

مدل ترکیبی فازی-عصبی برای کاربرد کاوی وب

چکیده

کاربرد کاوی وب شامل سه مرحله اصلی است: قبل از پردازش، کشف دانش و تجزیه و تحلیل الگو. اطلاعات به دست آمده از تجزیه و تحلیل می‌تواند توسط مدیران وبسایت برای مدیریت کارآمد و شخصی از وبسایت خود استفاده شود و در نتیجه نیازهای خاص جوامع خاصی از کاربران می‌تواند برآورده شود و سود افزایش یابد. همچنین، کاربرد کاوی وب تشخیص الگوهای پنهان نهفته در ورود داده‌های وب است. این الگوها نشان‌دهنده‌ی رفتار کاربر در حال جستجو است که می‌تواند برای تشخیص انحراف در رفتار کاربر در بانکداری مبتنی بر وب و برنامه‌های کاربردی دیگر که در آن حریم خصوصی داده‌ها و امنیت آن‌ها اهمیت زیادی برخوردار است به کار رود. روش پیشنهادی داده‌های وبسایت پلی تکنیک دکتر T.M.A.PAI را پردازش، کشف و تجزیه و تحلیل می‌کند. مدل ترکیبی براساس عصبی-فازی برای کشف دانش از لاگ وب استفاده شده است.

کلمات کلیدی: خروجی شبکه عصبی؛ خوشه‌بندی فازی C میانگین؛ شناسایی کاربر. شناسایی جلسه کاربر؛ وب داده‌ها.

1. معرفی

افزایش تعداد برنامه‌های کاربردی مبتنی بر وب که در مجموعه‌ای از مقدار عظیمی از اطلاعات در وب سرور منجر به پایگاه داده‌های مربوط می‌شود. این موضوع در کشف دانش با استفاده از تکنیک‌های وب کاوی استفاده شده است. در

کاربرد کاوی وب الگوهای جستجو کاربران برای استخراج اطلاعات مفید تجزیه و تحلیل می‌شود. جوامع کسب و کار از دانش کشف برای افزایش سود توسط وبسایت‌های شخصی برای مشتری استفاده می‌کند. با هدف درک رفتار کاربران و تنظیمات و در نتیجه افزایش سود برنامه‌های کاربردی مبتنی بر وب، کاربرد کاوی وب الگوهای استفاده شده‌ی پنهان و جالب در داده‌های وب را کشف می‌کند. معمولاً، کاربرد کاوی وب شامل پیش‌پردازش، کشف دانش و تجزیه و تحلیل الگو است.

لاگ‌های مربوط به سرور به عنوان ورودی برای فرآیند وب کاوی استفاده می‌شود. داده‌های لاگ بدون ساختار و مبهم هستند. به منظور استخراج الگوهای مفید از لاگ وب، باید برای رفع نویزهای آن از پیش پردازش شوند، در نتیجه داده‌های با حجم کاهش یافته پردازش شود. الگوریتم‌های داده کاوی می‌تواند برای لاگ‌های وب از پیش پردازش شده و برای استخراج الگوهای مفید به کار برده شود.

ویژگی نامشخص و مبهم رفتار کاربران در حال جستجو برای مدل‌سازی بسیار دشوار است. به طور کلی، یک وبسایت به بسیاری از گروه‌های مختلف کاربران حمله می‌کند. به عنوان مثال، یک گروه بازدیدکننده وبسایت ممکن است شامل دانش‌آموزان آینده‌نگر، والدین، بانکداران، پیمانکاران داخلی، ناشران کتاب و صاحبان فروشگاه کتاب و غیره همراه با عناصر ضد اجتماعی دیگر باشد. هر گروه می‌تواند برخی از نیازها یا اهداف خاصی داشته باشد. علاوه بر این، کاربران یک گروه می‌توانند مقاصد مختلفی از بازدید وبسایت داشته باشند. به عنوان مثال، ممکن است بانکدار از وبسایت دانشگاه و دسترسی به لینک‌های مربوط به تاریخ ورود بازدید کند به طوری که کمپینی برای وام‌های آموزشی داشته باشد. بانکدار مشابهی ممکن است از وبسایت جهت طرح وام‌های ماشین و مسکن بازدید کند. الگوریتم‌های خوشه‌بندی سخت موفق به گرفتن چنین رفتارهای متداخل و یا منافع کاربران به عنوان یک الگوریتم انحصاری نیستند. از این رو، الگوریتم‌های خوشه‌بندی فازی برای کاربرد کاوی وب مناسب‌تر می‌باشند. از این رو، در این کار پیش از ای، الگوریتم خوشه‌بندی فازی C- میانگین برای خوشه‌بندی جلسات کاربر وب به کار گرفته شده است.

