فصلنامه اقتصاد و تجارت نوین، شماره‌های ۱۷ و ۱۸، تابستان و پاییز ۱۳۸۸، صفحات ۱۹۳– ۲۱۴

شناسایی مستمر الگوهای مصرفی کاربران اینترنت با استفاده از یک مدل داده‌کاوی تکاملی

سید علی‌اصغر مصطفوی ثابت*
کارشناسی ارشد مهندسی صنایع دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران
محمدرضا صالحی راد**
استادیار آمار دانشگاه علامه طباطبایی، تهران
علی بیات***
کارشناسی ارشد مهندسی صنایع دانشگاه تربیت مدرس، تهران
هدا فیوض****
کارشناسی برق و الکترونیک دانشگاه آزاد اسلامی، بوشهر
تاریخ دریافت: ۱۳۸۹/۰۲/۰۱ تاریخ پذیرش: ۱۳۹۰/۰۶/۱۰

چکیده

شرکت‌های خدمات دهنده اینترنت به‌طور مداوم با حجم بسیار زیادی از داده‌های خام مربوط به‌نحوه استفاده کاربران روبرو هستند. از آنجایی که حجم این اطلاعات، بسیار زیاد می‌باشد بنابراین راهبردهای داده‌کاوی می‌توانند برای شناسایی الگوهای مصرفی پنهان کاربران در جهت اتخاذ راهبردهای مدیریتی مناسب بسیار مفید باشند. اما الگوهای مصرفی کاربران اینترنت در طول سال به‌دلیل شرایط مختلف محیطی تغییر می‌کند. از این‌رو در این مقاله، در چارچوب از یک مدل داده‌کاوی تکاملی، الگوهای مصرفی کاربران اینترنت، به‌طور پیوسته و در طول زمان با به‌کارگیری شاخص‌های مصرفی کاربران و با استفاده از تکنیک داده‌کاوی قواعد ارتباط شبکه‌ای، شناسایی می‌شود. جهت تشریح این مدل، داده‌های خام یک شرکت خدمات دهنده اینترنت مورد بررسی قرار گرفته است. نتایج بررسی نشان می‌دهد که این مدل داده‌کاوی با شناسایی الگوهای مصرفی کاربران اینترنت، می‌تواند سبب بهبود خدمات شرکت‌های خدمات دهنده اینترنت و کاهش هزینه‌های آنها شود.

* پست الکترونیکی: mostafavi.ie@gmail.com
** پست الکترونیکی: salehi_rad@hotmail.com
*** مسئول مکاتبات، پست الکترونیکی: bayat044@yahoo.com
**** پست الکترونیکی: foyouz_hoda@yahoo.com

Archive of SID
www.SID.ir

واژه‌های کلیدی: الگوی مصرفی اینترنت، قواعد ارتباط شبکه‌ای، مدل داده‌کاوی تکاملی

طبقه‌بندی JEL : C14 ,D12

۱. مقدمه

موج خصوصی‌سازی جهانی صنعت ارتباطات در اوایل دهه ۱۹۸۰ میلادی آغاز گردید.[1] این امر موجب شد تا شرکت‌های خصوصی زیادی جهت ارائه‌ی خدمات مخابرات و ارتباطات ایجاد شوند و برای جذب مشتری به رقابت با یکدیگر بپردازند. در کنار خصوصی شدن طیف وسیعی از خدمات ارتباطات، خصوصی شدن خدمات دسترسی به اینترنت هم آغاز شد و شرکت‌های ارائه‌دهنده خدمات اینترنت[2] (ISP) ایجاد شدند. در ایران نیز در اواسط دهه ۱۳۷۰ مجوز ایجاد این شرکت‌ها صادر شد و به‌تدریج شرکت‌های خصوصی ایجاد گردیدند. سال به سال افراد بیشتری متقاضی دریافت این خدمات شدند و به موازات آن، سرمایه‌داران بیشتری برای ایجاد ISP و کسب درآمد از بازار جدید و بزرگ ایجاد شده، رغبت پیدا کردند.

طی سال‌های اول که رشد متقاضیان اینترنت بسیار زیاد بود، راهبرد اصلی مدیران شرکت‌های ISP برای جذب بیشتر مشتریان، ارائه خدمات اینترنت با قیمت‌های کمتر نسبت به سایر رقبا بود.[3] این راهبرد در سال‌های اولیه ورود اینترنت به خانه‌های مردم بسیار کارساز بود. اما در حال حاضر با گذشت زمان و حرفه‌ای‌تر شدن نوع استفاده‌ی افراد از اینترنت، راهبرد قیمت تقریباً کارایی خود را از دست داده و خواسته‌های مردم به سمت کیفیت ارائه خدمات اینترنتی سوق پیدا کرده است. اکنون کاربران اینترنت به‌خوبی درک کرده‌اند که کیفیت این خدمات، بسیار مهم‌تر از تفاوت ناچیز در قیمت می‌باشد.[4] در سال ۲۰۰۶ میلادی طی یک بررسی درخصوص رضایت مشتریان، نشان داده شد که از نظر ۷۸٪ مشتریان، قابلیت اطمینان مهم‌ترین عامل در ارائه خدمات اینترنت می‌باشد.[5] بنابراین چگونگی توسعه یک راهبرد مناسب جهت افزایش قابلیت اطمینان خدمات، درک بهتر نیازهای مشتریان و برآورده ساختن آنها در راستای حفظ مشتریان موجود، به‌عنوان موضوعات بسیار مهم برای شرکت‌های ISP مطرح شده است. برخی نیازها و ترجیحات کاربران اینترنت را می‌توان از طریق الگوهای مصرفی آنها یافت. در بازار رقابتی موجود، هر شرکتی که به نوع رفتار مشتریانش آگاه‌تر باشد از برتری بیشتری

[1] Li et al. (2008)
[2] Internet Service Provider (ISP)
[3] Madden (1999)
[4] Robert and Cole (1997)
[5] Mccue (2006)

www.SID.ir

نسبت به سایر رقبا برخوردار است.

مدیران شرکت‌های خدمات دهنده اینترنت به‌طور مداوم با حجم بسیار زیادی از اطلاعات مربوط به نحوه استفاده کاربران روبرو هستند. این اطلاعات بر روی رایانه‌های مرکزی[1] آنها ذخیره می‌شود. داده‌هایی که از رایانه مرکزی یک ISP به‌دست می‌آیند شامل مواردی همچون شماره تلفن مشتری، زمان اتصال به شبکه، مدت زمان اتصال به شبکه، حجم اطلاعات فرستاده شده، حجم اطلاعات دریافت شده و غیره می‌باشد. از آنجایی که این حجم اطلاعات، بسیار زیاد است بنابراین راهبردهای داده‌کاوی[2] می‌توانند در جهت شناسایی الگوهای مصرفی پنهان کاربران برای اتخاذ راهبردهای مدیریتی مناسب، بسیار مفید باشند.

الگوهای مصرفی کاربران اینترنت در طول سال به‌دلیل شرایط مختلف محیطی از جمله تعطیلات نوروز، تعطیلات تابستان، ایام امتحانات دانش‌آموزان و دانشجویان و غیره در حال تغییر است. بنابراین نمی‌توان این الگوها را تنها بر اساس داده‌های یک دورهٔ زمانی به دست آورد و آن را برای تمام طول سال استفاده کرد. بلکه در هر زمان، باید الگوها را بر اساس داده‌های گذشته‌ی نزدیک به‌دست آورد تا نتایج به‌دست آمده برای زمان حال معتبرتر باشند. در این مقاله، با استفاده از یک مدل داده‌کاوی تکاملی، الگوهای مصرفی کاربران اینترنت را به‌طور پیوسته و در طول زمان با به‌کارگیری شاخص‌های مصرفی مشتریان و با استفاده از راهبرد داده‌کاوی قواعد ارتباط شبکه‌ای،[3] به‌دست می‌آوریم. از این طریق مدیران ISP می‌توانند همواره بر اساس الگوهای به‌دست آمده (به‌ویژه الگوهای ساعات اوج مصرف کاربران)، نسبت به مدیریت بهتر نیروهای متخصص شرکت و افزایش کارایی آنها، جهت بهبود ارائه خدمات و کاهش هزینه‌ها گام بردارند.

ساختار این مقاله به این صورت می‌باشد: در بخش دوم، ادبیات تحقیق در زمینهٔ شناسایی الگوهای رفتاری کاربران اینترنت مرور شده است. در بخش سوم، راهبرد داده‌کاوی قواعد ارتباط شبکه‌ای مورد استفاده در این مقاله توضیح داده می‌شود. در بخش چهارم، یک مدل داده‌کاوی تکاملی جهت شناسایی مستمر الگوهای مصرفی کاربران اینترنت، ارائه می‌شود. در بخش پنجم، این مدل برای داده‌های واقعی یک ISP اعمال گردیده و در نهایت در بخش ششم نتیجه‌گیری تحقیق ارائه می‌شود.

