

یادگیری به موقع معیار فاصله در محیطهای غیر ایستا

امید سجودی شیجانی

استادیار دانشگاه آزاد اسلامی، واحد قزوین، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، o_sojoodi@qiau.ac.ir

تاریخ دریافت: ۹۲/۱۲/۱۵ تاریخ پذیرش: ۹۴/۱/۲۴

چکیده

در این مقاله یک روش غیر پارامتری تطبیقی برای یادگیری به موقع معیار فاصله در محیطهای پویا ارائه می‌گردد. استفاده از این روش در محیطهای غیر ایستا، طبقه‌بندی بر اساس شباهت الگوهای ورودی که توسط یک فرآیند تصادفی غیر پارامتری تولید شده‌اند را با کارایی بالا امکان‌پذیر می‌سازد. منظور از کارایی در این تحقیق دقت طبقه‌بندی کننده، ابعاد فضای ویژگی و تعداد مراجعات برای بروزرسانی و تطبیق پارامترها در طول عمر سیستم است. برای این منظور از داده‌ها با برجسب زمانی برای یادگیری معیار فاصله و استخراج ویژگیهای جدید استفاده می‌شود. این یادگیری غیر پیوسته بوده و دقیقاً در زمان مورد نیاز انجام می‌گیرد. در حالیکه مدل‌های تطبیقی استخراج ویژگی برای کاهش فضای ابعاد و/یا افزایش دقت طبقه‌بندی کننده‌ها پیشنهاد شده‌اند، اما این روشها نسبت به تغییرات غیر پارامتریک داده‌های ورودی حساس بوده و نیاز به بروزرسانی پیوسته پارامترهای خود دارند. در روش ارائه شده یک ماتریس انتقال بهینه داده‌های دارای برجسب زمانی را از فضای اولیه به یک فضای ویژگی جدید به گونه‌ای انتقال می‌دهد که احتمال طبقه‌بندی صحیح داده‌های جدید با استفاده از طبقه‌بندی کننده مبتنی بر شباهت پیشینه گردد. همچنین با استفاده از داده‌های زمانی، توزیع داخل کلاسی و توزیع خارج کلاسی غیر پارامتری ارائه می‌گردد. نتایج آزمایشات روی مجموعه داده‌های واقعی و مصنوعی که شامل تغییرات طبیعی و مصنوعی می‌باشند عملکرد روش پیشنهادی را از نظر دقت، کاهش ابعاد و تعداد مراجعات برای بروزرسانی پارامترها در محیطهای پویا نشان می‌دهد.

کلید واژه‌ها

یادگیری تطبیقی، یادگیری به موقع معیار فاصله، طبقه‌بندی کننده تطبیقی

مقدمه

تاکنون سه راهکار برای مواجهه با این مساله پیشنهاد شده است. طبقه بندی کننده‌های تطبیقی با هدف بروز رسانی پارامترهای دسته بند و/یا انتخاب یک پنجره بهینه از داده‌های گذشته برای افزایش عملکرد الگوریتم یکی از این رویکردها است [۱-۳]. رویکرد دیگر برای مقابله با رویه‌ی تغییرات در طول زمان، اختصاص وزن به داده‌های آموزشی متناسب با زمان ورود آنها میباشد. در اینصورت داده‌های جدید حاوی اطلاعات بروزتری از مفهوم کلاسها بوده و وزن بیشتری به آنها تخصیص داده می‌شود [۴، ۵]. آخرین رویکرد مطرح شده برای بررسی تغییر مفهوم، استفاده از طبقه بندی کننده‌های تجمیعی است که در آن ترکیبی از خروجی طبقه‌بندی کننده ها نتیجه‌ی نهائی طبقه‌بندی را مشخص می‌کند [۶، ۷]. از طرف دیگر، راهکار پیشنهاد شده در این مقاله جهت مواجهه با تغییر مفهوم، تطبیق معیار فاصله با شرایط محیطی توسط استخراج ویژگیهای جدید در طول عمر سیستم می‌باشد.

در کاربردهای دنیای طبیعی گذشت زمان ممکن است باعث تغییر در طبیعت و رفتار سیستم‌ها گردد. این تغییرات می‌توانند به دلایل مختلفی از جمله تکامل یک فرآیند، تاثیر سالخوردگی^۱، رانش فرآیند^۲، از تنظیم خارج شدن سنسورها و یا پدیده‌های محیطی دیگر بوجد آیند [۱]. محیطهای پویا، تغییر مفهوم^۳ و غیرایستایی معمولاً بعنوان جایگزینی برای عرضه‌ی مفهوم تغییرات در حوزه‌های مختلف تحقیقاتی استفاده می‌شوند. یک مسئله‌ی اصلی در کاربردهای محیط‌های پویا، استفاده از طبقه‌بندی کننده‌هایی است که فرضیه‌های غیر ایستا را برای فرآیند تولید داده‌ها در نظر می‌گیرند.

1 Aging effect
2 Process drift
3 Concept drift

به گونه‌ای انتقال می‌دهد که خطای طبقه‌بندی کننده در فضای ویژگی با ابعاد پائین کاهش یابد. از جمله مزایای دیگر این روش عدم نیاز به بروز رسانی مداوم ماتریس انتقال می‌باشد. در ادامه مقاله بصورت بخش‌های زیر ارائه می‌شود: بخش ۲ روش JITDML و پارامترهای تخمینی را ارائه کرده است. در بخش ۳ تکنیک کاهش ابعاد در JITDML بوسیله‌ی تعریف یک ماتریس غیرمربعی توضیح داده شده است. در بخش نهایی عملکرد روش پیشنهادی نشان داده شده و با روش های دیگر مقایسه گردیده است.

یادگیری به موقع معیار فاصله

از دیدگاه استخراج ویژگی، در JITDML یک معیار فاصله بوسیله‌ی آموزش یک انتقال خطی از فضای داده‌ای ورودی به یک فضای جدید تخمین زده می‌شود. در این فضای جدید likelihood انتخاب کلاس صحیح در فرضیه یادگیری ماشین توسط طبقه‌بندی کننده KNN پیشینه می‌گردد. برای این منظور یک ماتریس انتقال بهینه A^{DT} - که در آن DT زمان تصمیم‌گیری یا زمان جاری می‌باشد- تعریف می‌گردد. این ماتریس بوسیله‌ی پیشینه کردن یک تابع هدف مشتق‌پذیر بر پایه اطلاعات متقابل^۵ در فضای جدید بدست می‌آید. تابع هدف تعریف شده در این تحقیق نسبت فاصله‌ی نقاط با کلاس یکسان به فاصله‌ی نقاط با کلاس متفاوت در مکان‌های با احتمال بالا برای هر کلاس است. برای این کار ابتدا نقاط ورودی جدید برای هر کلاس بر حسب مقدار احتمال بصورت آماری تخمین زده می‌شود. در شکل ۱ فضای ویژگی دو بعدی قبل و بعد از انتقال بوسیله‌ی ماتریس انتقال A در زمان DT نشان داده شده است.