در این کار، یک مدل ترکیبی براساس خوشه‌بندی عصبی-فازی برای خوشه‌بندی موثر کاربران وب‌سایت پلی‌تکنیک براساس الگوهای جستجو مشابه پیاده‌سازی شده است. لاگ وب با استفاده از تکنیک‌های کاهش ابعاد و روش‌های ترکیب پیش پردازش شده است.

ادامه مقاله به شرح زیر سازماندهی شده است. بخش 2 کارهای مرتبط؛ بخش 3 ارائه روش پیشنهادی؛ بخش 4 راه‌اندازی تجربی و تجزیه و تحلیل نتیجه و در نهایت، بخش 5 نتیجه کار را بیان می‌کند.

2. کارهای مرتبط

A. Bhargav و همکارانش یک چارچوب برای کاربرد کاوی وب متشکل از پیش پردازش، کشف الگو و طبقه‌بندی کاربران پیشنهاد داده است. این چارچوب کاربران را براساس کشور، ورود به سایت و زمان دسترسی طبقه‌بندی می‌کند. M. A. Eltahir و همکارانش و Sanjar و همکارانش به بررسی و بحث در مورد استخراج اطلاعات از تاریخ ورود کاربر با استفاده از کاربرد کاوی وب پرداخته‌اند. بررسی دقیق جمع‌آوری داده‌ها و مرحله قبل از پردازش از کاربرد کاوی وب توسط Varnagar بحث شده است. Sudheer و همکارانش چند روش آماده‌سازی داده‌ها از جریان دسترسی برای شناسایی جلسات منحصر به فرد و کاربران منحصر به فرد پیشنهاد داده است. تجزیه و تحلیل رفتار یادگیرنده برای کمک به ارزیابی یادگیری و به منظور افزایش ساختار یک دوره است، تکنیک‌های داده کاوی آموزشی به کار برده شده است. Maheswar و همکارانش یک الگوریتم جدید برای پیش پردازش و خوشه‌بندی لاگ وب ارائه داده است.

تکنیک‌های داده کاوی و تجزیه و تحلیل براساس عبارات منظم در اطلاعات تولید شده توسط گزارش‌های ارائه شده‌ی دانشگاه HTTP سرور توسط Adamov ارائه شده است. Joshi و همکارانش عملکردهای مختلف الگوریتم‌های محاسبات نرم بر روی وب‌سایت آموزشی را اجرا و تجزیه و تحلیل کرد. CLIQUE الگوریتمی برای خوشه‌بندی جلسات وب برای شخصی‌سازی وب‌سایت توسط Santhisree است. Nadi و همکارانش یک مدل پویا براساس تکنیک‌های خوشه‌بندی فازی، قابل اجرا بر روی خط کاربران پیشنهاد داده است. روش خوشه‌بندی فازی، در این مطالعه، امکان گرفتن عدم قطعیت در میان رفتار کاربران وب را فراهم می‌کند. خوشه‌بندی فازی در رویکرد medoids توسط پیرپائلو

و همکاری‌ها برای طبقه‌بندی توالی مرتب (مسیرهای) نمایش الگوهای رفتار فردی در یک فضای واقعی یا مجازی و حوزه زمان ارائه شده است. انصاری و همکاری‌ها یک تابع عضویت فازی برای تخصیص وزن به جلسه‌ها براساس تعداد آدرس‌های در دسترس توسط جلسات به کار برده‌اند و سپس از الگوریتم خوشه‌بندی فازی C-میانگین برای کشف خوشه‌های پروفایل کاربر استفاده کرده است. مدل فازی عصبی برای خوشه‌بندی داده‌ها توسط فرهاد روحی معرفی شده است.

3. متودولوژی

3.1. روش جمع‌آوری داده‌های وب

هر درخواست کاربر به سرور در لاگ وب سرور ثبت خواهد شد. هر خط در لاگ نشان‌دهنده‌ی درخواست به سرور است و معمولاً در فرمت قابل‌تعمیم لاگ شامل موارد زیر است:

1. میزبان / آدرس IP از راه دور: شناسایی کسی که از وب‌سایت بازدید کرده بود.
2. تأیید هویت کاربر: نام کاربری و رمز عبور در صورتی که سرور نیاز به احراز هویت کاربران داشته باشد (به‌طور کلی " -")
3. تاریخ و زمان درخواست: برای تعیین مدت زمانی که بازدیدکننده در یک صفحه داده به سر برده است مورد استفاده قرار می‌گیرد.
4. درخواست HTTP: روش (GET، POST، HEAD، و غیره) که برای انتقال اطلاعات همراه با ذکر نام درخواست منابع (یک صفحه HTML، یک فایل تصویری و یا یک اسکریپت) و نسخه پروتکل (پروتکل HTTP مورد استفاده قرار گرفته) استفاده می‌شود.
5. وضعیت درخواست پاسخ: کد وضعیت HTTP به مشتری بازگردانده می‌شود (200، 404 و غیره).
6. اندازه صفحه: شامل محتوا یا طول سند منتقل شده است.

```
124.40.247.196 - - [09/Jan/2015:10:04:32 +0000] "GET /courses/automobile-  
engineering HTTP/1.1" 200 2116 "http://tmapaipolytechnic.com/" "Mozilla/5.0  
(Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)  
Chrome/41.0.2267.0 Safari/537.36"
```