---

[1] Server Computer
[2] Data Mining
[3] Association Rules

www.SID.ir

۲. ادبیات تحقیق

در ارتباط با مبحث مدیریت ترک مشتری که یکی از زمینه‌های مدیریت ارتباط مشتری می‌باشد تحقیقات زیادی انجام شده است که بسیاری از کاربردهای آنها برای مدیریت مشتریان مراکز خدمات ارتباطات مطرح شده است. پیترسون[1] (۲۰۰۴) مدلی را پیشنهاد کرده است که در آن از روش‌های کنترل کیفیت آماری جهت افزایش دقت در کشف نرخ ترک مشتریان استفاده شده است. جیانگ[2] (۲۰۰۷) مدل پیترسون را گسترش داد و از ابزارهای آمار بیزی و نمودارهای کنترلی EWMA و همچنین نمودارهای رفتاری مشتریان، جهت کنترل دقیق‌تر نرخ ترک مشتریان استفاده کرد.

در حوزه تحلیل اطلاعات مصرفی کاربران اینترنت جهت شناسایی الگوهای رفتاری آنها، کارهای کمی نیز انجام شده است. مَدن[3] (۱۹۹۹) مدلی را پیشنهاد کرد که در آن، احتمال ترک کاربران اینترنت از یک ISP را به ویژگی‌های مختلف خدمات ISP و همچنین ویژگی‌های جمعیت‌شناختی مشتریان مرتبط می‌ساخت. نتایج مدل او نشان داد که احتمال ترک یک کاربر از یک ISP، رابطه مستقیمی با هزینه‌های ماهانه ISP و رابطه معکوسی با میزان درآمد کاربر دارد. همچنین در این تحقیق نشان داده شد که قیمت خدمات اینترنت یک ISP به‌عنوان یک معیار مهم برای کاربران جهت انتخاب آن ISP مطرح می‌باشد. لی شنگ تون[4] (۲۰۰۸) تحقیقاتی را برای شناسایی الگوهای مصرفی کاربران اینترنت ADSL در تایوان انجام داد. برای انجام این تحقیق، از مدل استاندارد داده‌کاوی CRISP-DM استفاده شده است. هدف او از این تحقیق، شناسایی الگوهای رفتاری کاربران بود تا بر اساس آن الگوها، نسبت به پیشنهاد خدمات متناسب با الگوی رفتاری کاربران، جهت افزایش رضایت آنان اقدام شود. او ابتدا داده‌های مربوط به مصرف اینترنت را از یک ISP جمع‌آوری کرد. بعد از پیش پردازش آنها، با استفاده از مدل شبکه عصبی SOM، الگوهای مصرف هفتگی کاربران را به ۹ خوشه دسته‌بندی نمود. سپس بر اساس خوشه‌های به‌دست آمده، پیشنهاداتی را به مدیریت ISP جهت بهبود راهبردها ارائه کرد.

۳. قواعد ارتباط شبکه‌ای

علم داده‌کاوی، یکی از حوزه‌های تحقیقاتی جذاب و پر کاربردی است که طی دو دهه گذشته بسیار توسعه پیدا کرده است. داده‌کاوی فرآیندی است که به جستجوی الگوهای معنادار در

---

[1] Pettersson
[2] Jiang
[3] Madden
[4] Li Sheng-Tun

www.SID.ir

پایگاه داده‌های بزرگ می‌پردازد. در میان راهبردهای مختلف داده‌کاوی، قواعد ارتباط شبکه‌ای به‌دلیل کاربرد وسیع در بازاریابی و همچنین تحلیل سبد خرید مشتریان بسیار مورد توجه قرار گرفته است.[۱] علاوه بر مباحث بازاریابی، محققان موارد بسیار زیادی از کاربردهای قواعد ارتباط شبکه‌ای در حوزه‌های مختلف علمی یافته‌اند که در این میان می‌توان به کاربرد این راهبرد در حوزه‌های پزشکی، اقتصادی، و همچنین مدیریت تولید و بهبود کیفیت تولیدات اشاره نمود.

۳-۱. الگوریتم شناسایی قواعد ارتباط شبکه‌ای استاندارد

قواعد ارتباط شبکه‌ای اولین بار توسط اگراوال[۲] (۱۹۹۳) معرفی شد. پس از آن بسیاری از محققان این الگوریتم را توسعه دادند. یک قاعده ارتباط شبکه‌ای به‌صورت یک عبارت $X \rightarrow Y$ بیان می‌شود که در آن X مجموعه‌ای از اقلام[۳] و Y یک قلم می‌باشد؛ و به این معناست که اگر مجموعه اقلام X رخ دهد آنگاه قلم Y با احتمال مشخصی رخ خواهد داد. وظیفه اصلی قواعد ارتباط شبکه‌ای، شناسایی الگوهای پنهان میان اقلام موجود در پایگاه داده‌های بزرگ می‌باشد. در الگوریتم قواعد ارتباط شبکه‌ای استاندارد جهت یافتن قواعد جذاب و کارا، برای هر قاعده دو مقدار عددی پشتیبانی[۴] و اطمینان[۵] به‌صورت زیر محاسبه می‌شود:

پشتیبانی (sup): نسبت تعداد تراکنش‌ها[۶] با همه اقلام موجود در قاعده به تعداد کل تراکنش‌ها

اطمینان (conf): نسبت تعداد تراکنش‌ها با همه اقلام موجود در قاعده به تعداد تراکنش‌ها با اقلام موجود در جمله شرطی

در واقع، مقدار پشتیبانی یک قاعده نمایانگر میزان فراوانی، و مقدار اطمینان نمایانگر قدرت آن می‌باشد. جهت انتخاب قواعد با میزان فراوانی و قدرت مطلوب، دو مقدار عددی حداقل پشتیبانی (minsup) و حداقل اطمینان (minconf) به‌عنوان مقادیر آستانه توسط تحلیل‌گر تعیین می‌گردد. اگر برای یک قاعده، دو شرط $Sup \geq minsup$ و $Conf \geq minconf$ برقرار باشد، آن قاعده به‌عنوان یک قاعده ارتباط شبکه‌ای مطلوب انتخاب می‌شود، در غیر این صورت رد می‌گردد.

---

[1] Bodon (2003)
[2] Agrawal
[3] Itemset
[4] Support
[5] Confidence
[6] Transactions

www.SID.ir

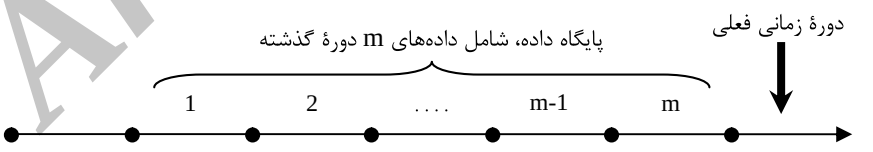
۳-۲. شناسایی قواعد ارتباط شبکه‌ای از داده‌های کمّی وزن‌دار در چارچوب مدل داده‌کاوی تکاملی

الگوریتم[1] WUARM یک الگوریتم بسیار کارا جهت شناسایی قواعد ارتباط شبکه‌ای از داده‌های کمّی وزن‌دار می‌باشد که در سال ۲۰۰۸ توسط سلیمان خان و همکاران پیشنهاد شد.[2] این الگوریتم در سال ۲۰۰۹ توسط مصطفوی و همکاران[3] در چارچوب یک مدل داده‌کاوی تکاملی توسعه یافت. توضیحات مربوط به علائم و گام‌های این الگوریتم به‌صورت زیر می‌باشد:

علائم

p تعداد اقلام؛ m تعداد دوره زمانی گذشته می‌باشد که در ایجاد قواعد ارتباط شبکه‌ای در هر دوره زمانی استفاده می‌شوند؛ n_i تعداد تراکنش‌ها در دوره زمانی iاُم، $1 \le i \le m$؛ t_{ij} تراکنش jاُم در دوره زمانی iاُم، $1 \le j \le n_i$؛ $T_i = \{t_{i1}, t_{i2}, t_{i3}, \dots, t_{in_i}\}$ مجموعهٔ تراکنش‌های دوره زمانی iاُم؛ i_k قلم k اُم، $1 \le k \le$ ؛ $I = \{i_1, i_2, i_3, \dots, i_p\}$ مجموعه‌ی اقلام؛ $|t_{ij}|$ تعداد اقلام مربوط به تراکنش j اُم در دوره‌زمانی iاُم که مقدارشان بزرگتر از صفر می‌باشد؛ w_k وزن مربوط به قلم kاُم، $0 \le w_k \le 1$؛ v_i وزن مربوط به دوره زمانی i اُم، $0 \le v_i \le 1$؛ q_{ijk} تعداد قلم kاُم در تراکنش jاُم در دوره زمانی iاُم، شامل اعداد صحیح غیرمنفی؛ $wu(q_{ijk})$ مقدار وزن‌دار قلم kاُم در تراکنش jاُم در دوره زمانی iاُم؛ $mwu(t_{ij})$ میانگین مقادیر وزن‌دار اقلام مربوط به تراکنش jاُم در دوره زمانی iاُم؛ Sup مقدار پشتیبانی محاسبه شده برای هر مجموعه اقلام؛ Conf مقدار اطمینان محاسبه شده برای هر قاعده می‌باشد. جهت روشن‌تر شدن توضیحات مربوط به علائم، شکل ۱ و جدول ۱ ارائه شده است.