برای این منظور از قانون بکار گرفته شده در طبقه‌بندی کننده KNN بعنوان معیار اصلی جهت استخراج ویژگیها استفاده شده است. دلیل استفاده از این قانون به عنوان یک معیار، قالب غیر پارامتری تغییرات در محیط‌های پویا می‌باشد که باعث استفاده هر چه بیشتر از طبقه‌بندی کننده‌های مبتنی بر شباهت می‌گردد. در طبقه‌بندی کننده‌های بر اساس شباهت، یک معیار شباهت مانند ضرب داخلی در فضای هیلبرت یا یک معیار مغایرت همانند فاصله در فضای اقلیدسی تعریف شده است [۸]. از سوی دیگر مشکل بالا بودن فضای ابعاد یکی دیگر از مشکلات اساسی در اکثر کاربردهای برخط شناسائی الگو و داده‌کاوی می‌باشد. این کاربردها شامل تشخیص حروف دستنویس [۹]، دنبال کردن هدف و کشف تغییر در دنباله‌های ویدئویی [۱۰-۱۲]، پردازش سیگنال بیولوژیکی متغیر در زمان [۱۳، ۱۴]، مکان‌یابی خودکار ربات متحرک [۱۵] و فیلترکردن سند یا نامه بصورت برخط می‌باشد [۱۶، ۱۷].

تاکنون تحقیقات اندکی بر روی تکنیک‌های کاهش فضای ویژگی در محیط‌های پویا انجام شده است. در ALDA [۱۸] یک ماتریس انتقال بهینه $(\Sigma^{-1/2})$ ، در حالتی که توزیع تمامی کلاس‌ها بصورت نرمال و کوواریانس‌ها مساوی هستند، بصورت تکراری تولید می‌شود. جوینگ و همکاران [۱۹] روش CCIPCA را ارائه کردند که بصورت افزایش پله‌ای مولفه‌های اصلی یک جریان از داده‌ها را محاسبه می‌کند. از دیگر روشهای موثر که بصورت وسیع در زمینه‌ی پردازش سیگنال و شناسائی گفتار از آن استفاده شده است می‌توان به ICA اشاره کرد. نسخه‌ی اصلاح شده‌ی ICA که با عنوان FastICA شناخته شده است [۲۰]، از یک طرح تکرار نقطه ثابت برای محاسبه‌ی ماتریس تصویر غیر پارامتری استفاده می‌کند. مشکل تمام این تکنیک‌ها، نیاز به محاسبه و تخمین ماتریس انتقال در هر گام از طول عمر سیستم می‌باشد.

با توجه به مطالب ارائه شده در بالا، این مقاله در جستجو برای پاسخ به این سوال است که چگونه می‌توان یک معیار فاصله تطبیقی و به موقع را برای بهبود دقت الگوریتم طبقه‌بندی کننده KNN همزمان با کاهش ابعاد فضای ویژگی‌ها در محیط‌های پویا غیر پارامتری توسعه داد؟ برای این هدف یک ماتریس انتقال بهینه A بوسیله‌ی پیشینه سازی ضریب واگرایی کولبک-لیبر^۴ بین توزیع‌های داخل کلاسی و خارج کلاسی در محتملترین مکان نمونه ورودی آینده محاسبه می‌گردد. بعلاوه برای حفظ یکپارچگی و مقاوم نگه‌داشتن طبقه‌بندی کننده تطبیقی KNN در برابر تغییرات، ماتریس A در مناسبترین زمان، مجدداً محاسبه شده و بروز می‌گردد.

این روش که اختصاراً JITDML نامیده می‌شود، پایگاه دانش سیستم های یادگیر غیر پارامتری و تطبیقی را به یک فضای بهینه

5 Mutual information

4 Kullback-Leibler divergence

$$d(X_i, X_j) = \|A^{DT} x_i, A^{DT} x_j\|_F$$

$$= tr((x_i - x_j)A^{DT} A^{DTT} (x_i - x_j)^T) \quad (3)$$

بطوریکه $X_i = A^{DT} x_i$ و $X_j = A^{DT} x_j$ نقاط متناظر با x_i و x_j در فضای انتقال داده شده در زمان DT می باشند. از دیدگاه همسایگی، احتمال انتخاب X_i بعنوان همسایه X_j در فضای انتقال داده شده می تواند همانند زیر تعریف شود:

$$P(X_j | X_i) = \frac{d^{-1}(X_i, X_j)}{\sum_{c'=1..C} \sum_{x_m \in c'} d^{-1}(X_i, X_m)} \quad (4)$$

بطوریکه $d^{-1}(X_i, X_j)$ نشان دهنده $1/d(X_i, X_j)$ می باشد. بر اساس احتمال همسایگی ذکر شده، توزیع داخل کلاسی⁷ همسایگی کلاس c برای داده ی X بوسیله $\zeta_c^I(X)$ نشان داده شده و بصورت مجموع احتمال انتخاب داده های با کلاس یکسان در اطراف X تعریف می شود. در حقیقت توزیع داخل کلاسی⁸ احتمال کلاس بندی صحیح داده ی X از لحاظ مفهوم همسایگی را در بر دارد. در همین راستا، توزیع همسایگی خارج کلاسی⁹ برای کلاس c داده ی X بوسیله $\zeta_c^E(X)$ نشان داده شده که بعنوان توزیع داده های با کلاس متفاوت در اطراف X تعریف می شود. توزیع بر مبنای همسایگی داخل کلاسی و خارج کلاسی داده ی X از کلاس c بصورت زیر نشان داده می شود:

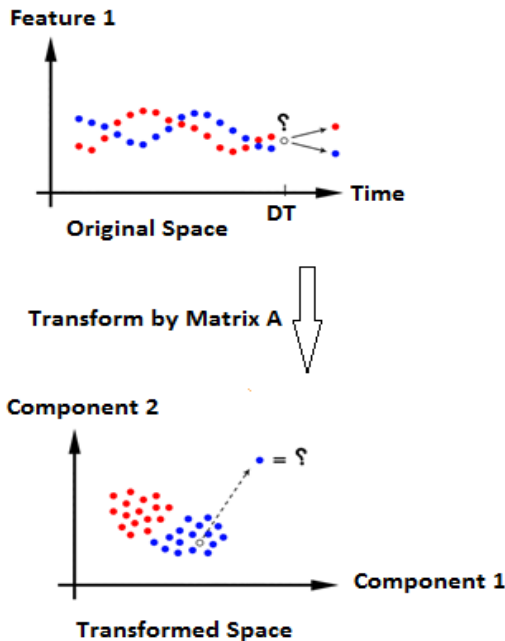
$$\zeta_c^I(X) = \sum_{X_j \in c} P(X_j | X) \quad (5)$$

$$\zeta_c^E(X) = \sum_{c'=1..C, c' \neq c} (\sum_{X_j \in c'} P(X_j | X)) \quad (6)$$

تابع هدفی که برای بیشینه کردن احتمال کلاس بندی صحیح در زمان تصمیم (DT) استفاده شده، بصورت جمع واگرایی کولبک-لیبر مابین توزیع های داخل کلاسی و خارج کلاسی \hat{x}^k برای هر کلاس پیشنهاد شده است:

$$f(A^{DT}) = \sum_{c=1..C} k l(\zeta_c^I(\hat{X}^c) \| \zeta_c^E(\hat{X}^c))$$

$$= \sum_{c=1..C} (\zeta_c^I(\hat{X}^c) \cdot \log \frac{\zeta_c^I(\hat{X}^c)}{\zeta_c^E(\hat{X}^c)}) \quad (7)$$