شکل 1: نمونه‌ای از لاگ وب

1. ارجاع درست: لیست آدرس سایت قبلی بازدید شده توسط مشتری است که به صفحه فعلی لینک شده است (اگر این اطلاعات از دست رفته باشد).

2. عامل کاربر: اطلاعاتی در مورد مرورگر مشتری، مدل مرورگر و سیستم عامل مشتری ارائه می‌دهد (اگر این اطلاعات از دست رفته باشد).

لاگ ورود به سیستم از وب سرور وبسایت پلی‌تکنیک دکتر T.M. A از 31 دسامبر 2014 از ساعات 12:09:56 تا 15 11:18:07 ژانویه 2015، در مجموع در 15 روز جمع‌آوری شده است. نمونه رکورد لاگ وب در شکل 1 داده شده است.

لاگ ورود به سیستم بالا نشان می‌دهد که کاربر با IP آدرس 124.40.247.196 درخواست لینک اتومبیل - مهندسی تحت دوره‌های آموزشی در 9 ژانویه 2015 در 10:04:32 کرده است و او از لینک <http://tmapaipolytechnic.com> عبور کرده است. درخواست موفقیت‌آمیز بوده است و در مجموع 2116 بایت دانلود شده است. همچنین، مرورگر موزیلا (سازگار) 5.0 و سیستم‌عامل ویندوز 6.3NT استفاده می‌شود.

3.2. مرحله قبل از پردازش لاگ وب

ورود به سیستم وب شامل تمام تعاملات با وب سرور است. که شامل درخواست واقعی کاربر و درخواست با توجه به فایل‌های تصویری موجود در صفحه وب است. علاوه‌براین به درخواست کاربر، ورود به سیستم متشکل از درخواست خودکار آغاز شده توسط روبات و یا برنامه‌ها است. به‌منظور کشف الگوهای مفید از لاگ وب، ورود به سیستم وب باید برای شناسایی کاربران و جلسات کاربر و یا فعالیت‌های کاربران در وبسایت پیش پردازش شود. پیش پردازش لاگ

وب متشکل از تمیز کردن لاگ وب، شناسایی کاربر و شناسایی جلسه است. قبل از تمیز کردن لاگ وب، ویژگی‌های وب نیاز به استخراج و زمان دارند، یک عدد صحیح برای نمایش زمان سپری شده از تاریخ پایه برای محاسبات بیشتر استفاده می‌شود.

پیش پردازش ورود به وب از تکنیک‌های کاهش ابعاد مختلف و روش‌های ترکیب، به نحو احسن برای شناسایی کاربران و جلسات کاربر استفاده می‌کند.

3.2.1. تمیز کردن لاگ وب

یک الگوریتم کلی برای تمیز کردن لاگ وب در زیر ارائه شده است. الگوریتم، لاگ فایل وب را به‌عنوان ورودی (WebLogFile) دریافت می‌کند و یک لاگ فایل تمیز به‌عنوان خروجی (New-LogFile) ارائه می‌دهد، که عاری از ورودی‌های اضافی است. الگوریتم روال خواستار چک کردن این است که آیا درخواست فعلی یک درخواست تصویر، یک درخواست ربات، یک درخواست GET موفق با توابع بولی با خروجی TRUE و FALSE است.

مرحله 1. تکرار مراحل 2-3 تا پایان فایل (WebLogFile) قابل دسترسی نیست.

مرحله 2. رکورد بعدی از WebLogFile خوانده می‌شود.

مرحله 3. اگر Log-Record یک درخواست ربات و یک درخواست تصویر نیست و یک GET موفق است -Log-Record را در New-LogFile بنویس.

مرحله 4. توقف

3.2.2. شناسایی کاربر

شناسایی کاربر براساس آدرس IP، عامل کاربر و ارجاع درست اجرا شده برای شناسایی منحصر به فرد کاربران است. الگوریتم شناسایی کاربر است که در زیر بیان می‌شود لاگ فایل تمیز را به‌عنوان ورودی دریافت می‌کند و هر کاربر مجزا را تشخیص می‌دهد.

مرحله 1. تنظیم کاربر - تعداد ← 0.

