سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران
جزیره کیش، خلیج فارس، ایران ۱۹ الی ۲۱ اسفند ۱۳۸۶
انجمن کامپیوتر ایران
Computer Society of Iran
دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر
پوستر

استفاده از ترکیب SVM فازی با برچسب گذاری صوری در بازیابی متون

محمد رحیمی
دانشکده مهندسی کامپیوتر و فناوری اطلاعات
دانشگاه صنعتی امیرکبیر
m_rahimi@aut.ac.ir

سعید شیری
دانشکده مهندسی کامپیوتر و فناوری اطلاعات
دانشگاه صنعتی امیرکبیر
shiry@ce.aut.ac.ir

چکیده: در مسائل بازیابی اطلاعات مبتنی بر بازخورد کاربر، یکی از مشکلات اصلی، کمبود تعداد نمونه های آموزشی است. دلیل این امر، عدم امکان دریافت تعداد نمونه های زیاد برچسب خورده توسط کاربر است. برای رفع این مشکل، ما در این مقاله، از روشی برای برچسب گذاری صوری متون استفاده می‌کنیم. با این کار از مزیت تعداد نمونه های آموزشی بیشتر، با نیاز کمتر به بازخورد کاربر، بهره‌مند خواهیم شد. برای دخیل کردن عدم دقت ذاتی موجود در برچسب های صوری، از مفاهیم فازی استفاده کرده و برای دسته بندی نمونه ها جهت بازیابی، از SVM که یک روش قدرتمند دسته بندی داده ها محسوب می شود استفاده کرده ایم. در ضمن جهت بهبود کارایی ، برای انتخاب نمونه های کاندید برای برچسب گذاری صوری، از روش خاصی استفاده شده است. این الگوریتم برای اولین بار در حوزه بازیابی متون مورد استفاده قرار گرفته و نتایج آزمایشات نشان می دهد که این روش، نسبت به SVM با یادگیری فعال و مبتنی بر بازخورد کاربر، نتایج بسیار بهتری ارائه می کند.

واژه های کلیدی: SVM ، منطق فازی، برچسب گذاری صوری ، بازخورد کاربر، بازیابی متون.

۱- مقدمه

در سیستم های بازیابی متن، معمولاً پرس و جویی[1] که توسط کاربر در قالب چند کلمه کلیدی وارد می شود، دارای ابهام است. مثال ساده آن استفاده از کلمه «جاوا» است که هم به عنوان زبان برنامه نویسی و هم جزیره ای در اندونزی شناخته می‌شود. معمولاً نتایج سیستم های بازیابی، تعداد کمی متن «مربوط» و تعداد زیادی متن «نامربوط» است. یکی از روش هایی که جهت رفع این مشکلات ارائه شده است، استفاده از بازخورد کاربر است. در این روش ها، با ارائه تعداد محدودی متن، از کاربر خواسته می شود تا متون مربوط و نامربوط را مشخص نماید. مشکلی که در این روش معمولاً بروز می کند این است که سیستم بازیابی، نمی تواند با دریافت تعدادی نمونه مربوط، مدل مربوط به متون مربوط و نامربوط را استخراج نماید. روشی که برای رفع این مشکل مورد استفاده قرار می گیرد، استفاده از یادگیری فعال است[10،11].

دراین روش، سیستم مجموعه‌ای از نمونه های برچسب نخورده توسط کاربر را در اختیار داشته و در هر مرحله با انتخاب برخی از نمونه ها(که بیشترین شک در مورد دسته بندی آنها وجود دارد) ، آنها را جهت برچسب گذاری در اختیار کاربر قرار می دهند. بنابراین در روش بدون بازخورد با ارائه «مربوط»ترین نمونه ها سیستم به کندی همگرا شده و در روش استفاده از بازخورد، با انتخاب «مشکوک ترین» نمونه ها، کاربر را مجبور می کنیم که تعداد زیادی نمونه نامربوط را نیز مشاهده کند. در این مقاله، ما از یک طرف برای سرعت بخشیدن به فرایند یادگیری از یادگیری فعال مبتنی بر بازخورد کاربر استفاده کرده و از طرف دیگر با برچسب گذاری صوری برخی نمونه ها، جلوی درگیر شدن کاربر با تعداد زیادی از نمونه های نامربوط گرفته می شود. برای انجام عملیات دسته بندی، از ماشین های بردار پشتیبان(SVM) استفاده کرده ایم که دارای قدرت تعمیم فوق العاده و پشتوانه تئوریک بسیار محکمی هستند.