شکل ۱. فضای ویژگی قبل و بعد از انتقال توسط ماتریس A در زمان DT

همانطور که در شکل ۱ دیده می شود، بعد از انتقال احتمال تعیین بر چسب کلاس صحیح برای داده ی جدید در زمان تصمیم گیری بصورت چشمگیری افزایش می یابد. به دلیل کمبود اطلاعات پیشین توزیع کلاسها در محیطهای پویا، بهترین تخمین از داده های ورودی استفاده از پیش بینی یک گام به جلو غیر پارامتری با اندازه گام غیر یکنواخت است که بطور مستقیم از امید ریاضی شرطی $E(x|DT = t)$ بدست می آید. بنابراین بهترین تخمین داده ورودی \hat{x}^k برای کلاس k در زمان t با در نظر گرفتن داده های مشاهده شده در جریان داده بصورت زیر بدست می آید:

$$\hat{x}^k = E(x | DT = t) = \sum_{x_i \in k} x_i \cdot p(x_i | DT = t) \quad (1)$$

بطوریکه x_i ($i=1 \dots N$) مجموعه داده در فضای R^D و $c_i \in \{1, \dots, C\}$ برچسب کلاس داده ی x_i است. احتمال مشاهده x_i در زمان t، $p(x_i | DT = t)$ ، می تواند بوسیله ی تشابه نرمال بین x_i و داده های در نظر گرفته شده در زمان تصمیم گیری t تعریف شود بنابراین:

$$\hat{x}^k = \frac{1}{\sum_{x_m \in k} 1/(t - t_m)} \sum_{x_i \in k} \frac{x_i}{(t - t_i)} \quad (2)$$

از سوی دیگر فاصله x_i و x_j در فضای انتقال داده شده بوسیله ی نرم فروبنیوس $A^{DT} x_i$ و $A^{DT} x_j$ تعریف می شود:

7 Intra-class distribution
8 Extra-class distribution

6 One-step-ahead prediction

بیشتر باشد، تعداد روزرسانی ها افزایش می یابد. در نتیجه حجم محاسبات افزایش یافته (کارایی منابع محاسباتی کاهش می یابد) و دقت طبقه بندی کننده افزایش پیدا می کند و بالعکس.

$\lambda \uparrow \rightarrow \# \text{ of adaptation} \uparrow \rightarrow \text{resource performance} \downarrow$
 $\rightarrow \text{accuracy} \uparrow$
 $\lambda \downarrow \rightarrow \# \text{ of adaptation} \downarrow \rightarrow \text{resource performance} \uparrow$
 $\rightarrow \text{accuracy} \downarrow$

کاهش ابعاد با استفاده از ماتریس انتقال غیر مربعی

تصویر انتقال خطی داده ها از فضای ویژگی با ابعاد بالا به فضایی با ابعاد پائین تر معمولاً بوسیله ی یک ماتریس غیر مربعی که یک ترکیب خطی از ویژگی ها در فضای اصلی است، صورت می گیرد. در NCA [۲۱] هنگام تولید ماتریس انتقال خصوصیات موقعیتی (مکانی) داده ها مورد توجه قرار می گیرد. در RCA [۲۲] از ارتباط بین اجزاء برای تشکیل ماتریس انتقال استفاده می شود و CCIPCA از یک افزایش پله ای مولفه های اصلی یک جریان از داده ها استفاده کرده و ماتریس انتقال را محاسبه می نماید. اما هیچکدام از این روشها فرضیه ی تغییرات در بستر زمان را در نظر نمی گیرند. در JITDML ابعاد فضای ویژگی بوسیله ی محدود کردن ماتریس انتقال A به یک ماتریس غیر مربعی $D \times d$ ، بصورت خطی از D به d ($d < D$) کاهش داده می شود. ماتریس انتقال A جهت کاهش پیچیدگی محاسباتی و حافظه ی مورد نیاز، داده ها را از فضای R^D به فضای R^d نگاشت می کند. شکل ۲ مجموعه داده های Elec2 [۲۷] و Luxembourg [۲۳، ۲۴] را که با استفاده از روشهای CCIPCA، RCA و JITDML در زمان $t=200$ به دو بعد نگاشت شده اند را نشان می دهد. در این شکل، نقاط آینده توسط ستاره مشخص شده اند و γ همسایگی آنها بوسیله دایره نشان داده شده است. همانطور که در شکل ۲- (e) و ۲- (f) دیده می شود، تعداد نقاط هم کلاس در همسایگی نقاط آینده بیشتر از موارد مشابه در شکل‌های ۲- (a)، ۲- (b)، ۲- (c) و ۲- (d) می باشد. این نشاندهنده احتمال بالاتر در طبقه بندی صحیح داده ی بدون برچسب بعدی (نقاط نمایش داده شده توسط ستاره) بوسیله ی KNN در JITDML است.

نتایج آزمایشات

برای اعتبارسنجی روش پیشنهادی از یک قاب کاری آزمایش که شامل مجموعه داده های مصنوعی و واقعی، پنج الگوریتم مقایسه ای و دو معیار مقایسه است، استفاده شده است.

مجموعه داده های مصنوعی

مجموعه داده های ساختگی با توزیع نرمال غالباً بوسیله ی محققان برای ارزیابی روش ها در محیط های پویا مورد استفاده قرار می گیرد [۱، ۲۵، ۲۶]. در اینجا دو نوع تغییرات در زمان را

بدیهی است که برای هر کلاس داریم:

$$\zeta_c^E(\hat{X}^c) = 1 - \zeta_c^I(\hat{X}^c) \quad (8)$$

بنابراین

$$\begin{aligned} A^{DT} &= \arg \max_{A^{DT} \in R^D \times R^D} (f(A^{DT})) \\ &= \arg \max_{A^{DT} \in R^D \times R^D} \left(\sum_{c=1..C} \zeta_c^I(\hat{X}^c) \right) \\ &= \arg \max_{A^{DT} \in R^D \times R^D} \left(\sum_{c=1..C} \frac{\sum_{x_j \in c_c} d^{-1}(X_j, \hat{X}^c)}{\sum_{c'=1..C} \sum_{x_j \in c'} d^{-1}(X_j, \hat{X}^c)} \right) \end{aligned} \quad (9)$$

که در آن \hat{x}^k تخمین بهینه برای داده‌های ورودی از کلاس c در فضای انتقال داده شده است. در مورد مشکل محاسبه‌ی گرادیان تابع $f(A^{DT})$ نسبت به A^{DT} یک بهینه‌ساز بر اساس گرادیان مانند بهینه‌ساز گرادیان توأم^۹ استفاده شده است. بطور خلاصه تکنیک JITDML را می‌توان بصورت سه گام زیر بیان کرد:

۱. مکان داده‌ی ورودی بعدی \hat{x}^k برای هر کلاس c بوسیله‌ی رابطه‌ی ۲ پیش بینی می‌شود.
۲. توزیع‌های داخل کلاسی و خارج کلاسی برای داده‌های تخمین زده شده در فضای انتقال داده شده با استفاده از روابط ۵ و ۶ تخمین زده می‌شوند.
۳. ماتریس انتقال A^{DT} بوسیله‌ی بهینه کردن تابع هدف $f(A^{DT})$ رابطه‌ی ۷ بدست می‌آید.

