5e − 04
	std					
D7	*error*	0.0421 ± 4.7298e − 04	0.0573 ± 0.0011	0.0471 ± 5.4e − 04	**0.0417 ± 6.4e − 04**	0.0553 ± 9.9e − 04
	std					

جدول ۲: نتایج آزمایش‌های مختلف بر روی مجموعه داده‌های معرفی‌شده در جدول ۱، همان‌طور که مشخص است، کارایی روش‌های پیشنهادی به‌طور قابل‌توجهی از سایر روش‌های موجود بهتر است

[9] Semidefinite optimization

www.SID.ir

Archive of SID



شکل ۱: قالب روش پیشنهادی برای ترکیب طبقه‌بندهای هسته در قالب شبکه عصبی دو لایه

در همه آزمایشات، برای داشتن نتایج عادلانه‌تر هر الگوریتم ۲۰ بار بر روی مجموعه داده‌ها تکرار می‌شود، و صحت نهایی از میانگین صحت تمام تکرارها حاصل می‌شود. همچنین داده‌های آموزشی با میانگین صفر و واریانس یک نرمال شده‌اند و مطابق این میانگین و واریانس داده‌های آموزش، داده‌های آزمون نیز نرمال می‌شوند. همان‌طور که در جدول ۲ پیداست، روش‌های پیشنهادی از نظر صحت طبقه‌بندی به‌طور قابل‌توجهی از سایر روش‌های موجود بهتر هستند.

۵- نتیجه‌گیری

در این مقاله سعی شده است با ارائه روش جدید برای ترکیب طبقه‌بندها، صحت طبقه‌بندی روش‌های یادگیری چندگانه هسته بهبود یابد. در این روش خروجی ماشین‌های یادگیر (SVM) به‌صورت غیرخطی با مدل شبکه عصبی ترکیب می‌شوند. توجه شود روش پیشنهادی ازاین‌جهت که تابع هسته را عمیق نمی‌کند و هم‌چنین مدل جدیدی برای تابع هسته ارائه نمی‌کند با روش‌های قبلی متفاوت است. در آزمایشات مشخص شد که این روش ازلحاظ صحت به‌صورت قابل‌توجهی از روش‌های جدید و موفق یادگیری چندگانه هسته مناسب‌تر است. در این روش ضرایب هر یک از هسته‌های مورداستفاده در مسئله یادگیری چندگانه هسته، طی عمل باز انتشار خطا یاد گرفته می‌شوند. درنتیجه این قالب، درگیر حل مسائل بهینه‌سازی پیچیده نخواهد بود. هم‌چنین می‌توان اجازه داد، خطا به توابع هسته نیز پس انتشار پیدا کند. به‌عبارت‌دیگر توابع هسته هم یاد گرفته شوند. یکی از مزیت‌های روش پیشنهادی این است که از دو بخش هسته و شبکه عصبی تشکیل‌شده است و بنابراین به‌سادگی می‌توان از بخش شبکه عصبی برای بازشناسی چندکلاسی بهره برد.

مراجع

[1] S. Sonnenburg, G. Rätsch, and C. Schäfer, "A general and efficient multiple kernel learning algorithm," in *NIPS*, 2005, pp. 1275-1282.

[2] F. R. Bach, G. R. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in Proceedings of the twenty-first international conference on Machine learning, 2004, p. 6.

[3] C. Longworth and M. J. Gales, "Combining derivative and parametric kernels for speaker verification," Audio, Speech, and Language Processing, IEEE Transactions on, vol. 17, pp. 748-757, 2009.

[4] A. Shrivastava, V. M. Patel, and R. Chellappa, "Multiple kernel learning for sparse representation-based classification," Image Processing, IEEE Transactions on, vol. 23, pp. 3013-3024, 2014.

[5] J. M. Fossaceca, T. A. Mazzuchi, and S. Sarkani, "MARK-ELM: Application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection," Expert Systems with Applications, vol. 42, pp. 4062-4080, 2015.

[6] M. Szafranski, Y. Grandvalet, and A. Rakotomamonjy, "Composite kernel learning," Machine learning, vol. 79, pp. 73-103, 2010.

[7] H. Xia and S. C. Hoi, "Mkboost: A framework of multiple kernel boosting," Knowledge and Data Engineering, IEEE Transactions on, vol. 25, pp. 1574-1586, 2013.

[8] Y. Cho and L. K. Saul, "Kernel methods for deep learning," in Advances in neural information processing systems, 2009, pp. 342-350.

[9] J. Zhuang, I. W. Tsang, and S. C. Hoi, "Two-Layer Multiple Kernel Learning," in AISTATS, 2011, pp. 909-917.

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, et al., "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the ACM International Conference on Multimedia, 2014, pp. 675-678.

[11] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," Journal of Machine Learning Research, vol. 9, pp. 2491-2521, 2008.

www.SID.ir