شکل ۱. دوره‌های زمانی مورد استفاده جهت شناسایی قواعد ارتباط شبکه‌ای



[1] Weighted Utility Association Rule Mining
[2] Sulaiman Khan et al. (2008)
[3] Mostafavi and Hosein Sabbari (2009)

Archive of SID
www.SID.ir

جدول ۱. جدول مربوط به داده‌های یک دورهٔ زمانی

	اقلام →	i_1	i_2	...	i_p
	(وزن) →	(w_1)	(w_2)		(w_p)
تراکنش‌های دورهٔ زمانی i (v_i)	t_{i1}	q_{i11}	q_{i12}	...	q_{i1p}
	t_{i2}	q_{i21}	q_{i22}	...	q_{i2p}
	t_{i3}	q_{i31}	q_{i32}	...	q_{i3p}
	⋮	⋮	⋮	⋮	⋮
	t_{in_i}	q_{in_i1}	q_{in_i2}	...	q_{in_ip}

الگوریتم

در هر دورهٔ زمانی، مراحل زیر تکرار می‌شود:

ورودی: m، v_i، w_k، **minsup**، **minconf**، پایگاه داده

خروجی: لیستی از قواعد ارتباط شبکه‌ای معتبر برای دوره زمانی فعلی

گام ۱: مقادیر کمی q_{ijk} را با استفاده از معادله ۱، به مقادیر وزن‌دار $wu(q_{ijk})$ تبدیل کنید:

$$wu(q_{ijk}) = v_i w_k q_{ijk} \quad , \quad 1 \le l \le m \, , \quad 1 \le j \le n_i \, , \quad 1 \le k \le p \qquad (۱)$$

گام ۲: میانگین مقادیر وزن‌دار اقلام را برای هر تراکنش در هر دوره با استفاده از معادله ۲ محاسبه کنید:

$$mwu(t_{ij}) = \frac{\sum_{k=1}^{p} wu(q_{ijk})}{|t_{ij}|} \quad , \quad 1 \le i \le m \, , \quad 1 \le j \le n_i \qquad (۲)$$

گام ۳: برای هر مجموعه اقلام (X,Y)، مقدار پشتیبانی را از رابطه ۳ به‌دست آورید:

$$Sup(X,Y) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{|s_i|} mwu(t_{ij})}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} mwu(t_{ij})} \qquad (۳)$$

به‌طوری‌که

$$S_i = \{S_i | S_i \subseteq T_i \, , X \cup Y \in S_i\}$$

گام ۴: مجموعه‌های اقلامی که محدودیت $Sup \ge minsup$ را برآورده می‌سازند، انتخاب کرده و به گام بعد بروید.

گام ۵: برای هر مجموعه اقلام انتخاب شده با اقلام $x_1, x_2, x_3, \dots, x_r$ که $r \ge 2$، تمام قواعد ممکن را به‌صورت رابطه ۴ به‌دست آورید:

www.SID.ir

$$x_1, x_2, \ldots, x_a, x_c, \ldots, x_r \rightarrow x_b \qquad (b = 1 \text{ to } r \ , a = b-1 \ , c = b+1) \tag{۴}$$

گام ۶: برای هر قاعده $X \rightarrow Y$ ، مقدار اطمینان را از معادله ۵ به‌دست آورید:

$$Conf(X \rightarrow Y) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{|S_i|} mwu(t_{ij})}{\sum_{i=1}^{m} \sum_{j=1}^{|Z_i|} mwu(t_{ij})} \tag{۵}$$

به‌طوری‌که

$$S_i = \{S_i | S_i \subseteq T_i \ , X \cup Y \in S_i\}$$

$$Z_i = \{Z_i | Z_i \subseteq T_i \ , X \in Z_i\}$$

گام ۷: در نهایت، قواعدی که محدودیت $Conf \geq minconf$ را برآورده می‌سازند، به‌عنوان قواعد نهایی انتخاب می‌شوند.

۴. مدل پیشنهادی جهت شناسایی الگوی مصرفی کاربران اینترنت

در این بخش یک مدل داده‌کاوی پویا جهت شناسایی پیوسته الگوهای رفتاری کاربران اینترنت با استفاده از راهبرد قواعد ارتباط شبکه‌ای معرفی می‌شود.

۴-۱. نوع خدمات اینترنت و انتخاب شاخص مناسب برای تحلیل رفتار کاربران

شرکت‌های ISP، خدمات متنوعی را جهت استفاده از اینترنت، به مشتریان خود عرضه می‌کنند. در ابتدای ورود این فناوری به خانه‌ها، دسترسی به اینترنت تنها به‌صورت Dial up عرضه می‌شد که همچنان نیز ادامه دارد. نوع دیگری از اینترنت به‌صورت ADSL عرضه می‌شود. در این نوع اینترنت، کاربر می‌تواند از اینترنت پرسرعت بدون مشغولی خط تلفن استفاده کند. با اینکه اینترنت ADSL به‌تدریج در حال باز کردن جای خود در بین خانواده‌ها است، همچنان سهم بزرگی از درآمد شرکت‌های ISP ناشی از عرضه خدمات اینترنت Dial up می‌باشد به‌همین دلیل در این تحقیق به بررسی الگوهای مصرفی کاربران اینترنت Dial up پرداخته شده است.

در سیستم اینترنت Dial up، مبنای قیمت بر اساس مدت زمانی است که کاربر از اینترنت استفاده می‌کند. در اینگونه خدمات اینترنت، فرقی نمی‌کند که کاربر چه حجمی از اطلاعات را رد و بدل کرده است بلکه تنها برای مدت زمان استفاده از اینترنت مبلغی را پرداخت می‌کند. بنابراین می‌توان نتیجه گرفت که مدت زمان استفاده از اینترنت Dial up با درآمد کسب شده در این مدت زمان، رابطه مستقیمی دارد. بدین معنا که هر چه مدت زمان استفاده‌ی مشتری‌ها از اینترنت بیشتر باشد، درآمد ISP هم بیشتر می‌باشد. بنابراین شاخص مدت زمان استفاده از اینترنت را می‌توان به‌عنوان شاخص مناسبی جهت تحلیل الگوی رفتاری کاربران در نظر گرفت.

Archive of SID

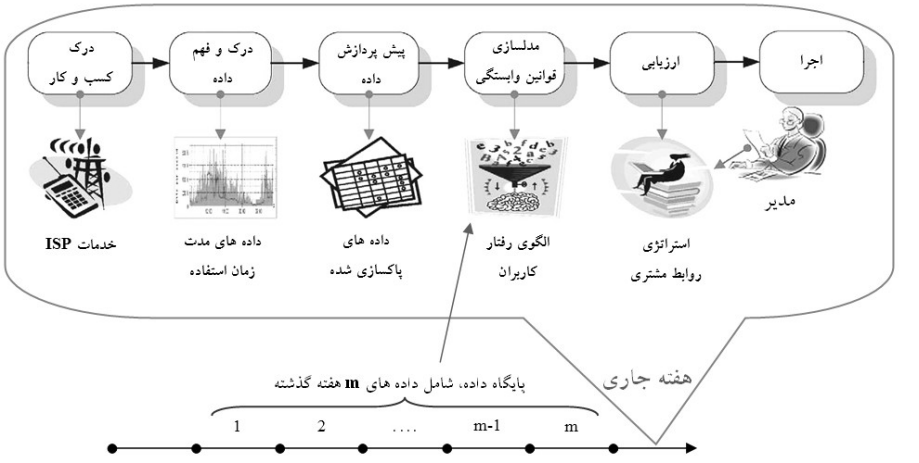
www.SID.ir

۴-۲. مدل داده‌کاوی تکاملی جهت شناسایی مستمر الگوی مصرفی کاربران اینترنت

مدل ارائه شده در این قسمت به‌گونه‌ای است که بخش فنی ISP در اول هر هفته با توجه به داده‌های خام مربوط به میزان استفاده‌ی کاربران از اینترنت در m هفته قبل و با استفاده از الگوریتم قواعد ارتباط شبکه‌ای، الگوی مصرفی کاربران اینترنت خود را به دست آورد. ضرورت شناسایی مستمر الگوها این است که رفتار مصرفی کاربران همیشه یکسان نیست، بلکه در روزهای مختلف سال و همچنین شرایط هر زمان، می‌تواند تغییر کند. علاوه بر این، مدیر ISP می‌تواند تأثیر برخی از خدمات جدید خود بر رفتار کاربران را از روی الگوهای به‌دست آمده مشاهده کند. همچنین می‌توانند با شناسایی بازه‌های زمانی که میزان استفاده کاربران بسیار کم می‌باشد، طرح‌های تشویقی (مثلاً کاهش قیمت در ساعات کم ترافیک) را به‌گونه‌ای ایجاد کنند که قسمتی از بار ترافیک ساعاتی که اوج مصرف کاربران می‌باشد به زمان‌های خلوت منتقل شود.