مرحله 2. تکرار مراحل 3-8 در حالی که انتهای فایل (New-LogFile) قابل دسترس نیست.

مرحله 3. رکورد جاری از New-LogFile به Cur-Record خوانده شود.

مرحله 4. رکورد بعدی New-LogFile به Next-Record خوانده شود.

مرحله 5. فرض کنید IPcur و IPnext آدرس IP در Cur-Record و Next-Record باشند.

مرحله 6. فرض کنید UAcur و UAnext عامل کاربر در Cur-Record و Next-Record باشند.

مرحله 7. فرض کنید RUcur و RUnext URL ارجاع در Cur-Record و Next-Record باشند.

مرحله 8.

If IPcur <> IPnext [IP addresses are different]

Identify both entries as belonging to different user.

Increment the User-Count by 1.

Otherwise

If UAcur <> UAnext [Both browser and OS are unique]

Identify both entries as belonging to different user.

Increment the User-Count by 1.

Otherwise If RUcur<>"-"and RUnext = "-"

Identify both entries as belonging to different user.

Increment the User-Count by 1.

otherwise

Assume as same user.

مرحله 9. توقف

3.2.3 شناسایی جلسه

شناسایی جلسه با هدف شناسایی فعالیت‌های کاربر در یک وبسایت است. الگوریتم برای شناسایی جلسه براساس هر

دو روش اکتشافی و ناوبری در زیر ارائه شده است.

مرحله 1. تنظیم PageStayTime ← 10 دقیقه.

مرحله 2. فرض کنید $\{\}$ = SessionSet

مرحله 3. تنظیم $K \downarrow 0$ و $J \downarrow 0$.

مرحله 4. فرض کنید LJ ، $URIj$ ، TJ ، $RUjz$ و UJ نشان‌دهنده‌ی ورود به سیستم، URI ، زمان، ارجاع URL و کاربر باشند.

مرحله 5. برای هر کاربر منحصر به فرد UJ مراحل 6-8 را انجام می‌دهد.

مرحله 6. فرض کنید $1 \downarrow 0$.

مرحله 7. برای هر ورود LJ مرحله 8 را انجام می‌دهد.

مرحله 8.

If referring URL, $RUj = \text{“-“}$ and $(tj - tj-1) > \text{PageStayTime}$

$K \leftarrow K+1$

$Sk \leftarrow URIj$

$\text{SessionSet} \leftarrow \text{SessionSet} \cup Sk$

Otherwise

If RUj is Present in any Si of Uj

$Si \leftarrow URIj$

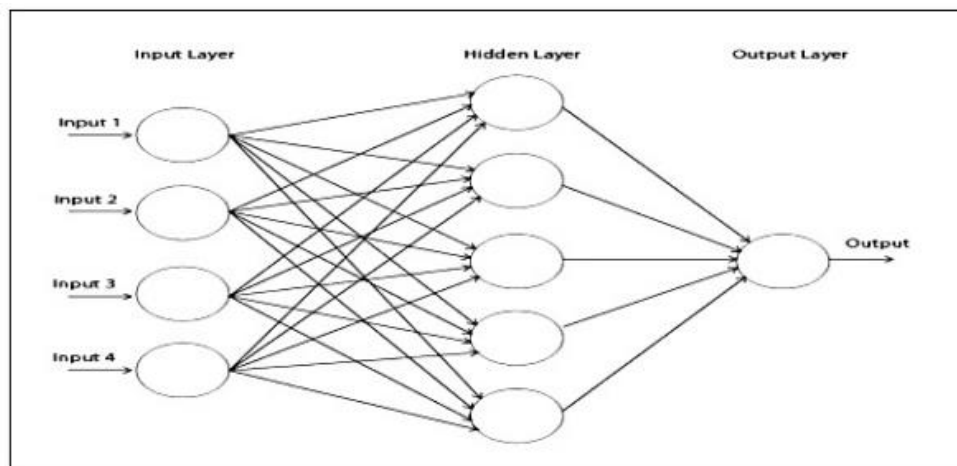
Else

$K \leftarrow K+1$

$Sk \leftarrow URIj$

$\text{SessionSet} = \text{SessionSet} \cup Sk$

مرحله 9. توقف



شکل 2: شبکه عصبی رو به جلو چند لایه

3.3. کشف دانش با استفاده از الگوریتم خوشه‌بندی فازی - عصبی

هنگامی که جلسات کاربر شناسایی شدند، یک فرایند خوشه‌بندی به منظور گروه‌بندی جلسات مشابه در خوشه یکسان به کار برده می‌شود. هر خوشه شامل کاربران با رفتار جستجوی مشترک و منافع مشابه است. در کار پیش رو، الگوریتم خوشه‌بندی فازی C- میانگین برای خوشه‌بندی جلسات کاربر استفاده شد.