بروز رسانی به موقع

بر خلاف روشهای تطبیقی معمول که بطور مداوم در هر گام ماتریس انتقال را محاسبه می‌نمایند فقط در زمان مورد نیاز ماتریس انتقال را محاسبه و بروز می‌رساند. بطور مشخص، ماتریس A^{DT} توسط حل معادله ۷ هنگامی محاسبه می‌گردد که ماتریس انتقال قبلی A^{DT-1} توانایی انتقال داده‌ها به یک فضای ویژگی مطمئن را نداشته باشد. فضای ویژگی را مطمئن مینامیم هرگاه احتمال طبقه بندی درست داده ورودی از یک حد آستانه λ ($0 \leq \lambda \leq 1$) بیشتر باشد. به عبارت دیگر، برای تعیین ضرورت روزرسانی مجدد ماتریس انتقال، میانگین توزیع داخل کلاسی برای \hat{X}^c (با توجه به ماتریس انتقال A^{DT-1}) محاسبه شده و با λ مقایسه میگردد. اگرچه λ پارامتری است که باید توسط کاربر مقداردهی اولیه شود اما بطور معنی داری تعیین کننده یک مصالحه بین میزان دقت و حجم محاسبات می باشد. هر چه λ

⁹ Conjugate gradient optimizer

در فضای R^{34} می باشد. در حالی که مجموعه های Vote و Ionosphere بطور معمول در محیط های ایستا مورد استفاده قرار می گیرند، Zliobaite در [۳] نشان داد که می توان از این مجموعه ها به عنوان محک هایی در محیط های پویا استفاده کرد.

برای تولید مجموعه داده های مصنوعی بررسی می نمائیم: تغییرات ناگهانی و تغییرات تدریجی. در مورد تغییرات ناگهانی ۴۰۰ نمونه در دو کلاس گوسی با احتمال برابر و میانگین و کوواریانس ۰ و ۳ در فضای R^5 با دو سیگنال ضربه واحد^{۱۰} ترکیب شده اند. این مجموعه دارای سه تغییر ناگهانی در زمانهای ۵۰، ۱۰۰ و ۱۵۰ می باشد. جریان داده ی دو کلاسه برای یک مجموعه ضربه واحد در یک بعد در شکل ۳ نشان داده شده است. در مقابل، برای تغییرات تدریجی یک مجموعه ضربه سینوسی^{۱۱} در فضای R^3 در نظر گرفته شده است. مجموعه ی ضربه سینوسی یک جریان داده ی دو کلاسه است که هر بعد آن بوسیله ی ترکیب نمونه ها با یک توزیع دو متغیره گوسی با میانگین و واریانس ۱ و ۰٫۳ و سیگنال ضربه سینوسی با اختلاف فاز $\pi/2$ تولید شده است. در شکل ۳ مجموعه داده های ضربه سینوسی در یک بعد نشان داده است.

مجموعه داده های واقعی

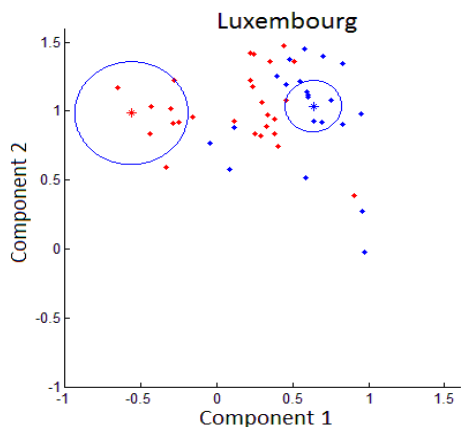
برای بررسی عملکرد روش پیشنهادی روی داده های دنیای واقعی، همانگونه که در جدول ۱ نشان داده شده است از پنج مجموعه داده استفاده شده است. مجموعه ی Luxembourg که از European Social Survey [۲۳] ساخته شده است شامل جواب های شخصی به سوالات یک پرسشنامه می باشد. این مجموعه در طول یک دوره ۵ ساله جمع آوری شده است. بعد از تبدیل ویژگی ها همانگونه که در [۲۴] آمده است، مجموعه شامل ۱۹۰۱ نمونه با ۳۱ ویژگی با مقدار حقیقی می باشد.

مجموعه ی Elec2 بوسیله ی هرییس [۲۷] بر اساس بازار الکترونیسته ی نیو ساوت ولز استرالیا عرضه گردید. این مجموعه اشاره به قیمت برق تحت تأثیر ذخیره و تقاضا در این بازار دارد. برای آزمایش ۲۹۵۶ نمونه از یک دوره ی ۳ ماهه از ۱۱ می ۱۹۹۷ تا ۱۱ جولای ۱۹۹۷ انتخاب شده است. هر نمونه دارای ۶ ویژگی می باشد.

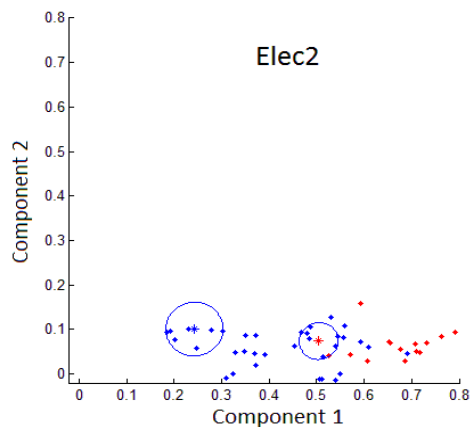
مجموعه داده ی واقعی دیگری که برای آزمایشات ما مورد استفاده قرار گرفته است، مجموعه ی Chess می باشد. این مجموعه داده که در [۲۸] معرفی شده است، شامل اطلاعات ثبت شده بازی یک بازیکن شطرنج در یک دوره زمانی از دسامبر ۲۰۰۷ تا مارس ۲۰۱۰ می باشد.

داده ها از درگاه chess.com جمع آوری شده و شامل ۵۳۲ نمونه بوده و هر نمونه دارای ۶ ویژگی می باشد. دو مجموعه داده واقعی آخر، Ionosphere و Vote می باشند که از مخزن یادگیری ماشین UCI [۲۹] انتخاب شده اند. مجموعه ی vote شامل ۴۳۵ نمونه در فضای R^{16} بوده و مجموعه ی Ionosphere نیز دارای ۳۵۱ نمونه

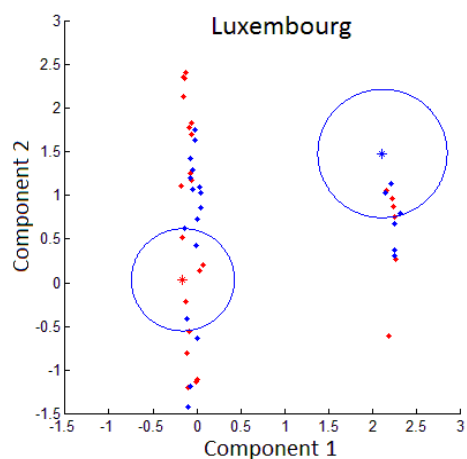
10 Step pulse
11 Sinusoidal pulse



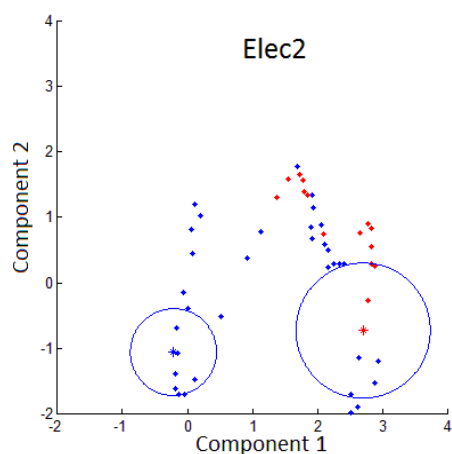
(a)