مدل ارائه شده شامل شش گام می‌باشد که به‌ترتیب عبارتند از: درک خدمات ISP، جمع‌آوری و درک داده‌های خام مربوط به استفاده کاربران از اینترنت، پیش پردازش داده‌های خام، استفاده از الگوریتم قواعد ارتباط شبکه‌ای جهت شناسایی الگوهای مفید، پیشنهاد راهبرد مناسب برای هفته‌ی جاری بر اساس الگوهای به‌دست آمده، و در نهایت اجرای راهبردهای پیشنهاد شده. شکل ۲ مدل داده‌کاوی شناسایی مستمر الگوی مصرف کاربران اینترنت را نشان می‌دهد.

شکل ۲. مدل پیشنهادی جهت شناسایی مستمر الگوی مصرفی کاربران اینترنت



www.SID.ir

در ابتدای هر هفته، این مراحل ششگانه اجرا می‌شوند. در ادامه در مورد هر مرحله توضیحاتی داده شده است:

- **درک کسب‌وکار:** برای شروع فرآیند داده‌کاوی، ابتدا باید اهداف و خواسته‌های مدیران شرکت ISP کاملاً مشخص شده و سپس هدف داده‌کاوی تعیین گردد. هدف این مقاله، شناسایی الگوهای مصرف اینترنت کاربران یک ISP در طول هفته می‌باشد تا با شناسایی ساعت‌های پرمصرف در روزهای مختلف هفته، نسبت به برنامه‌ریزی جهت ارائه خدمات بهتر در آن زمان‌ها اقدام شود. با توجه به این هدف مدیریتی، هدف از این داده‌کاوی عبارت است از:

  شناسایی قواعد ارتباط شبکه‌ای به فرم (ساعت ⟶ روز)

- **جمع‌آوری داده:** با توجه به هدف داده‌کاوی که در مرحلهٔ قبل بیان شد، در این مرحله داده‌های خام مربوط به اتصالات کاربران طی m هفته‌ی گذشته جمع‌آوری می‌شوند.
- **پیش پردازش داده:** در این مرحله، داده‌های مربوط به زمان ورود و زمان خروج در هر تراکنش باید به‌گونه‌ای پردازش شوند که برای استفاده در الگوریتم شناسایی مستمر قواعد ارتباط شبکه‌ای از داده‌های کمّی وزن‌دار به شکل مناسبی درآیند. برای این کار بایستی جدول جدیدی ایجاد شود که سطرهای آن شامل تراکنش‌ها و ستون‌های آن شامل ۷ ستون برای روزهای هفته و ۲۴ ستون برای بازه‌های یک ساعت در شبانه‌روز باشد. مقادیر ستون‌های روزهای هفته به‌صورت دودویی می‌باشد و مقادیر ستون‌های ساعت، میزان استفادهٔ کاربر بر حسب واحد دقیقه می‌باشد.
- **مدلسازی:** در این مرحله، الگوریتم قواعد ارتباط شبکه‌ای ارائه شده را بر روی داده‌های جدول آماده شده در مرحله قبل اعمال کرده و قوانین نهایی استخراج می‌شود. در این راستا بایستی به چند نکته توجه کرد:
  - تمام مقادیر وزن wk برابر با ۱ در نظر گرفته می‌شود.
  - وزن vi تنها در مقادیر ستون‌های بازه‌ی ساعتی ضرب می‌شود.
  - مقدار mwu هر تراکنش فقط با استفاده از مقادیر ستون‌های بازهٔ ساعتی محاسبه می‌شود و ستون‌های دودویی مربوط به روزها در این محاسبات دخیل نمی‌شود.
  - از مقادیر ستون‌های مربوط به روزهای هفته تنها جهت ایجاد قواعد استفاده می‌شوند.
  - مقادیر حداقل آستانه پشتیبانی و حداقل آستانه اطمینان بر اساس تجربه و نظر فرد متخصص تعیین می‌شوند. زمانی که نیاز باشد قوانین قوی‌تری استخراج شود، می‌توان مقادیر حداقل آستانه را بزرگ‌تر از حد معمول در نظر گرفت.

Archive of SID

www.SID.ir

۵. مطالعه موردی

در این بخش، فرآیند شناسایی مستمر الگوهای مصرفی کاربران اینترنت بر روی داده‌های واقعی شرکت ISP دام‌نتینگ که یکی از با سابقه‌ترین شرکت‌های ISP شهر بوشهر است اعمال گردیده است. این شرکت در ارائه خدمات پشتیبانی با مشکل مواجه بود و اکثر کاربران از ضعف پاسخ‌گویی مرکز تلفن پشتیبانی این شرکت گلایه می‌کردند. از این‌رو مدیریت شرکت قصد داشت کیفیت عرضه خدمات خود را با افزایش کارایی مرکز تلفن بالا ببرد تا با افزایش سطح رضایت مشتریان خود، نرخ ترک آنها را کاهش دهد. با توجه به اهداف مدیریت، هدف از این داده‌کاوی عبارت است از «شناسایی قواعد ارتباط شبکه‌ای به فرم (ساعت/روز)»، تا با شناسایی ساعت‌های پرمصرف در روزهای مختلف هفته، نسبت به برنامه‌ریزی جهت ارائه خدمات بهتر در آن زمان‌ها اقدام شود.

جهت شناسایی الگوی مصرفی کاربران برای کمک به مدیریت ISP در هفته جاری، از مدیر فنی شرکت درخواست شد که تمام اطلاعات متصل شوندگان[1] مربوط به ۴ هفتهٔ گذشته را در اختیار قرار دهد. در نهایت داده‌های مورد نظر به‌صورت فایل‌های اکسل از رایانه مرکزی ISP استخراج شد. داده‌های به‌دست آمده شامل ۴۱۸۵۱ تراکنش (سطر) مربوط به کل اتصال کاربران به شبکه از ساعت ۰۰:۰۰ روز شنبه ۸۸/۱۰/۱ تا ساعت ۲۳:۵۹ روز جمعه ۸۸/۱۰/۲۸ بود که هر تراکنش نمایانگر یک فرآیند اتصال به شبکه اینترنت و خروج از آن توسط هر کاربر می‌باشد. شکل ۳ نمونه‌ای از داده‌های خام به‌دست آمده از اتصال کاربران به شبکه را نشان می‌دهد.

شکل ۳. نمونه‌ای از اطلاعات به‌دست آمده از اتصال کاربران شبکه

	B	C	D	E	F	G	H	I	J	K
1	User ID	Username	Credit	Login Time	Logout Time	Duration	Successful	Caller ID	Bytes IN	Bytes OUT
2	119580	d106uiqus4	408.6	1386-10-01 00:00	1386-10-01 00:49	2944	t	77145xxx63	5474194	1017993
3	115843	d20fma5ywk	300.5	1386-10-01 00:00	1386-10-01 00:37	2165	t	77135xxx01	4207875	408445
4	116105	d20qciitp8	555.7	1386-10-01 00:01	1386-10-01 01:08	4003	t	async	7797275	5678978
5	92935	d20jdnh39e	387.7	1386-10-01 00:02	1386-10-01 00:48	2793	t	77155xxx72	4729692	361910
6	109486	d203nnstji	0	1386-10-01 00:02	1386-10-01 00:02	0	f	77145xxx62	0	0
7	114885	d10geams4s	1341	1386-10-01 00:02	1386-10-01 02:43	9658	t	77125xxx24	1.7E+07	3280176
8	119710	d10afpwa5j	309	1386-10-01 00:04	1386-10-01 00:41	2226	t	77125xxx64	8074515	989082
9	126858	d10eskjvvq	402.5	1386-10-01 00:05	1386-10-01 00:53	2899	t	7713?5xxx62	4692679	912742
10	115968	d20juq4de7	0	1386-10-01 00:06	1386-10-01 00:06	0	f	77125xxx46	0	0
11	120241	d10ufc2gwe	125.2	1386-10-01 00:07	1386-10-01 00:22	902	t	77145xxx95	141470	70489
12	115908	d20i25h3rt	565.9	1386-10-01 00:08	1386-10-01 01:16	4076	t	77135xxx17	6214638	1206699
13	96097	d07qubtdbi	0	1386-10-01 00:10	1386-10-01 01:45	5677	t	77135xxx66	4151850	311006