در ادامه این مقاله، ابتدا در بخش ۲ مروری بر روش های بازیابی متن مبتنی بر SVM انجام می گیرد. در بخش ۳، مسئله تعداد کم نمونه های آموزشی توصیف و تشریح شده و در بخش ۴، الگوریتم PLFSVM معرفی می‌گردد. در بخش ۵ نحوه ارزیابی نتایج الگوریتم و در بخش ۶ نتایج حاصل از آزمایشات را بیان می‌کنیم. در نهایت در بخش ۷، جمع بندی برروی روش استفاده شده، انجام خواهیم داد.

۲- کارهای انجام شده

استفاده از ماشین های بردار پشتیبان در بازیابی متون، اولین بار در سال ۱۹۹۷ در [3] پیشنهاد گردید. در [4] بحث کاملی در مورد مزیت های استفاده از SVM و دلایل کارایی آن در بازیابی متون انجام شده است. تا کنون پژوهش‌های متعددی برای بازیابی متون با استفاده از SVM انجام گرفته است. در [5،6] برروی نحوه گزینش ویژگی ها در بازیاب متن با SVM بررسی هایی انجام گرفته است.

در [7] از ترکیب SVM به عنوان دسته بندی کننده و PSO برای گزینش از میان مجموعه ای از دسته بندی کننده های ایجاد شده توسط SVM ، برای بازیابی متون استفاده شده است. در [8] با ارائه روشی برای تنظیم مقدار آستانه ای در SVM ، کارایی سیستم‌های

بازیابی متن مبتنی بر SVM ، افزایش داده شده است. در [9] روش بازیابی متن مبتنی بر SVM و بازخورد کاربر با برخی روش های متداول دیگر در بازیابی متون مقایسه شده و با ارائه نتایج، کارایی بهتر آن نسبت به سایر روش ها ، گزارش شده است. در [10] روش هایی برای انتخاب بهترین نمونه برای برچسب گذاری در الگوریتم SVM با یادگیری فعال ارائه شده و این روش ها با یکدیگر مقایسه شده اند. کارایی هر یک از روش‌ها در حوزه بازیابی متون مورد بررسی قرار گرفته اند. در [14] روشی برای گزینش بهترین نمونه ها برای برچسب گذاری در SVM مبتنی بر بازخورد کاربر، ارائه شده (به نام HRFSVM) و با دو روش دیگر SVM مقایسه شده است. نتایج این روش برای بازیابی متون مورد استفاده قرار گرفته و برتری آن نشان داده شده است.

تمامی روش های مبتنی بر بازخوردی که در فوق به آنها اشاره کردیم، با مشکل تعداد نمونه هایی آموزشی مواجه هستند. در آن روش‌ها، الگوریتم هایی برای انتخاب نمونه مناسب برای برچسب گذاری ارائه شده است اما در هر صورت، نیاز به دریافت تعداد زیادی نمونه آموزشی از کاربر وجود دارد. در روشی که ما در این مقاله استفاده می کنیم، سعی شده است تا این مشکل با برچسب گذاری صوری برخی نمونه های برچسب گذاری نشده توسط کاربر رفع شده، مجموعه آموزشی گسترش پیدا کرده و کارایی افزایش پیدا کند.

۳- مسئله تعداد کم نمونه های آموزشی

در سیستم های بازیابی اطلاعات مبتنی بر بازخورد کاربر، این مشکل وجود دارد که یا کاربر حوصله و تمایلی به برچسب گذاری تعداد زیادی از نمونه ها ندارد و یا اینکه اصولاً واسط کاربری ارائه شده، ظرفیت برچسب گذاری این تعداد زیاد از نمونه ها را ندارد. نتیجه این امر این است که تعداد نمونه های برچسب خورده ای که برای آموزش بکار خواهند رفت، ناکافی خواهد بود. این مسئله حتی در روش یادگیری قدرتمندی مثل SVM نیز خود را نشان می دهد. بنابراین لازم است که راه حلی برای مشکل تعداد کم نمونه ها که در روش های مبتنی بر بازخورد کاربر ایجاد می شود، پیدا شود.