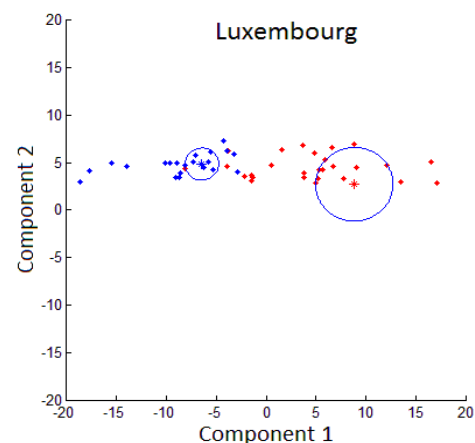
(b)



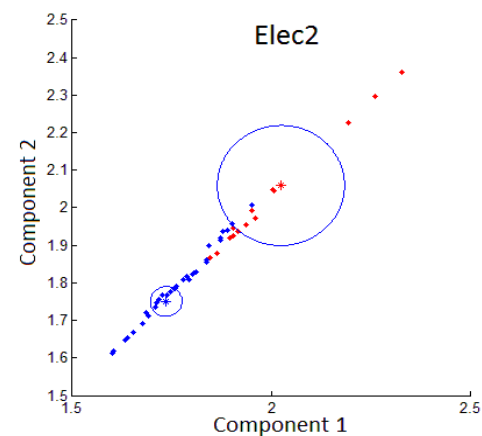
(c)



(d)

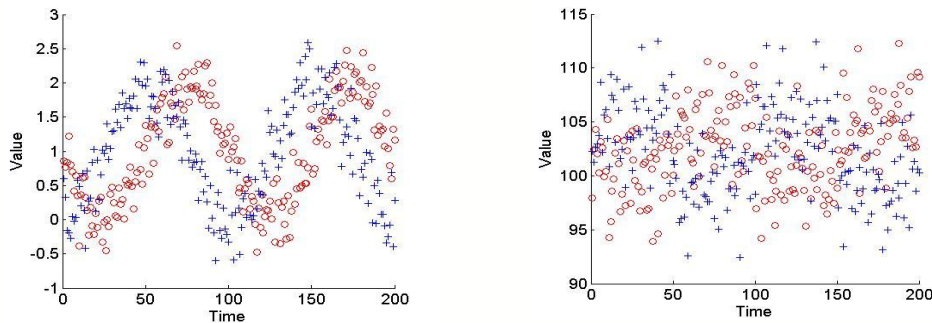


(e)



(f)

شکل ۲: نمایش ۲ بعدی نتیجه اعمال CCIPCA (a,b), RCA (c,d) و JITDML (e,f) بر روی مجموعه داده های آموزشی Luxembourg (۳۱ بعدی) و Elec2 (۶ بعدی). نقاط نمایش داده شده توسط ستاره نمایانگر نقاط آینده در هر کلاس بوده و محدوده ۷ نزدیکترین همسایه آنها توسط دایره نشان داده شده است. در شکل فقط ۵۰ نقطه در زمان بین ۱۵۰ تا ۲۰۰ نمایش داده شده اند.



شکل ۳: مجموعه داده‌های گوسی یک بعدی در جریان داده. (راست) مجموعه ضربه واحد با تغییرات ناگهانی در زمانهای ۵۰، ۱۰۰ و ۱۵۰. (چپ) مجموعه ضربه سینوسی با تغییرات تدریجی در زمان.

جدول ۱: مجموعه داده‌ها و مشخصات آنها

مجموعه	نوع داده	نوع تغییر	منبع تغییر	تعداد نمونه‌ها	ابعاد
ضربه واحد	مصنوعی	۳ تغییر ناگهانی	ساختگی	۴۰۰	۵
ضربه سینوسی	مصنوعی	تدریجی	ساختگی	۴۰۰	۳
Vote	واقعی	تدریجی	شبیه سازی شده	۴۳۵	۱۶
Ionosphere	واقعی	تدریجی	شبیه سازی شده	۳۵۱	۳۴
Luxembourg	واقعی	تدریجی	واقعی	۱۹۰۱	۳۱
Elec2	واقعی	تدریجی	واقعی	۲۹۵۶	۶
Chess	واقعی	تدریجی	واقعی	۳۵۲	۶

معیارهای ارزیابی

روش در طبقه‌بندی را نشان می‌دهد. خطای Cross Entropy در فضای تصمیم برای نمونه ی X_{N+1} که متعلق به کلاس C_{N+1} بوده و در زمان t_{N+1} بر اساس مشاهدات قبلی (X_1 تا X_N) بدست آمده است، بصورت زیر تعریف می‌شود:

$$CrossEntropy(X_{N+1}) = - \sum_{i=1..N} M_i^{C_{N+1}} \cdot \ln(L(X_i, X_{N+1})) \quad (11)$$

بطوریکه N تعداد داده‌های مجموعه‌ی آموزشی می‌باشد و

$$M_i^{C_{N+1}} = \begin{cases} 1 & X_i \in C_{N+1} \\ 0 & else \end{cases} \quad (12)$$

همچنین $L(X_i, X_{N+1})$ مشابهت X_i و X_{N+1} را نشان می‌دهد:

$$L(X_i, X_{N+1}) = e^{-\frac{d(X_i, X_{N+1})}{\sum_{j=1..N} d(X_j, X_{N+1})}} \quad (13)$$

همانطور که قبلاً ذکر شد، هدف اصلی یادگیری معیار فاصله ایجاد یک متریک مناسب جهت افزایش کارایی طبقه‌بندی کننده‌ها می‌باشد. بنابراین معیارهای در نظر گرفته شده برای ارزیابی، میزان عملکرد طبقه بندی کننده بر روی داده‌ها در فضای جدید را مورد بررسی قرار می‌دهند. معیار ارزیابی ساده‌ای که غالباً برای ارزیابی عملکرد یک طبقه‌بندی کننده استفاده می‌شود دقت طبقه‌بندی می‌باشد [۳۰]. این معیار گاهی اوقات بصورت درصد خطای طبقه‌بندی نادرست استفاده می‌شود:

$$Misclassification Error = 1 - p/q \quad (10)$$

بطوریکه p تعداد داده‌هایی که بصورت صحیح طبقه‌بندی شده‌اند و q تعداد کل داده‌ها می‌باشد. دومین معیار در نظر گرفته شده در این مقاله جهت تعیین خطا، میزان Cross Entropy می‌باشد. در قالب همسایگی می‌توان تابع خطای Cross Entropy را منفی لگاریتم تابع شباهت تعریف کرد. در آزمایشات مقدار پائین این معیار پتانسیل بالای