[1] Connection Log

www.SID.ir

حال باید داده‌های ستون‌های زمان ورود به شبکه[1] و زمان خروج از شبکه[2] به‌گونه‌ای پردازش شوند که برای استفاده در الگوریتم شناسایی مستمر قواعد ارتباط شبکه‌ای از داده‌های کمّی وزن‌دار به شکل مناسبی آماده شود. جدول ۲ نمونه‌ای شامل ۴ تراکنش از اتصال کاربران به شبکه را نشان می‌دهد. حال بایستی جدول جدیدی ایجاد شود که سطرهای آن شامل تراکنش‌ها و ستون‌های آن شامل ۷ ستون برای روزهای هفته و ۲۴ ستون برای بازه‌های یک ساعت در شبانه روز باشد (جدول ۳). حال باید برای هر تراکنش جدول اول، بر اساس زمان ورود و خروج کاربر، تراکنشی در جدول دوم ایجاد شود. برای انجام این کار، دو حالت رخ می‌دهد:

جدول ۲. نمونه‌ای از تراکنش‌های مربوط به اتصال کاربران به شبکه

تراکنش	زمان ورود	زمان خروج
۱	۲۲:۲۳ ۸۸-۱۰-۱۲	۲۲:۵۵ ۸۸-۱۰-۱۲
۲	۲۲:۲۵ ۸۸-۱۰-۱۲	۲۳:۴۰ ۸۸-۱۰-۱۲
۳	۲۲:۲۶ ۸۸-۱۰-۱۲	۲۳:۱۰ ۸۸-۱۰-۱۲
۴	۲۲:۲۶ ۸۸-۱۰-۱۲	۰۰:۳۱ ۸۸-۱۰-۱۳

جدول ۳. نمونه‌ای از جدول داده‌های خام پردازش شده

تراکنش	روزهای هفته				ساعت‌های شبانه‌روز				
	شنبه	یکشنبه	...	جمعه	۰ - ۱	۱ - ۲	...	۲۲ - ۲۳	۲۳ - ۲۴
۱	۱	۰	...	۰	۰	۰	...	۳۲	۰
۲	۱	۰	...	۰	۰	۰	...	۳۵	۴۰
۳	۱	۰	...	۰	۰	۰	...	۳۴	۱۰
۴	۱	۰	...	۰	۰	۰	...	۳۴	۶۰
۵	۰	۱	...	۰	۳۱	۰	...	۰	۰

۱. **زمان ورود و خروج آن تراکنش در یک روز باشد:** برای هر تراکنش با توجه به اینکه در چه روزی از هفته رخ داده است در خانهٔ مربوط به آن روز عدد یک و برای بقیه روزها عدد صفر قرار داده می‌شود. سپس با توجه به اینکه تراکنش در چه ساعت‌هایی رخ داده است و در هر کدام از ساعت‌ها، چند دقیقه طول کشیده است، میزان دقیقه مربوط به هر بازهٔ زمانی یک

[1] Login Time
[2] Logout Time

www.SID.ir

ساعته در خانه مربوط به آن نوشته می‌شود (تراکنش‌های ۱، ۲ و ۳ در جدول ۲).

۲. **زمان ورود و خروج آن تراکنش در دو روز باشد:** این حالت زمانی رخ می‌دهد که کاربر در یک روز به شبکه وصل شود و در روز بعد از شبکه خارج شود. در این حالت آن تراکنش بایستی به دو تراکنش مجزا تبدیل شود. تراکنش اول مربوط به روز اول است و بازه ساعتی اتصال آن به شبکه از زمان ورود به شبکه تا ساعت ۲۴ همان روز می‌باشد. تراکنش دوم مربوط به روز دوم بوده و بازه‌ی ساعتی اتصال آن به شبکه از ساعت ۰۰ : ۰۰ تا زمان خروج از شبکه می‌باشد. نمونه‌ای از این حالت، تراکنش ۴ در جدول ۳ می‌باشد که به دو تراکنش ۴ و ۵ در جدول ۳ تبدیل شده است.

این پردازش‌ها برای داده‌های خام مربوط به چهار هفته گذشته، با استفاده از برنامه‌نویسی در برنامه اکسل انجام پذیرفت (جدول ۴). حال بایستی برای به‌دست آوردن قواعد جذاب و مفیدی به‌صورت (ساعت/روز)، الگوریتم شناسایی مستمر قواعد ارتباط شبکه‌ای از داده‌های کمّی وزن‌دار، بر روی جدول تراکنش‌های پردازش شده اعمال شود.

برای اجرای الگوریتم ابتدا بایستی ارزش داده‌های هفته‌های گذشته و همچنین مقادیر حداقل آستانه پشتیبانی و حداقل آستانه اطمینان تعیین گردند. این مقادیر پس از مشورت متخصصان داده‌کاوی و ISP و بر اساس تجربیات گذشته در زمینه الگوی مصرفی کاربران، به این صورت تعیین شدند: ارزش داده‌های یک هفته گذشته ۱، دو هفته گذشته ۰/۶، سه هفته گذشته ۰/۳ و چهار هفته گذشته ۰/۲ در نظر گرفته شد. همچنین حداقل آستانه پشتیبانی برابر با ۰/۰۱۲ و حداقل آستانه اطمینان برابر با ۰/۰۹ در نظر گرفته شد. با توجه به مدل پیشنهادی، به تمام اوزان w_k مقدار ۱ تخصیص گرفت.

حال مقادیر کمی جدول داده‌های خام پردازش شده با استفاده از وزن هر تراکنش و با استفاده از رابطه ۱ به مقادیر کمی وزن‌دار تبدیل شده و سپس مقدار mwu ی هر تراکنش با استفاده از رابطه ۲ محاسبه می‌گردد (جدول ۵). توجه شود که این محاسبات فقط برای مقادیر ستون‌های بازه‌های ساعتی انجام می‌شود و ستون دودویی مربوط به روزها در این محاسبات دخیل نمی‌شوند. در گام بعد، برای تمام زوج‌های به‌صورت (ساعت/روز)، مقدار پشتیبانی با استفاده از رابطه ۳ به‌دست آورده می‌شوند. سپس برای آن زوج‌هایی که مقدار حداقل آستانه پشتیبانی را برآورده می‌سازند قواعدی به‌صورت (ساعت/روز) نوشته و مقدار اطمینان بر اساس رابطه ۵ محاسبه می‌شود. در نهایت، قواعد ارتباط شبکه‌ای که مقدار حداقل آستانه اطمینان را برآورده می‌سازند، به‌عنوان قواعد نهایی انتخاب می‌شوند (جدول ۶).

Archive of SID

www.SID.ir

جدول ۴. داده‌های خام پردازش شده (مقادیر کمّی) چهار هفتهٔ گذشته، آماده برای الگوریتم قواعد ارتباط شبکه‌ای

ساعت‌های شبانه روز																								روزهای هفته							وزن	هفته	تراکنش
۲۳ – ۲۴	۲۲ – ۲۳	۲۱ – ۲۲	۲۰ – ۲۱	۱۹ – ۲۰	۱۸ – ۱۹	۱۷ – ۱۸	۱۶ – ۱۷	۱۵ – ۱۶	۱۴ – ۱۵	۱۳ – ۱۴	۱۲ – ۱۳	۱۱ – ۱۲	۱۰ – ۱۱	...	۰۵ – ۰۴	۰۴ – ۰۳	۰۳ – ۰۲	۰۲ – ۰۱	۰۱ – ۰۰	جمعه	پنج شنبه	چهارشنبه	سه شنبه	دوشنبه	یکشنبه	شنبه			
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۵	۶۰	۵۱	۰	...	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰/۲	۱	۱
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۲۰	۱۰	۰	...	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰/۲	۱	۲
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۰	۱۷	۴۷	۰	۰	۰	۱	۰	۰	۰	۰/۳	۲	۱۱۸۳۵
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۲۸	۶۰	۴۶	۰	۰	۰	۱	۰	۰	۰	۰/۳	۲	۱۱۸۳۶
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۰	۰	۰	۰	۰	۰	۰	۲۲	۰	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰/۶	۳	۱۷۶۴۰
۰	۰	۰	۰	۱۴	۶۰	۶۰	۴۹	۰	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰/۶	۳	۱۷۶۴۱
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۶۰	۱۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۱	۴	۴۱۸۵۰
۱۹	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۱	۴	۴۱۸۵۱