روشی که ما در اینجا استفاده می کنیم، با توجه به این مسئله به وجود آمده است که اولاً برچسب گذاری تعداد زیادی از نمونه ها، عملی است وقت گیر و مستلزم صرف نیروی انسانی خبره و ثانیاً مجموعه بزرگی از نمونه های برچسب نخورده را در اختیار داریم. بنابراین سعی می کنیم تا به روشی ، این نمونه های برچسب نخورده را همراه با نمونه های برچسب گذاری شده، برای انجام فرایند آموزش به صورت مؤثر، به کار بگیریم. البته باید توجه شود که انتخاب نمونه های برچسب نخورده برای برچسب گذاری صوری، باید با دقت زیاد انجام شود تا مفید واقع شوند. در غیراینصورت ممکن است در کارایی سیستم بازیابی حتی اثر منفی داشته باشند. این روش برچسب گذاری که براساس «مرتبط بودن» یا «نامربوط بودن» نمونه تحت بررسی با نمونه پرس وجو شده انجام می شود، در بخش های آینده به صورت دقیق تر توضیح داده خواهد شد. همانطور که در بخش قبل نیز ذکر شد، این برچسب گذاری دارای ماهیت غیردقیق و فازی خواهد بود.

۴- روش PLFSVM

ماشین بردار پشتیبان که توسط وپنیک[ii] براساس ایده حداقل سازی ریسک شکل گرفته است، یک روش یادگیری ماشین فوق العاده قدرتمند محسوب می شود. این روش تاکنون در حوزه های متعددی مثل دسته بندی متون، شناسایی گفتار، داده کاوی و ... به کار رفته و نشان داده است که نسبت به بسیاری روش های متعارف یادگیری ماشینی دارای برتری است[2]. ایده اصلی در SVM (برای حالت دو کلاسه)، این است که ابرصفحه بهینه برای تفکیک کلاس ها به گونه ای انتخاب شود که حاشیه بین دو کلاس، حداکثر شود. صرف نظر از اینکه این روش دارای قابلیت تعمیم بسیار زیادی است، عامل محدود کننده آن ، لزوم تعلق هر نمونه به فقط یکی از دو کلاس مذکور و آن هم با میزان اهمیت یکسان است. درحالی که در عمل، بسیاری از اوقات، نمونه های آموزشی دارای درجه اهمیت متفاوتی نسبت به یکدیگر هستند و لازم است که این مسئله در فرایند یادگیری ، مدنظر قرار بگیرد. بنابراین ماشین های بردار پشتیبان فازی پیشنهاد شده اند[8]. FSVM در واقع توسعه‌ی SVM است به گونه ای که نمونه های آموزشی دارای درجه اهمیت متفاوتی باشند. دلایلی که باعث شده است که در روش مورد بررسی از FSVM استفاده شود، اول اعمال مفاهیم فازی به مسئله، دوم وجود پایه تئوریک قوی برای آن و سوم قدرت تعمیم بسیار خوب آن است. بنابراین در روش مورد استفاده، مزایای برچسب گذاری صوری نمونه ها را با FSVM ترکیب شده است. باید توجه شود که روش PLFSVM[iii] ارائه شده، با SVM استاندارد از جنبه های مختلف، تفاوت دارد. PLFSVM سعی می کند مسئله تعداد کم نمونه های آموزشی را حل کند در حالی که SVM استاندارد فقط توانایی کار با نمونه های برچسب گذاری شده را دارد. علاوه بر این، PLFSVM نیاز به نیروی کار کمتری نسبت به SVM دارد بنابراین برای بسیاری از کاربردها که کاربر خبره به صورت گسترده و آسان در دسترس نیست(مثل سیستم های بازیابی تصویر در شبکه هایی با محدودیت پهنای باند) مناسب تر خواهد بود. همچنین PLFSVM میزان اهمیت نمونه ها را نیز در فرایند یادگیری لحاظ می‌کند که چنین اتفاقی در SVM استاندارد نمی افتد.