بنابراین:

مقدار P-value در جدول نشان داده شده است. نتایج این آزمون نشان می‌دهد که بهبود کارایی در روش JITDML نسبت به سایر روشهای مقایسه شده با ۹۳٪ سطح اطمینان از نظر آماری قابل اهمیت است.

$$CrossEntropy(X_{N+1}) = \sum_{i=1..N} M_i^{c_{N+1}} \cdot \frac{d(X_i, X_{N+1})}{\sum_{j=1..N} d(X_j, X_{N+1})} \quad (14)$$

الگوریتمهای مقایسه‌ای و نتایج آزمایشات

در ادامه برای مقایسه کارایی روش پیشنهادی در محیطهای پویا از چهار الگوریتم مواجهه با تغییر مفهوم در محیطهای غیر ایستا و یک الگوریتم استخراج ویژگی غیر پارامتری که بطور پیوسته در طول عمر سیستم بروز می‌گردد، استفاده شده است. تمام این روشها به منظور فراهم آوردن مجموعه داده‌های آموزشی مناسب جهت افزایش کارایی طبقه‌بندی کننده ها پیشنهاد شده‌اند.

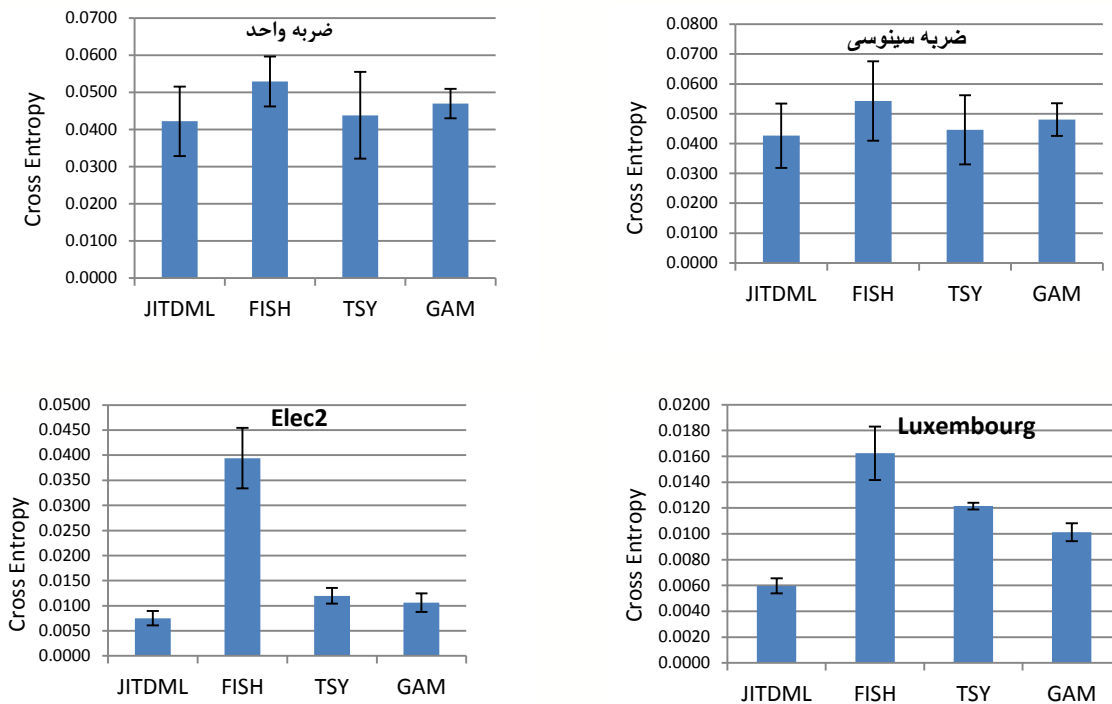
اولین الگوریتم برای مواجهه با تغییر مفهوم، Unified Instance Selection Algorithm (FISH) است که از میزان شباهت در دامنه زمان و مکان برای انتخاب مجموعه داده‌های آموزشی استفاده می‌کند [۳]. در الگوریتم انتخاب مجموعه داده آموزشی [۷] که به اختصار TSY نامیده می‌شود، از انتخاب بهترین طبقه‌بندی کننده در میان مجموعه طبقه‌بندی کننده‌ها برای مواجهه با تغییر مفهوم استفاده می‌گردد. الگوریتم پنجره آموزشی [۲] که در این مقاله که به اختصار GAM نامیده می‌شود، یک آشکار کننده نقطه شکست را برای ایجاد پنجره بهینه‌ای از داده‌های آموزشی بکار می‌برد. روش اکثریت وزن پویا (DWM) [۶] از طبقه‌بندی کننده تجمیعی پویا برای پاسخ به تغییرات محیطی بهره می‌برد. از سوی دیگر تحلیل مولفه همسایگی NCA یکی از روشهای غیر پارامتری استخراج ویژگی جهت افزایش کارایی طبقه بندی کننده‌ها در محیطهای ایستا می‌باشد و با بروز رسانی پیوسته پارامترهای آن در طول عمر سیستم به عنوان یک محک مناسب در آزمایشات مورد استفاده قرار می‌گیرد. برای انجام آزمایشات از طبقه بندی کننده KNN (K=3, 5, 7, 11) به عنوان طبقه‌بندی کننده پایه ای استفاده شده است. همچنین از قالب آزمایشی Prequential که برای آزمایشات در محیطهای پویا پیشنهاد شده است [۳۱] به عنوان تکنیک اعتبار سنجی استفاده می‌گردد. جدول ۲ نتیجه آزمایشات انجام شده بر روی مجموعه داده‌های آموزشی ذکر شده را نشان می‌دهد. همانطور که در جدول دیده می‌شود، JITDML در بیشتر حالات به غیر از حالتی که مجموعه داده آموزشی ضربه واحد و Luxembourg باشد، درصد خطای طبقه‌بندی کمتری نسبت به روشهای دیگر خواهد داشت. همچنین میانگین خطای طبقه‌بندی کننده‌ها بر روی مجموعه داده‌های آموزشی فراهم شده توسط روشهای مختلف نشان می‌دهد که JITDML در ایجاد داده‌های آموزشی جهت یادگیری طبقه‌بندی کننده KNN موفق‌تر عمل می‌نماید. برای مشخص کردن اهمیت بهبود کارایی در روش JITDML به لحاظ آماری، از آزمون علامت ویلکاکسون^{۱۲} بین نتایج JITDML و سایر تکنیک‌ها استفاده شده و