شبکه عصبی مصنوعی در جستجوی معماری و نمایش اطلاعات الگوهای از انسان مغز است. شبکه عصبی مصنوعی برای انجام هر هدف انجام می‌شود. الگوها در ورودی ارائه شده‌اند، و با گره‌های خروجی با وزن‌های متفاوت همراه هستند. فرآیند تکرار شونده برای تنظیم وزن بین گره‌های ورودی و خروجی دنبال می‌شود. فرایند تنظیم وزن، به نام یادگیری، ارائه، یادگیری مستمر و یا قابلیت یادگیری مصنوعی به سیستم است، که می‌تواند یادگیری تحت نظارت و یا بدون نظارت در شبکه‌های عصبی مصنوعی باشد. یادگیری تحت نظارت خواستار تعریف کلاس خروجی برای هر یک از ورودی‌ها است.

در کار فعلی، یک مدل ترکیبی براساس شبکه‌های عصبی و خوشه‌بندی فازی برای خوشه‌بندی کاربران با توجه به الگوهای جستجو ارائه شده است. ابتدا، مجموعه نمونه‌ای از اطلاعات ورود به سیستم وب پیش پردازش با استفاده از الگوریتم خوشه‌بندی فازی C میانگین خوشه‌بندی می‌شود. سپس، ورودی الگوریتم خوشه‌بندی به‌عنوان ورودی به شبکه عصبی داده شده و خروجی الگوریتم خوشه‌بندی به‌عنوان خروجی هدف و آموزش شبکه عصبی انجام شده است. میانگین مربعات خطا - متوسط مربع خطاها بین خروجی شبکه و خروجی مورد نظر است و به‌عنوان اندازه‌گیری عملکرد استفاده می‌شود.

ما از شبکه عصبی رو به جلو چندلایه، آموزش دیده با الگوریتم آموزش در شکل 2 نشان داده شده است.

4. راه‌اندازی تجربی و تحلیل نتایج

لاگ ورود به سیستم از وب سرور وبسایت پلی‌تکنیک دکتر T.M. A از 31 دسامبر 2014 از ساعات 12:09:56 تا 15 11:18:07 ژانویه 2015، در مجموع در 15 روز جمع‌آوری شده است. در مجموع 5817 درخواست لاگ وب در این دوره ثبت شده است.

4.1. لاگ وب قبل از پردازش

الگوریتم تمیز کردن لاگ وب تمام سوابق بی‌ربط از لاگ وب را حذف می‌کند، تنها 1169 سابقه، حدود 20٪ از سوابق برای پردازش بیشتر حفظ می‌شود. آمار درخواست‌های بی‌ربط ثبت شده تحت دسته‌بندی‌های مختلف در جدول 1 نشان داده شده است.

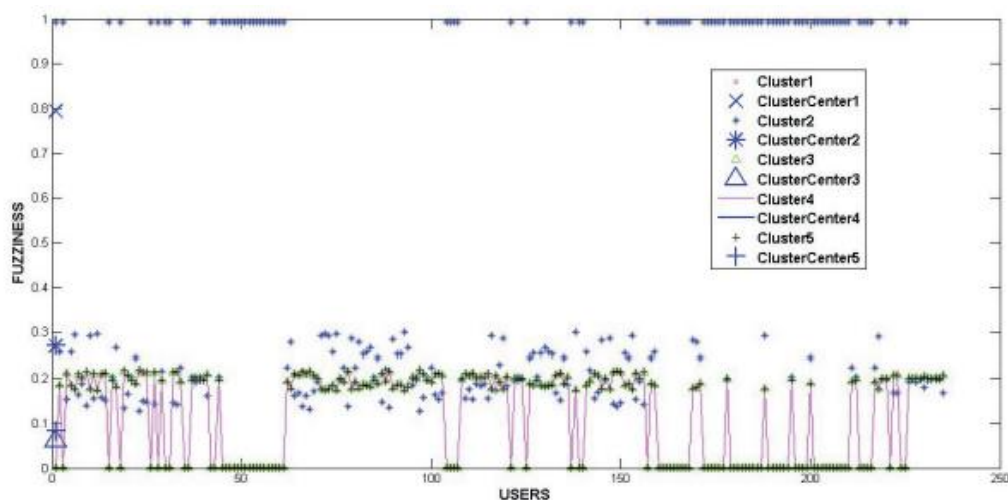
می‌توان مشاهده کرد که حدود 50٪ درخواست از لاگ وب را دانلود تصویر تشکیل می‌دهد، که توسط الگوریتم تمیز کردن لاگ وب حذف شده است. جدول 2 نتایج کل الگوریتم تمیز کردن از ورود به سیستم وب را نشان می‌دهد. برنامه‌کاربردی الگوریتم شناسایی کاربر در تمیز کردن لاگ وب در مجموع 235 کاربر منحصر به فرد را شناسایی کرده است. که توسط الگوریتم شناسایی جلسه انجام می‌شود. الگوریتم شناسایی یک جلسه برای هر کاربر با حداکثر تعداد صفحات در یک جلسه = 25 است.