در روش PLFSVM پیشنهاد شده در [1] ، راه های حل چند مسئله مورد بررسی قرار می گیرد: اول، انتخاب نمونه های مناسب برای برچسب گذاری صوری، دوم، تعیین برچسب های صوری، سوم، تخمین میزان عدم قطعیت برچسب های صوری تعیین شده و چهارم، تجمیع برچسب های صوری با روش یادگیری فعال FSVM . همانطور که ذکر شد، دراین مقاله، ما از روش یادگیری فعال به صورت ترکیبی با روش

SVM استفاده می کنیم(ترکیب این دو روش در حوزه بازیابی تصاویر در [13] انجام شده است). بنابراین به کارایی بالاتر با نیاز کمتر به بازخورد کاربر خواهیم رسید.

مراحل کلی الگوریتم PLFSVM پیشنهادی در [1] عبارتند از:

- پرس وجو از کاربر دریافت شده و سیستم یک مرحله الگوریتم KNN را با درنظر گرفتن فاصله اقلیدسی اجرا می کند. نتیجه این کار پیدا کردن l_0 عدد نمونه که بیشترین شباهت را به پرس وجو دارند می باشد. سپس این نمونه ها برای برچسب گذاری در اختیار کاربر قرار داده می شوند.
- کاربر برروی l_0 نمونه دریافت شده، برچسب گذاری کرده و هریک از آنها را به عنوان «مربوط»(کلاس ۱) و «نامربوط»(کلاس ۲) شناسایی می کند. براساس این برچسب گذاری، یک SVM ساده اولیه آموزش داده می شود.
- با استفاده از ایده یادگیری فعال در SVM، تعداد l نمونه که نزدیک تر از بقیه نمونه ها به مرز تصمیم گیری(ابرصفحه تفکیک کننده) هستند، برای برچسب گذاری مجدد کاربر انتخاب می-شوند. ما در این مقاله برای افزایش کارایی سیستم، بجای استفاده صرف از نمونه های نزدیک تر به مرز تصمیم گیری که «مشکوک ترین» نمونه ها از لحاظ دسته بندی هستند، از دو دسته نمونه استفاده کرده ایم. اول نمونه هایی که در جهت موافق بیشترین فاصله را با مرز دارند و دوم، نمونه هایی که کمترین فاصله را از مرز تصمیمی گیری دارند[14]. نسبت نمونه‌های اول به دوم بیشتر خواهد بود. در حالت SVM با یادگیری فعال فقط از نمونه های دوم استفاده می شود.
- پس از برچسب گذاری نمونه ها توسط کاربر، آنها را به مجموعه نمونه های برچسب خورده قبلی برای آموزش، اضافه می‌کنیم.
- دو مرحله عملیات خوشه بندی برروی مجموعه نمونه های مربوط و نامربوط انجام می شود. خوشه های شکل گرفته برای انتخاب نمونه های کاندید جهت برچسب گذاری صوری ، مورداستفاده قرار می گیرند.
- یک معیار فازی برای تعیین درجه مربوط یا نامربوط بودن نمونه‌ها مورد استفاده قرار می گیرد.
- از یک ماشین بردار پشتیبان فازی برای آموزش نمونه های برچسب گذاری شده اصلی وصوری به صورت توأم استفاده می-شود.
- مراحل ۳ تا ۷ آنقدر تکرار می شوند تا نتیجه مطلوب حاصل شود.

در ادامه به توضیح برخی از مراحل اصلی ذکر شده فوق می‌پردازیم.

۱-۴ FSVM

در این بخش، پیش از معرفی FSVM ، برای پرهیز از ذکر موارد اضافی، ما با فرض اینکه خواننده با SVM استاندارد آشنایی دارد، فقط نحوه فرمول بندی FSVM را نشان خواهیم داد. در FSVM معیار(۱) باید تحت شرایط (۲) حداقل شود.

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} m_i x_i \qquad (۱)$$

$$y_i(w.z_i + b) \geq 1 - x_i \qquad x_i \geq 0 \quad i = 1,...,n \qquad (۲)$$

در فرمول فوق، μ_i میزان تعلق فازی را نشان می دهد. درواقع از مقادیر تعلق به منظور وزن دهی به نمونه‌های آموزشی به کار می رود. نمونه های با مقادیر تعلق بالاتر، نقش بیشتری در فرایند آموزش خواهند داشت تا نمونه های کم اهمیت تر. مشابه حالت معمولی، در اینجا نیز فرم دوگان مسئله به شکل فرمول های (۳) و (۴) خواهد بود.

$$\sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j K(X_i, X_j) \qquad (۳)$$

$$\sum_{i=1}^{n} y_i a_j = 0, \qquad 0 \leq a_i \leq m_i C \quad i = 1,...,n \qquad (۴)$$

نتیجه این روش، یافتن ابرصفحه تفکیک کننده مشابه حالت SVM استاندارد است اما با بردارهای پشتیبان و α_i های متفاوت. بنابراین با یافتن μ_i های مناسب می توان میزان اثر نمونه i ام در فرایند یادگیری را کنترل کرد.