Archive of SID
www.SID.ir

جدول ۵. مقادیر کمّی وزن‌دار به همراه مقادیر mwu محاسبه شده برای هر تراکنش

mwu	ساعت‌های شبانه روز																							روز های هفته							وزن	هفته	تراکنش
	۲۳ – ۲۴	۲۲ – ۲۳	۲۱ – ۲۲	۲۰ – ۲۱	۱۹ – ۲۰	۱۸ – ۱۹	۱۷ – ۱۸	۱۶ – ۱۷	۱۵ – ۱۶	۱۴ – ۱۵	۱۳ – ۱۴	۱۲ – ۱۳	۱۱ – ۱۲	...	۰۳ – ۰۴	۰۲ – ۰۳	۰۱ – ۰۲	۰۰ – ۰۱	جمعه	پنج شنبه	چهارشنبه	سه شنبه	دوشنبه	یکشنبه	شنبه								
۷/۷۳	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۱۲	۱۰/۲	...	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰/۲	۱	۱
۲	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۴	۲	...	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰/۲	۱	۲
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۹/۶	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۵/۱	۱۴/۱	۰	۰	۰	۱	۰	۰	۰	۰/۳	۲	۱۱۸۳۵
۱۳/۴	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۸/۴	۱۸	۱۳/۸	۰	۰	۰	۱	۰	۰	۰	۰/۳	۲	۱۱۸۳۶
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۱۳/۲	۰	۰	۰	۰	۰	۰	۰	۱۳/۲	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰/۶	۳	۱۷۶۴۰
۲۷/۵	۰	۰	۰	۸/۴	۳۶	۳۶	۳۶	۲۹/۴	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰/۶	۳	۱۷۶۴۱
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۳۵/۵	۶۰	۱۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰/۶	۴	۴۱۸۵۰
۱۴/۵	۱۹	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	...	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰/۶	۴	۴۱۸۵۱
۳۳۶۶۸۷/۷	مجموع																										

Archive of SID
www.SID.ir

جدول ۶. قواعد ارتباط شبکه‌ای نهایی

conf%	sup%	قانون (ساعت/روز)	conf%	sup%	قانون (ساعت/روز)
۹/۵	۱/۳	(۲۰ - ۱۹) ⟶ سه‌شنبه	۱۰/۹	۱/۵	(۰۱ - ۰۰) ⟶ شنبه
۹/۶	۱/۴	(۲۱ - ۲۰) ⟶ سه‌شنبه	۹/۸	۱/۴	(۱۴ - ۱۳) ⟶ شنبه
۱۰/۲	۱/۴	(۲۲ - ۲۱) ⟶ سه‌شنبه	۱۰/۳	۱/۴	(۱۵ - ۱۴) ⟶ شنبه
۱۲/۱	۱/۷	(۲۳ - ۲۲) ⟶ سه‌شنبه	۹/۹	۱/۴	(۱۶ - ۱۵) ⟶ شنبه
۱۳/۳	۱/۹	(۲۴ - ۲۳) ⟶ سه‌شنبه	۹/۵	۱/۳	(۱۷ - ۱۶) ⟶ شنبه
۱۳/۱	۱/۹	(۰۱ - ۰۰) ⟶ چهارشنبه	۹	۱/۳	(۱۸ - ۱۷) ⟶ شنبه
۱۰/۲	۱/۵	(۰۲ - ۰۱) ⟶ چهارشنبه	۹/۴	۱/۳	(۲۰ - ۱۹) ⟶ شنبه
۹/۸	۱/۴	(۱۶ - ۱۵) ⟶ چهارشنبه	۹/۶	۱/۳	(۲۱ - ۲۰) ⟶ شنبه
۹/۷	۱/۴	(۱۷ - ۱۶) ⟶ چهارشنبه	۱۰/۸	۱/۵	(۲۲ - ۲۱) ⟶ شنبه
۹/۷	۱/۴	(۲۰ - ۱۹) ⟶ چهارشنبه	۱۳/۱	۱/۸	(۲۳ - ۲۲) ⟶ شنبه
۹/۵	۱/۴	(۲۱ - ۲۰) ⟶ چهارشنبه	۱۲/۹	۱/۸	(۲۴ - ۲۳) ⟶ شنبه
۹/۴	۱/۴	(۲۲ - ۲۱) ⟶ چهارشنبه	۱۲/۴	۱/۸	(۰۱ - ۰۰) ⟶ یکشنبه
۱۳/۱	۱/۹	(۲۳ - ۲۲) ⟶ چهارشنبه	۱۰/۶	۱/۵	(۱۵ - ۱۴) ⟶ یکشنبه
۱۳	۱/۹	(۲۴ - ۲۳) ⟶ چهارشنبه	۱۰/۳	۱/۵	(۱۶ - ۱۵) ⟶ یکشنبه
۱۳	۱/۹	(۰۱ - ۰۰) ⟶ پنجشنبه	۹	۱/۳	(۱۷ - ۱۶) ⟶ یکشنبه
۱۰/۱	۱/۵	(۰۲ - ۰۱) ⟶ پنجشنبه	۹/۴	۱/۴	(۲۱ - ۲۰) ⟶ یکشنبه
۱۰/۵	۱/۵	(۱۵ - ۱۴) ⟶ پنجشنبه	۹/۹	۱/۴	(۲۲ - ۲۱) ⟶ یکشنبه
۱۰/۵	۱/۵	(۱۶ - ۱۵) ⟶ پنجشنبه	۱۱/۹	۱/۷	(۲۳ - ۲۲) ⟶ یکشنبه
۹	۱/۳	(۱۷ - ۱۶) ⟶ پنجشنبه	۱۳	۱/۹	(۲۴ - ۲۳) ⟶ یکشنبه
۹/۴	۱/۴	(۲۱ - ۲۰) ⟶ پنجشنبه	۱۳/۲	۱/۸	(۰۱ - ۰۰) ⟶ دوشنبه
۹/۹	۱/۵	(۲۲ - ۲۱) ⟶ پنجشنبه	۱۰/۱	۱/۴	(۰۲ - ۰۱) ⟶ دوشنبه
۱۳/۱	۱/۹	(۲۳ - ۲۲) ⟶ پنجشنبه	۹	۱/۲	(۱۳ - ۱۲) ⟶ دوشنبه
۱۲/۲	۱/۸	(۲۴ - ۲۳) ⟶ پنجشنبه	۹/۶	۱/۳	(۱۴ - ۱۳) ⟶ دوشنبه
۱۲/۳	۱/۸	(۰۱ - ۰۰) ⟶ جمعه	۱۰/۲	۱/۴	(۲۰ - ۱۹) ⟶ دوشنبه
۱۱	۱/۶	(۰۲ - ۰۱) ⟶ جمعه	۹/۸	۱/۳	(۲۱ - ۲۰) ⟶ دوشنبه
۱۰/۲	۱/۵	(۱۳ - ۱۲) ⟶ جمعه	۱۰/۵	۱/۴	(۲۲ - ۲۱) ⟶ دوشنبه
۱۱/۵	۱/۷	(۱۴ - ۱۳) ⟶ جمعه	۱۱/۷	۱/۶	(۲۳ - ۲۲) ⟶ دوشنبه
۱۱/۵	۱/۷	(۱۵ - ۱۴) ⟶ جمعه	۱۱/۷	۱/۶	(۲۴ - ۲۳) ⟶ دوشنبه
۱۱/۸	۱/۸	(۱۶ - ۱۵) ⟶ جمعه	۱۱/۷	۱/۷	(۰۱ - ۰۰) ⟶ سه‌شنبه
۱۱/۳	۱/۷	(۱۷ - ۱۶) ⟶ جمعه	۹/۳	۱/۳	(۰۲ - ۰۱) ⟶ سه‌شنبه
۹/۵	۱/۴	(۱۸ - ۱۷) ⟶ جمعه	۹/۳	۱/۳	(۱۵ - ۱۴) ⟶ سه‌شنبه
۱۰/۲	۱/۵	(۲۱ - ۲۰) ⟶ جمعه	۹/۱	۱/۳	(۱۶ - ۱۵) ⟶ سه‌شنبه
۹/۸	۱/۵	(۲۲ - ۲۱) ⟶ جمعه	۹	۱/۳	(۲۳ - ۲۲) ⟶ جمعه