جدول 1: آمار درخواست فردی (مربوط).

Request category	Number of records	Percentage
PNG	408	7
JPEG	2501	43
CSS	291	5
JS	582	10
NOT GET	58	< 1
NOT 200	1164	20
Robot	931	16

جدول 2: نتایج کل تمیز کردن لاگ وب

Statistics	Number of records
Original size	5817
Corrupt requests	-
Failed requests (other than 200)	1164
Multimedia objects	3782
Cleaned log size	1169
Percentage in reduction	80



شکل 3: مراکز نهایی خوشه فازی

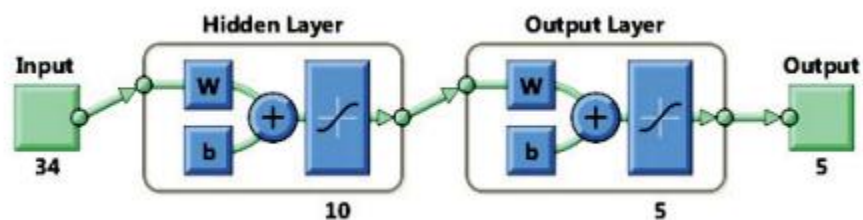
4.2. فاز کشف دانش

الگوریتم خوشه‌بندی فازی C-میانگین با استفاده از نرم‌افزار فازی MATLAB اجرا شده است. تابع MATLAB الگوریتم خوشه‌بندی فازی C-میانگین را انجام می‌دهد. تابع FCM یک مجموعه داده و تعداد مورد نظری از خوشه‌ها و مراکز خوشه مطلوب و امتیاز عضویت برای هر نقطه داده بر می‌گرداند. که با یک حدس اولیه برای مراکز خوشه شروع می‌شود، که بنا به محل متوسط هر خوشه در نظر گرفته شده است. حدس اولیه برای این مراکز خوشه به احتمال زیاد اشتباه است. سپس، FCM به هر نقطه یک درجه عضویتی برای هر خوشه اختصاص می‌دهد. با به‌روزرسانی مکرر مراکز خوشه و امتیازات عضویت برای هر نقطه داده، FCM مکرراً مراکز خوشه‌ها را به موقعیت مکانی راست در یک مجموعه داده حرکت می‌دهد. این تکرار بر مبنای به حداقل رساندن تابع هدفی است که نشان‌دهنده فاصله از هر نقطه از اطلاعات داده شده به یک مرکز خوشه توسط امتیاز عضویت آن نقطه است. خوشه‌بندی فازی C-میانگین یک فرایند تکراری

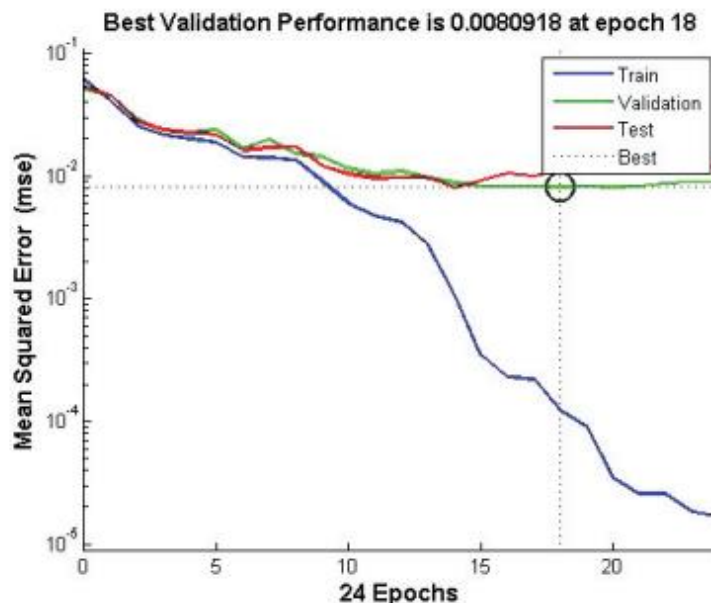
است. فرآیند زمانی متوقف می‌شود که به حداکثر تعداد تکرار رسیده باشیم و یا وقتی که بهبود عملکرد بین دو تکرار متوالی کمتر از حداقل میزان بهبودی مشخص شده باشد. ما با ارزش 2 برای تعداد خوشه‌ها، 100 برای تعداد تکرارها و $5E-1$ برای حداقل بهبود مراکز خوشه فازی نهایی آغاز خواهیم کرد.