۲-۴ انتخاب نمونه‌های کاندید برای برچسب‌گذاری صوری

برای این کار، از یک فرایند خوشه بندی دو مرحله ای برای یافتن خوشه های محلی استفاده می کنیم[2]. در مرحله اول، با انجام یک عملیات خوشه بندی کاهشی[iv]، تعداد خوشه های مناسب تخمین زده می شود. این روش علاوه بر سریع بودن، نیازی به دانش اولیه درمورد تعداد خوشه ها ندارد. سپس از این مقدار تخمینی به عنوان ورودی الگوریتم K-means استفاده می گردد.

پس از خوشه بندی، دو مجموعه مستقل از خوشه های مربوط و نامربوط در اختیار خواهیم داشت. در اینجا از ایده ای مشابه KNN استفاده می کنیم. برای انتخاب نمونه های برچسب نخورده برای برچسب گذاری صوری، نزدیک ترین نمونه های برچسب نخورده به مراکز خوشه ها را برمی‌گزینیم. یعنی به ازای هر خوشه، یک نمونه پیدا شده و برای برچسب گذاری صوری ارائه می گردد. سپس کلاس مربوط به آن خوشه، به نمونه مذکور تعلق می‌گیرد. از آنجایی که با افزایش تعداد نمونه های لازم برای برچسب گذاری صوری، هزینه محاسباتی زیاد خواهد شد، فقط از نزدیک ترین نمونه‌ی همسایه استفاده می‌شود.

۳-۴ تعیین میزان ابهام (تعلق فازی) نمونه‌ها

در [2]، یک تابع تعلق فازی برای این منظور پیشنهاد شده است. از آنجایی که نمونه های کاندید برای برچسب گذاری صوری، براساس خوشه بندی برروی نمونه های دارای برچسب انتخاب شده اند، می‌توان از همان خوشه ها برای تعیین میزان تعلق استفاده نمود. برای این منظور، فرض کنید که برچسب‌های هر یک از نمونه های مذکور، در مرحله قبل تعیین شده‌است. دراین حالت، نزدیک تر بودن یک نمونه به خوشه ای با کلاس موافق، به معنای میزان تعلق بیشتر آن به کلاس مذکور است. از طرف دیگر، نزدیک بودن آن به خوشه ای با کلاس مخالف، نشانه تعلق کمتر آن به کلاس موافق خواهد بود. براساس این موضوع، می توان یک تابع فازی نمایی تعریف کرد که این مفاهیم را در درون خود ایجاد کند[2]. این تابع را W_1 نامگذاری می کنیم.

از طرف دیگر، میزان تعلق فازی نمونه مذکور، باید به دسته بندی که توسط SVM اولیه صورت گرفته است نیز وابستگی داشته باشد. برای این منظور، از یک تابع سیگموئید طبق فرمول (۵) استفاده می شود.

$$w_2(X_p)=\begin{cases}\dfrac{1}{1+\exp(-a_2 y)} & class1\\ \dfrac{1}{1+\exp(a_2 y)} & class2\end{cases} \qquad (۵)$$

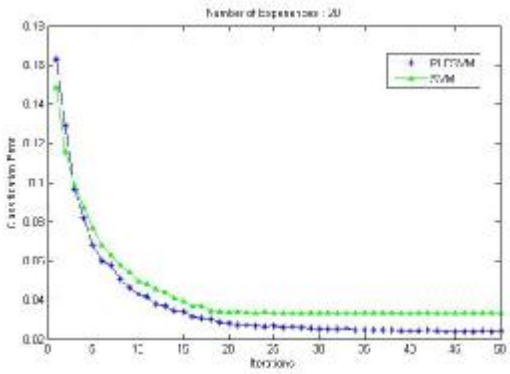
در این فرمول، a_2 فاکتور مقیاس و y فاصله مستقیم نمونه X_P از مرز SVM خواهد بود. ایده به کاررفته در این فرمول به این صورت است که اگر کلاس تعیین شده برای نمونه مورد بررسی(X_P) توسط خوشه بندی، «موافق» باشد و فاصله آن از مرز SVM ، درجهت موافق زیاد باشد، معنای آن این است که این نمونه به درستی برچسب خورده است. در نهایت برای اینکه هر دو عامل را در تعیین میزان تعلق یک نمونه به یک کلاس دخیل کرده باشیم، مقدار تعلق را براساس ضرب مقادیر بدست آمده از W_1 و W_2 ، محاسبه می کنیم.