12 Wilcoxon signed rank test

جدول ۲: دقت اندازه‌گیری شده طبقه‌بندی کننده KNN با مقادیر مختلف k

	k	JITDML	FISH	TSY	DWM	GAM	NCA
ضربه واحد	3	31.325	46.3	34.6875	32.65	32.525	30.425
	5	31.8375	43.8875	35.4375	32.9125	32.8625	30.9
	7	31.925	41.875	35.7375	32.925	32.7875	31.775
	11	32.7875	39.0375	36.3375	32.475	32.1625	32.65
ضربه سینوسی	3	34.505	48.5644	45.7673	46.9431	45.2475	35.6643
	5	35.4703	46.151	45.9282	46.4975	44.9381	36.2871
	7	36.3861	45.3465	46.5718	46.745	44.8267	36.9801
	11	37.2896	45.099	46.4728	46.5347	44.8267	39.9257
Vote	3	5.977	8.046	14.023	8.7356	8.9655	6.8966
	5	5.5172	8.7356	14.9425	8.7356	8.9655	6.6667
	7	5.7471	9.1954	15.4023	8.7356	8.9655	7.1264
	11	6.4368	10.1149	15.6322	8.7356	8.9655	7.5862
Ionosphere	3	8.2621	10.5413	10.5413	10.2564	11.6809	9.9715
	5	8.8319	10.5413	10.5413	10.2564	11.6809	9.6866
	7	9.1168	10.5413	10.5413	10.2564	11.6809	9.1168
	11	8.8319	10.5413	10.5413	10.2564	11.6809	10.2564
Luxembourg	3	7.6276	32.404	30.4577	28.8269	29.879	5.05
	5	7.7854	33.5613	31.9306	28.8269	29.879	5.313
	7	7.838	31.9306	31.1415	28.8269	29.879	5.6286
	11	7.6276	31.2993	31.4571	28.8269	29.879	5.4182
Elec2	3	13.2273	15.9675	18.3356	15.2571	13.9378	13.5656
	5	12.9905	16.8471	20.9066	15.2571	13.9378	13.8701
	7	13.092	17.8958	22.6996	15.2571	13.9378	14.6482
	11	12.7876	20.5683	26.8268	15.2571	13.9378	16.1367
Chess	3	30.0188	32.2702	34.334	33.5835	36.2101	30.9568
	5	29.2683	33.2083	34.334	33.5835	36.2101	29.6435
	7	27.955	31.8949	34.7092	33.5835	36.2101	30.2064
	11	28.1426	34.334	37.8987	33.5835	36.2101	30.5816
Average		18.8788	27.3821	28.0049	25.1543	25.4596	19.3905
P-Value			3.70E-06	3.70E-06	4.20E-06	4.20E-06	0.0679

1- Wilcoxon signed rank test



شکل ۴: میزان خطای Cross Entropy اندازه گیری شده در روشهای مختلف مواجهه با تغییر مفهوم در محیطهای غیر ایستا

با افزایش λ و بروزرسانی سریع تر ماتریس انتقال از خطای طبقه بندی کننده کاسته شده و کمترین خطای طبقه بندی کننده زمانی حاصل می گردد که مقدار λ برابر ۱ باشد. شکل ۶ نیز تعداد فراخوانی های معادله ۷ برای بروزرسانی ماتریس انتقال را بصورت نرمال شده نشان می دهد. همانطور که دیده می شود با افزایش λ تعداد فراخوانی های معادله ۷ افزایش یافته و در نتیجه حجم محاسبات افزایش می یابد. بنابراین λ را میتوان به عنوان یک پارامتر تنظیم کننده دقت طبقه بندی کننده و حجم محاسبات در نظر گرفت.

نتیجه گیری

در این مقاله یک روش تطبیقی به موقع برای یادگیری معیار فاصله در محیطهای پویا ارائه گردید. در این روش غیر پارامتری، یک تابع هدف مبتنی بر تئوری اطلاعات جهت یادگیری و بدست آوردن ماتریس انتقال بهینه پیشنهاد شد. با انتقال داده ها به فضای ویژگی جدید طبقه بندی داده ها با دقت بالایی انجام خواهد شد. همچنین برای جلوگیری از بروزرسانی مداوم و تطبیق به موقع ماتریس انتقال، کارایی تخمینی طبقه بندی کننده با مقدار حد آستانه λ مقایسه می گردد. از طرف دیگر برخلاف روش های بررسی

در آزمایش قبل روشهای مختلف از نقطه نظر کارایی با در نظر گرفتن KNN به عنوان طبقه بندی کننده پایه مورد ارزیابی قرار گرفتند. در آزمایش دیگر میانگین خطای Cross Entropy برای تکنیک های مختلف که نمایانگر کارایی آنها در ایجاد مجموعه داده ای مناسب مستقل از نوع طبقه بندی کننده می باشد محاسبه گردیده و نتایج در شکل ۴ نشان داده شده است. همانطور که در شکل دیده می شود JITDML میزان خطای Cross Entropy کمتری نسبت به روشهای دیگر بر روی داده های آموزشی Luxembourg, Elec2, ضربیه واحد و ضربیه سینوسی ایجاد می نماید. این به معنی ایجاد معیار فاصله و فضای ویژگی مناسبتر از طریق یادگیری داده های آموزشی برای طبقه بندی کننده ها می باشد.

در آزمایشات قبل برای دست یافتن به حداکثر دقت، مقدار λ برابر ۱ در نظر گرفته شده است. در این صورت حداکثر تطبیق پذیری ماتریس انتقال با تغییر شرایط محیطی بوجود خواهد آمد. در آزمایش دیگری برای نشان دادن نقش λ به عنوان نقطه تعادل بین میزان دقت و حجم محاسبات، میانگین نرخ خطای طبقه بندی کننده KNN(K=5) و تعداد بروزرسانی مورد نیاز برای تطبیق ماتریس انتقال با ازای λ های مختلف محاسبه و نتایج در شکلهای ۵ و ۶ نشان داده شده است. همانطور که در شکل ۵ دیده می شود