نتایج نشان می‌دهد که جلسات به 5 خوشه تقسیم شدند. کاربر - جلسات در میان این خوشه‌ها توزیع شده است. اکثر کاربران در خوشه 5 باید درجه بالایی از عضویت داشته باشند.

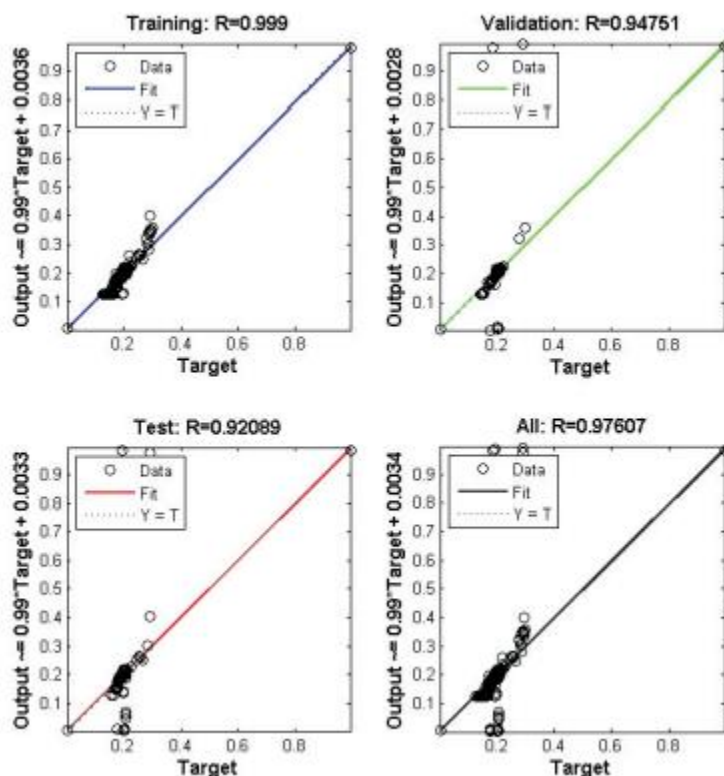
پس از این که الگوریتم خوشه‌بندی استفاده شد و خوشه‌ها تشکیل شدند، گام بعدی آموزش‌های عصبی خودکار شبکه برای خوشه‌بندی لاگ وب به خوشه شناخته شده است. برای این منظور، شبکه عصبی رو به جلو چندلایه با استفاده از نرم‌افزار شبکه عصبی MATLAB همانطور که در شکل 4 نشان داده شده است پیاده‌سازی می‌شود.



شکل 4: شبکه عصبی روبه جلو چندلایه با استفاده از نرم‌افزار شبکه عصبی MATLAB



شکل 5: کارایی شبکه عصبی



شکل 5: نمایش رگرسیون از شبکه

شبکه عصبی بهترین عملکرد اعتبارسنجی = 0.0080918 در دوران 18 همانطور که در شکل 5 نشان داده شده است به دست آورده است.

نمایش رگرسیون خروجی شبکه با توجه به اهداف آموزش، اعتبارسنجی، تست و تمام مجموعه در شکل 4 نشان داده شده است.

برای تناسب کامل، داده‌ها باید در امتداد یک خط 45 درجه، که در آن خروجی شبکه با اهداف زیر برابر هستند قرار می‌گیرند. برای این مشکل، مقدار R با 0.92 برای داده‌های آزمون و یک مقدار از 0.97 برای کل داده‌ها مناسب و معقول است. شبکه عصبی با دقت کاربران را به دسته‌های مختلف طبق الگوهای جستجوی خود تقسیم می‌کند.

5. نتیجه‌گیری

سیستم وب ورود اطلاعات، یک مخزن از الگوهای فعالیت کاربر در یک وبسایت است. کشف دانش از سیستم وب ورود اطلاعات نقش مهمی در ارائه خدمات بنا به نیازهای برنامه‌های کاربردی مبتنی بر وب دارد. در این کار، مدل ترکیبی براساس عصبی - فازی برای کشف الگوهای پنهان در لاگ وب وبسایت پلی‌تکنیک اجرا شده است. تکنیک‌های لاگ وب قبل از پردازش براساس تکنیک‌های کاهش ابعاد و ترکیب روش استفاده شده است. گام پیش پردازش تمام داده‌های بی‌ربط، با اندازه لاگ وب حاصل یعنی 20٪ از اندازه لاگ اصلی حذف می‌کند. مدل عصبی - فازی، شبکه‌های عصبی و مجموعه تئوری فازی را ترکیب می‌کند. خوشه‌بندی یک فرایند ذهنی است، که به معنی مجموعه یکسانی از اطلاعات است که بارها و بارها نیاز به تقسیم شدن برای کاربردهای مختلف دارد. این مسئله باعث می‌شود مشکل خوشه به عنوان یک الگوریتم یا رویکرد ناکافی برای حل تمام مشکلات خوشه‌بندی باشد. این مشکل از سیستم فازی - عصبی به عنوان یک سیستم یادگیری گرفته شده است و الگوها و قوانین خودکار تولید می‌کند. خوشه‌بندی عصبی - فازی، کاربران با الگوهای جستجوی مشابه را در یک خوشه گروه‌بندی می‌کند. همچنین، اطلاعات به دست آمده از تحلیل می‌تواند توسط مدیران وبسایت برای مدیریت کارآمد و شخصی وبسایت خود مورد استفاده قرار گیرد.