۵- نحوه ارزیابی نتایج

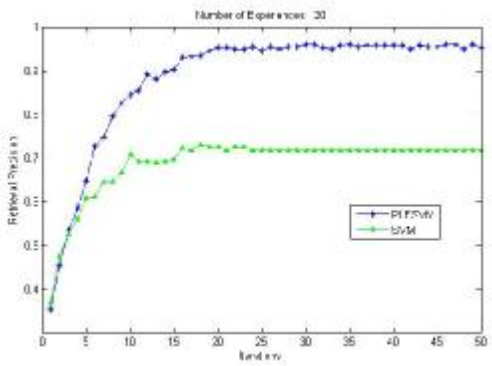
برای ارزیابی نتایج بدست آمده از روش استفاده شده، ما از دو معیار استفاده می کنیم. معیار اول، معیار دقت بازیابی است که در [14] نیز برای ارزیابی الگوریتم بازیابی متن مورد استفاده قرار گرفته است. معیار دوم استفاده شده، خطای دسته بندی است که در [10] نیز برای ارزیابی الگوریتم بازیابی متن مورد استفاده قرار گرفته است.

۶- آزمایشات

برای انجام آزمایشات، به عنوان مجموعه داده آزمون از مجموعه newsgroups [v] استفاده کردیم. این مجموعه از ۲۰ عنوان مختلف تشکیل شده است که ما از هر عنوان ۵۰ مستند(و در مجموع ۱۰۰۰ مستند) را به صورت تصادفی انتخاب کردیم. پس از آن با استفاده از نرم افزار pretext [vi] عملیات پیش پردازشی لازم برروی مستندات انجام شده و تعداد در حدود ۲۰۰۰۰ ویژگی از آنها استخراج شده و بردار مستندات تولید گردید. ویژگی های مستندات را کلمات موجود در آنها تشکیل می دهند. در طی عملیات تشکیل بردار مستندات، کلمات با تکرار کمتر از ۲ و بیشتر از ۱۰۰ حذف شده ، ریشه یابی برای حذف حالت های مختلف کلمات و افعال انجام شده و کلمات ایست نیز حذف شده اند. همچنین از tfidf برای وزن دهی به ویژگی ها(کلمات) استفاده شد. پس از تشکیل بردار های نهایی از مستندات، برای کاهش ابعاد مستندات از PCA [vii] استفاده کرده و تعداد ویژگی ها را به ۴۰۰ عدد رساندیم. پیاده سازی الگوریتم با نرم افزار MATLAB انجام گرفت و پارامترهای l_0 و l به ترتیب برابر با ۵ و ۱۰ درنظر گرفته شد. نتایج حاصل از اجرای الگوریتم های PLFSVM و SVM با یادگیری فعال در شکل های (۱) و (۲) دیده می شود.



شکل (۱): مقایسه خطای دسته بندی در PLFSVM و SVM



شکل (۲): مقایسه دقت بازیابی در PLFSVM و SVM

نمودارهای شکل‌های فوق برای تعداد ۵۰ تکرار فرایند آموزش SVM و با میانگین گیری از ۲۰ بار انجام آزمایش بدست آمده اند. در هر آزمایش، با انتخاب یک متن تصادفی به عنوان پرس وجو، عملیات بازیابی انجام گرفته است. همانطور که مشاهده می شود، هردو نمودار کارایی بهتر روش PLFSVM را نسبت به SVM با یادگیری فعال و بازخورد کاربر ، نشان داده اند(خطای درحدود ۰.۰۲ و دقت در حدود ۰.۹۵) . در هر دو نمودار ، در تکرار های اولیه، تفاوت چندانی میان دو روش مشاهده نمی شود اما با افزایش تکرارها، استفاده از برچسب های

صوری تولید شده، تأثیر بسیار زیاد خود را نشان می دهند. این نتایج در ضمن نشان می دهند که تخمین به کاررفته در نحوه انتخاب نمونه های کاندید و همچنین تعیین تابع تعلق فازی نمونه‌ها، به خوبی انجـام شـده است.