Archive of SID
www.SID.ir

جهت درک بهتر قواعد به‌دست آمده، آنها را به‌صورت یک جدول نشان می‌دهیم به‌طوری‌که سطرها نشان دهنده روزها و ستون‌ها نشان دهنده ساعت می‌باشند. اعداد درون هر خانه مربوط به مقدار پشتیبانی قاعده و رنگ هر خانه نشان دهنده میزان اطمینان می‌باشد. (جدول ۷)

جدول ۷. نمایش دیداری قواعد ارتباط شبکه‌ای به‌دست آمده

روزهای هفته	ساعت‌های شبانه روز																					
	۰۰ – ۰۱	۰۱ – ۰۲	۰۲ – ۰۳	.. – .	۱۱ – ۱۲	۱۲ – ۱۳	۱۳ – ۱۴	۱۴ – ۱۵	۱۵ – ۱۶	۱۶ – ۱۷	۱۷ – ۱۸	۱۸ – ۱۹	۱۹ – ۲۰	۲۰ – ۲۱	۲۱ – ۲۲	۲۲ – ۲۳	۲۳ – ۲۴
شنبه	۱/۵			...			۱/۴	۱/۴	۱/۴	۱/۳	۱/۳		۱/۳	۱/۳	۱/۵	۱/۸	۱/۸
یکشنبه	۱/۸			...				۱/۵	۱/۵	۱/۳				۱/۴	۱/۴	۱/۷	۱/۹
دوشنبه	۱/۸	۱/۴		...		۱/۲	۱/۳						۱/۴	۱/۳	۱/۴	۱/۶	۱/۶
سه شنبه	۱/۷	۱/۳		...				۱/۳	۱/۳				۱/۳	۱/۴	۱/۴	۱/۷	۱/۹
چهارشنبه	۱/۹	۱/۵		...					۱/۴	۱/۴			۱/۴	۱/۴	۱/۴	۱/۹	۱/۹
پنج شنبه	۱/۹	۱/۵		...				۱/۵	۱/۵	۱/۳				۱/۴	۱/۵	۱/۹	۱/۸
جمعه	۱/۸	۱/۶		...		۱/۵	۱/۷	۱/۷	۱/۸	۱/۷	۱/۴			۱/۵	۱/۵	۱/۳	

بیش از ۱۳ ٪ ۱۳ ٪ - ۱۲٪ ۱۲ ٪ - ۱۱ ٪ ۱۱ ٪ - ۱۰ ٪ ۱۰ ٪ - ۹ ٪ کمتر از ۹ ٪

با توجه به جدول ۷، الگوهای هفتگی مصرف اینترنت کاربران طی چهار هفتهٔ گذشته را می‌توان به این صورت تفسیر نمود:

- به‌طور کلی در تمام روزهای هفته، بازه ساعت ۲ بامداد الی ۱۲ ظهر نسبت به ساعت‌های دیگر شبانه روز دارای کمترین میزان استفاده توسط کاربران می‌باشد به‌طوری‌که در این بازهٔ زمانی، هیچ‌گونه قاعده‌ای به‌دست نیامده است.
- در روزهای شنبه تا پنجشنبه، بازهٔ ساعت ۱۴ الی ۲۲ دارای ترافیک متوسطی می‌باشد. البته همان‌طور که از جدول مشخص می‌باشد، در این روزها در بازهٔ ساعت ۱۷ الی ۱۹ میزان استفاده بسیار کاهش می‌یابد.
- بیشترین میزان استفاده کاربران در هفته را می‌توان در بازه ساعت ۲۲ الی ۲۴ روزهای شنبه تا پنجشنبه و بازهٔ ۰۰ الی ۰۱ بامداد روزهای یکشنبه تا جمعه و بازهٔ ساعت ۱۳ الی ۱۷ روز جمعه مشاهده نمود.
- نوع رفتار کاربران در روزهای شنبه تا پنجشنبه (روزهای کاری) تقریباً مشابه هم می‌باشند و بیشترین تمرکز استفاده در بازه‌های ۰۰ الی ۰۱ بامداد و ۲۲ الی ۲۴ می‌باشد. این در

Archive of SID
www.SID.ir

حالی است که در روز جمعه (روز تعطیل) بیشترین تمرکز استفاده کاربران در بازه‌های ساعت ۰۰ الی ۰۲ بامداد و ۱۳ الی ۱۷ می‌باشد.

- آخر شب روز جمعه میزان استفاده بسیار کاهش می‌یابد در حالی که آخر شب روزهای دیگر هفته، اوج مصرف کاربران می‌باشد.

در حال حاضر شیفت‌های مرکز پشتیبانی به‌ترتیب عبارتند از ۱۶-۸، ۲۴-۱۶ و ۸-۲۴. با توجه به الگوهای به‌دست آمده، می‌توان در شیفت‌های تعریف شده برای کارمندان مرکز پشتیبانی فنی تغییراتی را اعمال کرد. پس از مشورت با مدیر ISP پیشنهاد شد شیفت‌های روزهای کاری برای هفته‌ای که پیش رو است، به این صورت تغییر کنند:

- شیفت اول از ۲ بامداد الی ۱۲ ظهر (روز شنبه و یکشنبه از ۱ بامداد الی ۱۲ ظهر)
- شیفت دوم از ۱۲ ظهر الی ۲۰
- شیفت سوم از ۲۰ الی ۲ بامداد روز بعد

در شیفت اول نیازی نیست که بهترین فرد بخش پشتیبانی برای پاسخ‌گویی به کاربران انتخاب شود. برای این شیفت می‌توان از فردی که نسبت به دیگران از تخصص و کارایی کمتری برخوردار است به‌طور پاره وقت استفاده نمود. بنابراین برای این شیفت، شرکت هزینهٔ زیادی را متحمل نمی‌شود. اما کسانی که برای شیفت‌های دوم و سوم انتخاب می‌شوند باید از تخصص کافی در پاسخ‌گویی به کاربران و حل مشکلات مختلف اینترنتی آنان برخودار باشند. زیرا اکثر کاربران، در این شیفت‌ها از اینترنت استفاده می‌کنند. شیفت دوم بهتر است به‌صورت تمام وقت باشد و پیشنهاد می‌شود که فرد پاسخ‌گو در شرکت حضور داشته باشد نه در خارج از شرکت. اما برای شیفت سوم حتماً باید متخصص‌ترین فرد بخش پشتیبانی فنی انتخاب شود و باید به‌طور تمام وقت پاسخ‌گوی کاربران باشد. از آنجایی که این شیفت در شب می‌باشد، امکان اینکه فرد متخصص در شرکت حضور داشته باشد، نیست. بنابراین بایستی امکانات مناسبی در منزل آن فرد فراهم شود تا در مدت زمان مسئولیتش بتواند خدمات خوبی را به کاربران عرضه نماید.

از آنجایی که در روز جمعه الگوی مصرفی کاربران اینترنت متفاوت می‌باشد، بنابراین شیفت‌ها نیز متفاوت خواهند بود. در روز جمعه، ساعات اوج مصرف شامل بازهٔ ساعت ۰۰ الی ۰۲ بامداد و همچنین ۱۲ ظهر الی ۱۷ می‌باشد. بازه‌ی۰۰ الی ۰۲ بامداد توسط شیفت سوم روز پنجشنبه پوشش داده می‌شود. ۲۲ ساعت باقی‌مانده به سه شیفت تقسیم می‌شود:

- شیفت اول از ۲ بامداد الی ۱۲ ظهر (۱۰ ساعت)
- شیفت دوم از ۱۲ ظهر الی ۱۸ عصر (۶ ساعت)
- شیفت سوم از ۱۸ عصر الی ۱ بامداد روز بعد (۷ ساعت)

Archive of SID

www.SID.ir

در جدول ۸ شمای کلی شیفت‌های جدید کارمندان بخش پشتیبانی فنی نشان داده شده است. اگر کارمندان مرکز پشتیبانی را از نظر تخصص در سه سطح نشان دهیم آنگاه رنگ سفید نشان دهنده شیفت کاری فردی با سطح تخصص ۳، رنگ خاکستری نشان دهنده شیفت کاری فردی با سطح تخصص ۲ و رنگ سیاه نشان دهنده شیفت کاری فردی با سطح تخصص ۱ می‌باشد.