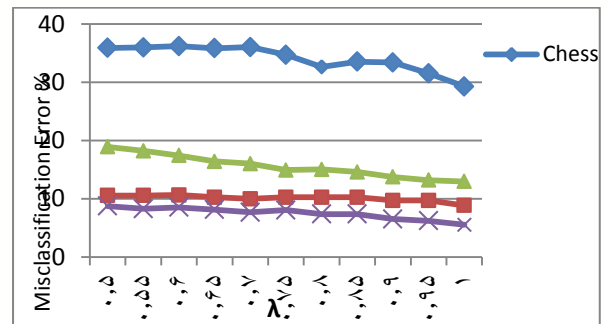
مراجع

- [1] C. Alippi and M. Roveri, "Just-in-Time Adaptive Classifiers-Part II: Designing the Classifier," *Neural Networks, IEEE Transactions on*, vol. 19, pp. 2053-2064, 2008.
- [2] J. o. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in *In SBIA Brazilian Symposium on Artificial Intelligence*, 2004, pp. 286-295.
- [3] I. Zliobaite, "Combining similarity in time and space for training set formation under concept drift.," *Intelligent Data Analysis* vol. 15, pp. 589-611, 2011.
- [4] B. J. Oommen and L. Rueda, "Stochastic learning-based weak estimation of multinomial random variables and its applications to pattern recognition in non-stationary environments," *Pattern Recognition*, vol. 39, pp. 328-341, 2006.
- [5] P. Zhang, X. Zhu, and Y. Shi, "Categorizing and mining concept drifting data streams," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining Las Vegas, Nevada, USA: ACM*, 2008.
- [6] J. Z. Kolter and M. A. Maloof, "Dynamic Weighted Majority: An Ensemble Method for Drifting Concepts," *Journal of Machine Learning Research*, vol. 8, pp. 2755-2790, 2007.
- [7] A. Tsymbal, M. Pechenizkiy, d. Cunningham, and S. Puuronen, "Dynamic integration of classifiers for handling concept drift," *Inf. Fusion*, vol. 9, pp. 56-68, 2008.
- [8] Y. Chen, E. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti, "Similarity-based Classification: Concepts and Algorithms," *The Journal of Machine Learning Research*, vol. 10, pp. 747-776, 2009.
- [9] A. M. Namboodiri and A. K. Jain, "Online handwritten script recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 124-130, 2004.
- [10] Z. Chen, L. Di, G. Yu, and N. Chen, "Real-Time On-Demand Motion Video Change Detection in the Sensor Web Environment," *The Computer Journal*, vol. 54, pp. 2000-2016, November 1, 2011.
- [11] J. J. Pantrigo, A. Sánchez, and J. Mira, "Representation spaces in a visual-based human action recognition system," *Neurocomputing*, vol. 72, pp. 901-915, 2009.
- [12] J. Wen, X. Gao, Y. Yuan, D. Tao, and J. Li, "Incremental tensor biased discriminant analysis: A new color-based visual tracking method," *Neurocomputing*, vol. 73, pp. 827-839, 2010.
- [13] B. Bahraminejad, S. Basri, M. Isa, and Z. Hambli, "Application of a sensor array based on capillary-attached conductive gas sensors for odor identification," *Measurement Science and Technology*, vol. 21, p. 085204, 2010.

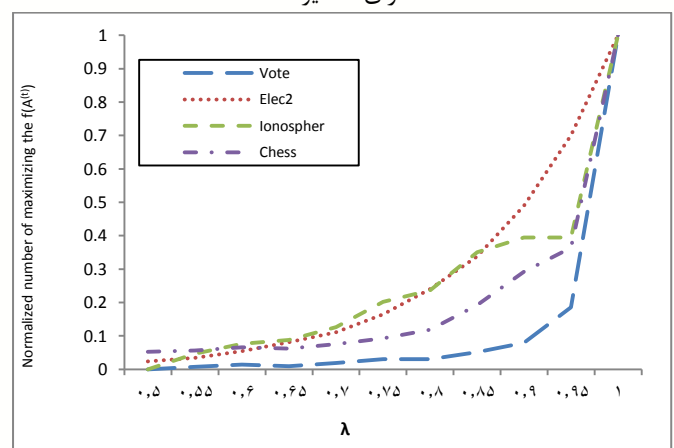
شده دیگر به عنوان روشهای مقایسه‌ای، تکنیک پیشنهادی نیازی به استفاده از روشهای‌های آشکار کننده نقطه شکست و ایجاد پنجره‌های بهینه از داده‌ها ندارد. نتایج آزمایشات بر روی داده‌های واقعی و مصنوعی نشان می‌دهد که معیار فاصله یادگیری شده توانایی ایجاد فضای مناسب‌تری جهت طبقه‌بندی داده‌ها در محیطهای پویا را دارا می‌باشد. این نتیجه توسط بررسی دو معیار ارزیابی اندازه گیری شده مانند خطای طبقه‌بندی کننده و خطای Cross Entropy که بطور مستقیم با دقت طبقه‌بندی در ارتباط می‌باشند، نشان داده شده است. تحقیقات آینده در این زمینه می‌تواند به بررسی و بهبود روش پیشنهادی در محیط‌های پویا هنگامی که احتمال مشاهده داده جدید ورودی در بیش از یک مکان باشد، اختصاص یابد.

تقدیر و تشکر

این مقاله از طرح تحقیقاتی که با بودجه پژوهشی و حمایت مالی دانشگاه آزاد اسلامی واحد قزوین به انجام رسیده است استخراج گردیده است.



شکل ۵: میانگین نرخ خطای طبقه‌بندی کننده KNN(K=5) به ازای مقادیر مختلف λ



شکل ۶: تعداد بروزسانی (نرمالیزه شده) مورد نیاز برای تطبیق ماتریس انتقال با ازای λ های مختلف

- [28] I. Zliobaite, "Adaptive Training Set Formation," in Physical sciences, informatics (09P). vol. PhD Thesis: VILNIUS UNIVERSITY, 2010.
- [29] A. Asuncion and D. Newman, "UCI machine learning repository, University of California, Irvine, School of Information and Computer Sciences," 2007.
- [30] C. Ferri, J. Hernandez-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," Pattern Recognition Letters, vol. 30, pp. 27-38, 2009.
- [31] J. Gama, R. Sebasti, and P. P. Rodrigues, "Issues in evaluation of stream learning algorithms," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining Paris, France: ACM, 2009.
- [14] B. Graimann, J. E. Huggins, A. Schlogl, S. P. Levine, and G. Pfurtscheller, "Detection of movement-related patterns in ongoing single-channel electrocorticogram," Neural Systems and Rehabilitation Engineering, IEEE Transactions on, vol. 11, pp. 276-281, 2003.
- [15] A. Rizzi and R. Cassinis, "A robot self-localization system based on omnidirectional color images," Robotics and Autonomous Systems, vol. 34, pp. 23-38, 2001.
- [16] I. Katakis, G. Tsoumakas, and I. Vlahavas, "Tracking recurring contexts using ensemble classifiers: an application to email filtering," Knowledge and Information Systems, vol. 22, pp. 371-391, 2008.
- [17] W. Lam and K. Leung Yu, "High-dimensional Learning Framework for Adaptive Document Filtering " Computational Intelligence, vol. 19, pp. 42-63, 2003.
- [18] H. A. Moghaddam, M. Matinfar, S. M. Sajad Sadough, and K. Amiri Zadeh, "Algorithms and networks for accelerated convergence of adaptive LDA," Pattern Recognition, vol. 38, pp. 473-483, 2005.
- [19] W. Juyang, Z. Yilu, and H. Wey-Shiuan, "Candid covariance-free incremental principal component analysis," Pattern Analysis and Machine Intelligence ,IEEE Transactions on, vol. 25, pp. 1034-1040, 2003.
- [20] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," International Journal of Neural Systems vol. 10, 2000.
- [21] J. Goldberger ,S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighborhood Component Analysis," in Neural Information Processing Systems, 2005.
- [22] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning Distance Functions using Equivalence Relations," in In Proceedings of the Twentieth International Conference on Machine Learning, 2003, pp. 11-18.
- [23] R. t. C. C. T. Jowell, "European social survey 2002/2003; 2004/2005; 2006/2007," in Technical Report. vol. 3: Technical Reports, London: Centre for Comparative Social Surveys, City University, 2007.
- [24] I. Zliobaite, "Combining Time and Space Similarity for Small Size Learning under Concept Drift," in Proceedings of the 18th International Symposium on Foundations of Intelligent Systems Prague, Czech Republic: Springer-Verlag, 2009.
- [25] M. M. Lazarescu, S. Venkatesh, and H. H. Bui, "Using multiple windows to track concept drift," Intelligent Data Analysis, vol. 8, pp. 29-59, 2004.
- [26] L. L. Minku, A. P. White, and Y. Xin, "The Impact of Diversity on Online Ensemble Learning in the Presence of Concept Drift," Knowledge and Data Engineering, IEEE Transactions on, vol. 22, pp. 730-742, 2010.
- [27] M. Harries, "SPICE-2 Comparative Evaluation: Electricity Pricing," Technical Report, The University of South Wales 1999.