۷- نتیجه گیری

در این مقاله، روشی مبتنی بر ماشین بردار پشـتیبان فـازی و برچسـب گذاری صوری برای رفـع مشـکل تعـداد کـم نمونـه هـای آموزشـی در عملیات بازیابی متون مبتنی بر بازخورد کاربر، معرفی شد. نتایج بدسـت آمده، کارایی بسیار خوب این روش نسبت به SVM با یادگیری فعـال و بازخورد کاربر را نشان می دادند.

از آنجـایی کـه در الگـوریتم PLFSVM اسـتفاده شـده، از روش هـای تخمینی برای انتخاب نمونه های کاندید برای برچسب گذاری صـوری و همچنین تعیین تابع تعلق مناسب برای آنها، استفاده گردیده اسـت، بـه نظر می رسد برای کارهای آینده، می تـوان از روش‌تخمینـی بهتـر نیـز برای این کار، استفاده نمود و نتایج را با این روش مقایسه کرد.

۸- مراجع

[1] K. Wu, K. H. Yap, "*A Pseudo-Labeling Framework for Content-based Image Retrieval*", IEEE Symposium on Computational Intelligence in Image and Signal Processing, 2007.

[2] K. Wu, K. H. Yap," *Fuzzy SVM for content-based image retrieval: a pseudo-label support vector machine framework*", IEEE Computational Intelligence Magazine, Volume 1, Issue 2, Page(s):10 – 16, 2006.

[3] T. Joachims, "*Text categorization with support vector machines: Learning with many relevant features*", Technical Report 23, Universitat Dortmund, LS VIII, 1997.

[4] T. Joachims."*Text categorization with support vector machines*". In Proceedings of the European Conference on Machine Learning. Springer Verlag, 1998.

[5] E. Leopold,J. Kindermann, "*Text Categorization with Support Vector Machines.How to Represent Texts in Input Space?*", Machine Learning, 46, 423–444, 2002.

[6] T. Masuyama, H. Nakagawa,"*Cascaded Feature Selection in SVMs Text Categorization*", LECTURE NOTES IN COMPUTER SCIENCE, ISSU 2588, pages 588-591, 2003.

[7] T. Peng · W. Zuo · F. He, "*SVM based adaptive learning method for text classification from positive and unlabeled documents*", Knowl Inf Syst, Springer-Verlag 2007.

[8] J. Shanahan , N. J. G. Roma , "*Improving SVM Text Classification Performance through Threshold Adjustment*" , LECTURE NOTES IN COMPUTER SCIENCE, ISSU 2837, pages 361-372, 2003.

[9] H. Drucker, B. Shahrary, D.C. Gibbon, "*Support vector machines: relevance feedback and information retrieval*" , Information Processing and Management 38, p305-323, 2002.

[10] S.Tong,D. Koller, "*Support Vector Machine Active Learning with Applications to Text Classification*", Journal of Machine Learning Research, p45-66, 2001.

[11] G. Schohn and D. Cohn. "*Less is more: Active learning with support vector machines*". In Proceedings of the Seventeenth International Conference on Machine Learning, 2000.

[12] C.F. Lin , S.D. Wang, "*Fuzzy support vector machines,''* IEEE Trans. Neural Networks, vol. 13, no. 2, pp. 464–471, Mar. 2002.

[13] S. Tong, E. Chang , "*Support vector machine active leaning for image retrieval,''Proc. ACM Int. Conf. Multimedia*, pp. 107–118, Ottawa Canada, 2001.

[14] Z. Xu, X. Xu, K. Yu, V. Tresp, "*A Hybrid Relevance-Feedback Approach to Text Retrieval*", LECTURE NOTES IN COMPUTER SCIENCE, ISSU 2633, pages 281-293, 2003.

---

[i] Query

[ii] Vapnik

[iii] Pseudo-Labeling Fuzzy SVM

[iv] Subtractive

[v] http://people.csail.mit.edu/jrennie/20Newsgroups/20news-bydate.tar.gz

[vi] www.icmc.usp.br/~edsontm/pretext/pretext.html

[vii] Principal Component Analysis