جدول ۸. شمای کلی شیفت‌های جدید کارمندان بخش پشتیبانی

روزهای هفته	ساعت های شبانه روز																							
	۰۰ - ۰۱	۰۱ - ۰۲	۰۲ - ۰۳	.. .	۱۱ - ۱۲	۱۲ - ۱۳	۱۳ - ۱۴	۱۴ - ۱۵	۱۵ - ۱۶	۱۶ - ۱۷	۱۷ - ۱۸	۱۸ - ۱۹	۱۹ - ۲۰	۲۰ - ۲۱	۲۱ - ۲۲	۲۲ - ۲۳	۲۳ - ۲۴	۲۲ - ۲۳	۲۱ - ۲۲	۲۰ - ۲۱	۱۹ - ۲۰	۱۸ - ۱۹	۱۷ - ۱۸	۱۶ - ۱۷
شنبه				...																				
یکشنبه				...																				
دوشنبه				...																				
سه شنبه				...																				
چهارشنبه				...																				
پنج شنبه				...																				
جمعه				...																				

جدول ۹. الگوهای مصرف کاربران پس از گذشت یک دورهٔ زمانی (هفته)

روزهای هفته	ساعت‌های شبانه روز																							
	۰۰ - ۰۱	۰۱ - ۰۲	۰۲ - ۰۳	...	۱۱ - ۱۲	۱۲ - ۱۳	۱۳ - ۱۴	۱۴ - ۱۵	۱۵ - ۱۶	۱۶ - ۱۷	۱۷ - ۱۸	۱۸ - ۱۹	۱۹ - ۲۰	۲۰ - ۲۱	۲۱ - ۲۲	۲۲ - ۲۳	۲۳ - ۲۴	۲۲ - ۲۳	۲۱ - ۲۲	۲۰ - ۲۱	۱۹ - ۲۰	۱۸ - ۱۹	۱۷ - ۱۸	۱۶ - ۱۷
شنبه	۱/۶			...		۱/۳	۱/۴	۱/۵	۱/۴	۱/۴		۱/۳	۱/۳	۱/۴	۱/۶	۲	۱/۹							
یکشنبه	۱/۹	۱/۴		...				۱/۶	۱/۶	۱/۴			۱/۴	۱/۵	۱/۷	۱/۹	۲							
دوشنبه	۱/۷	۱/۶		...			۱/۴	۱/۳	۱/۴				۱/۴		۱/۴	۱/۵	۱/۶							
سه شنبه	۱/۶	۱/۴		...				۱/۳					۱/۴	۱/۵	۱/۵	۱/۶	۱/۷							
چهارشنبه	۱/۸	۱/۴		...			۱/۴		۱/۵	۱/۶	۱/۵	۱/۶		۱/۴	۱/۵	۱/۸	۱/۷							
پنج شنبه	۱/۷	۱/۶		...					۱/۳					۱/۳	۱/۴	۱/۶	۱/۴							
جمعه	۱/۵	۱/۳		...		۱/۳	۱/۴	۱/۴	۱/۵	۱/۴	۱/۲	۱/۳	۱/۳	۱/۴	۱/۴	۱/۳								

بیش از ۱۳ ٪ | ۱۳ ٪ - ۱۲٪ | ۱۲ ٪ - ۱۱ ٪ | ۱۱ ٪ - ۱۰ ٪ | ۱۰ ٪ - ۹ ٪ | کمتر از ۹ ٪

Archive of SID
www.SID.ir

در دوره‌های زمانی (هفته) بعد نیز فرآیند شناسایی الگوهای مصرفی کاربران تکرار می‌شود. جدول ۹، الگوهای مصرف اینترنت را پس از گذشت یک دورهٔ زمانی نشان می‌دهد. این الگوها با استفاده از داده‌های چهار هفته از تاریخ ۸۶/۱۰/۸ الی ۸۶/۱۱/۵ به‌دست آمده‌اند. همانطور که از جدول‌ها دیده می‌شود، الگوهای مصرفی کاربران به‌دست آمده طی این دو دوره تقریباً دارای ساختاری مشابه هم می‌باشند درصورتی‌که این الگوها تغییرات مشخصی کنند آنگاه نیاز است که مدیریت بر اساس تغییرات به‌وجود آمده، تغییراتی را در نوع چیدمان کاری نیروهای متخصص خود ایجاد کند تا همواره کیفیت ارائهٔ خدمات با وجود تغییر در الگوی رفتاری کاربران، در سطح مناسبی قرار داشته باشد.

۶. نتیجه‌گیری و پیشنهادها

در این تحقیق، ابتدا یک مدل داده‌کاوی تکاملی جهت شناسایی مستمر الگوهای مصرفی کاربران اینترنت پیشنهاد داده شد و سپس از آن مدل جهت شناسایی مستمر الگوهای مصرفی یک شرکت خدمات دهنده اینترنت استفاده گردید. با بررسی قواعد به‌دست آمده مشاهده شد بیشترین میزان استفاده کاربران در هفته، در بازه ساعت ۲۲ الی ۲۴ روزهای شنبه تا پنجشنبه و بازهٔ ۰۰ الی ۰۱ بامداد روزهای یکشنبه تا جمعه و بازه ساعت ۱۳ الی ۱۷ روز جمعه است. در نهایت، با استفاده از الگوهای به‌دست آمده از مدل پیشنهادی و شناسایی ساعات اوج مصرف اینترنت، پیشنهاداتی جهت تغییر در شیفت‌های کاری کارمندان بخش پشتیبانی فنی ارائه شد. اجرای این تغییرات، دو مزیت را در پی داشت:

۱- در ساعات اوج مصرف کاربران، از افراد بسیار متخصص و به‌صورت تمام‌وقت، جهت پاسخ‌گویی هر چه بهتر به کاربران اینترنت استفاده می‌شود.

۲- دستمزد ساعتی افراد متخصص سطح بالای شرکت‌های ISP، نسبت به متخصصان سطح پایین‌تر بیشتر است. از آنجایی که ساعات اوج مصرف در طول ۲۴ ساعت یک روز معمولاً زیر ۶ ساعت می‌باشد بنابراین با شناسایی ساعات اوج مصرف شبکه، از متخصصان سطح بالا تنها در آن ساعات استفاده می‌شود و دیگر نیازی نیست در ساعت‌های دیگر شبانه روز که ترافیک شبکه کم می‌باشد از آنها استفاده شود. با این کار، در هزینه‌های ISP نیز صرفه‌جویی شده است.

با بررسی مقالات و تحقیقات انجام شده، تاکنون پژوهش‌های بسیار کمی در حوزهٔ کاربرد داده‌کاوی برای شناسایی الگوهای رفتارهای کاربران اینترنت انجام شده است. با توجه به کارهای انجام شده در این مقاله و همچنین زمینه‌های بالقوه‌ی دیگر در حوزه‌ی شناسایی الگوهای رفتاری کاربران اینترنت، در ادامه، پیشنهاداتی را برای تحقیقات آینده ارائه می‌کنیم:

Archive of SID
www.SID.ir

- شناسایی الگوهای مصرفی کاربران اینترنت در روزهای خاصی از سال با استفاده از داده‌های چندین سال (مثلاً ایام نوروز، ایام امتحانات دانشجویان، ایام ماه رمضان و غیره)
- شناسایی الگوهای مصرفی کاربران اینترنت در شهرهای مختلف کشور و مقایسه آنها
- تحلیل رفتارهای مصرفی مشتریان با استفاده از تحلیل فرآیندهای تصادفی و پیش‌بینی رفتارهای آینده با استفاده از این تحلیل
- استفاده از الگوریتم‌های خوشه‌بندی جهت شناسایی کاربران اینترنتی که قصد ترک ISP را دارند.

Archive of SID

www.SID.ir

مآخذ

Agrawal, R., Imielinski, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases*, Proceedings of the ACM SIGMOD Conference on Management of Data, 207–216.

Bodon, F. (2003) *A fast Apriori implementation*, In ICDM Workshop on Frequent Itemset Mining Implementation, vol.90, Melbourne, Florida, USA.

Jiang, W. (2007). A statistical process control approach to business activity monitoring. IIE *Transactions* 39. 235-249.

Li, S. T., & Shue, L. Y. (2008). *Business intelligence approach to supporting strategy-making of ISP service management*, Expert Systems with Applications, 35, 739–754

Madden, G. (1999). *Subscriber churn in the Australian ISP market*, Information Economics and Policy, 11, 195-207.

McCue, A. (2006). TalkTalk is UK's least-liked broadband provider. Broadband & ISPs.

Mostafavi, A., & Sabbari, H. (2009). *A dynamic model for mining association rules from weighted quantitative data*, Third Conference of Data Mining, Science and Technology University,Tehran, Iran (in Persian).

Pettersson, M. (2004). *SPC with application to churn Management*, Quality and Reliability Engineering International, 20, 397-406.

Sulaiman Khan, M., Muyeba, M., & Coenen, F. (2008). *A weighted utility framework for mining association rules*, Second UKSIM European Symposium on Computer Modeling and Simulation.

Wayland, R., Paul, E. & Cole, M. (1997). *Customer connections: New strategies for growth*, Harvard Business School Press.

Archive of SID

www.SID.ir