References

- [1] G. Neelima and Sireesha Rodda, An Overview on Web Usage Mining, Emerging ICT for Bridging the Future – Proceedings of the 49th Annual Convention of the Computer Society of India, vol. 2, Advances in Intelligent Systems and Computing, vol. 338, Springer International Publishing, pp. 647–655, (2015).
- [2] A. Bhargav and M. Bhargav, Pattern Discovery and Users Classification Through Web Usage Mining, Proceedings of the International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), IEEE, pp. 632–635, (2014).
- [3] M. A. Eltahir and A. F. A. Dafa-Alla, Extracting Knowledge from Web Server Logs using Web Usage Mining, Proceedings of the International Conference on Computing, Electrical and Electronics Engineering (ICCEEE), IEEE, pp. 413–417, (2013).
- [4] Sanjay Kumar Malik and SAM Rizvi, Information Extraction using Web Usage Mining, Web Scrapping and Semantic Annotation, Proceedings of the International Conference on Computational Intelligence and Communication Systems, IEEE, pp. 465–469, (2011).
- [5] C. R. Varnagar, N. N. Madhak, T. M. Kodinariya and J. N. Rathod, Web Usage Mining: A Review on Process, Methods and Techniques, Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES), IEEE, pp. 40–46, (2013).
- [6] K. Sudheer Reddy, M. Kantha Reddy and V. Sitaramulu, An Effective Data Preprocessing Method for Web Usage Mining, Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES), IEEE, pp. 7–10, (2013).
- [7] N. Sael, A. Marzak and H. Behja, Web Usage Mining Data Preprocessing and Multi Level Analysis on Moodle, Proceedings of the ACS International Conference on Computer Systems and Applications (AICCSA), IEEE, pp. 1–7, (2013).
- [8] B. U. Maheswari and P. Sumathi, A New Clustering and Preprocessing for Web Log Mining, Proceedings of the World Congress on Computing and Communication Technologies (WCCCT), IEEE, pp. 25–29, (2014).
- [9] A. Adamov, Data Mining and Analysis in Depth, Case Study of qafqaz University http Server Log Analysis, Proceedings of the 8th International Conference on Application of Information and Communication Technologies (AICT), IEEE, pp. 1–4, (2014).
- [10] M. Joshi, P. Lingras, Yiyu Yao and C. B. Virendrakumar, Rough, fuzzy, Interval Clustering for Web Usage Mining, Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE, pp. 397–402, (2010).
- [11] K. Santhisree and A. Damodaram, CLIQUE: Clustering Based on Density on Web Usage Data: Experiments and Test Results, Proceedings of the 3rd International Conference on Electronics Computer Technology (ICECT), IEEE, vol. 4, pp. 233–236, (2011).
- [12] S. Nadi, M. Saraee and M. Davarpanah-Jazi, A Fuzzy Recommender System for Dynamic Prediction of User's Behaviour, Proceedings of the International Conference for Internet Technology and Secured Transactions (ICITST), IEEE, pp. 1–5, (2010).
- [13] Pierpaolo D'Urso and Riccardo Massari, Fuzzy Clustering of Human Activity Patterns, Fuzzy Sets and Systems, vol. 215, Science Direct, pp. 29–54, (2013).
- [14] Z. Ansari, A. V. Babuy, W. Ahmed and M. F. Azeemz, A Fuzzy Set Theoretic Approach to Discover User Sessions from Web Navigational Data, Recent Advances in Intelligent Computational Systems (RAICS), IEEE, pp. 879–884, (2011).
- [15] Farhat Roohi, Neuro Fuzzy Approach to Data Clustering: A Framework for Analysis, European Scientific Journal March 2013 Edition, vol. 9, no. 9, pp. 183–192, (2013).
- [16] Dr. T. M. A. Pai Polytechnic Web Site URL <http://www.tmapaipolytechnic.com/>